A cura di Vincenzo Crupi, Gian Franco Gensini,
Matteo Motterlini

LA DIMENSIONE COGNITIVA DELL'ERRORE IN MEDICINA

Con scritti di Arkes, Dawes, Detweiler, Eddy, Fahey, Fischhoff, Forrow, Gigerenzer, Green, Gruppen, Kahneman, McNeil, Meehl, Pauker, Poses, Redelmeier, Salovey, Schwartz, Shafir, Sox, Tversky

FrancoAngeli

Vincenzo Crupi svolge attività di ricerca presso il Dipartimento di Scienze della Cognizione e della Formazione, il Laboratorio di Scienze Cognitive dell'Università di Trento e il Laboratoire de Sciences Cognitives dell'Università di Aix-Marsiglia I. Con Matteo Motterlini è autore di *Decisioni mediche. Un punto di vista cognitivo* (2005).

Gian Franco Gensini è preside della Facoltà di Medicina e Chirurgia dell'Università di Firenze. Professore di Medicina interna e Cardiologia e direttore dell'Unità ospedaliera di Clinica medica e Cardiologia dell'Università di Firenze, dirige l'edizione italiana della rivista "Evidence-based medicine" ed è fondatore e presidente del Centro italiano di Evidence-based medicine.

Matteo Motterlini è professore associato di Filosofia della Scienza e di Economia cognitiva presso la Facoltà di Filosofia dell'Università Vita-Salute San Raffaele di Milano. È membro fondatore del Centro di ricerca in Epistemologia sperimentale e applicata (CRESA) presso la stessa università. Fra le sue ultime pubblicazioni: *Economia cognitiva e sperimentale* (a cura di, con Francesco Guala) (2005) e *Critica della ragione economica* (a cura di, con Massimo Piattelli Palmarini) (2005).

La traduzione dei saggi è di Vincenzo Crupi.
Le traduzioni non sono da intendersi versione ufficiale degli editori di riferimento,
per qualsiasi chiarimento si rimanda all'originale.

Copyright © 2006 by FrancoAngeli s.r.l., Milano, Italy

Ristampa							Anno							
0	1	2	3	4	5	6	2006	2007	2008	2009	2010	2011	2012	2013

È vietata la riproduzione, anche parziale, effettuata a qualsiasi titolo, eccetto quella ad uso personale.
Quest'ultima è consentita nel limite massimo del 15% delle pagine dell'opera, anche se effettuata
in più volte, e alla condizione che vengano pagati i compensi stabiliti dall'art. 2 della legge vigente.
Ogni fotocopia che eviti l'acquisto di un libro è illecita ed è severamente punita.
Chiunque fotocopia un libro, chi mette a disposizione i mezzi per farlo, chi comunque favorisce
questa pratica commette un reato e opera ai danni della cultura.
Stampa: Tipomonza, via Merano 18, Milano.

Indice

Premessa dei curatori pag. 7

Introduzione. Errori e decisioni in medicina, di *Matteo Motterlini, Vincenzo Crupi* » 11

I. Giudizi e probabilità

1. Il ragionamento probabilistico nella medicina clinica: problemi e opportunità, di *David M. Eddy* » 45

2. Giudizi di probabilità in medicina: trascurare le possibilità implicite, di *Donald A. Redelmeier, Derek J. Koehler, Varda Liberman, Amos Tversky* » 69

3. La raccolta e l'integrazione di informazioni come fonti di errore nelle decisioni diagnostiche, di *Larry D. Gruppen, Fredric M. Wolf, John E. Billi* » 75

4. I giudizi diagnostici in casi di sospetta batteriemia: euristica della disponibilità e *wishful thinking*, di *Roy M. Poses, Michele Anthony* » 91

5. L'influenza di informazioni pseudodiagnostiche nella valutazione dell'ischemia miocardica, di *Lee A. Green, J. Frank Yates* » 111

6. La distorsione retrospettiva nella valutazione della probabilità delle diagnosi, di *Hal R. Arkes, Paul D. Saville, Robert L. Wortmann, Allan R. Harkness* » 123

II. Preferenze e decisioni

7. L'elicitazione delle preferenze per terapie alternative, di *Barbara J. McNeil, Stephen G. Pauker, Harold C. Sox, Amos Tversky* pag. 131

8. È sempre preferibile avere più opzioni? L'effetto di attrazione nelle scelte di trattamento, di *Janet A. Schwartz, Gretchen B. Chapman* » 141

9. L'ingannevole ricerca di più informazioni, di *Donald A. Redelmeier, Eldar Shafir, Prince S. Aujla* » 159

10. Assolutamente relativo: come la presentazione dei risultati della ricerca medica influenza le scelte di trattamento, di *Lachlan Forrow, William C. Taylor, Robert M. Arnold* » 169

11. Finanziamenti basati sull'evidenza: la comprensione dei risultati della ricerca clinica, di *Tom Fahey, Sian Griffiths, Tim J. Peters* » 175

III. Valutazioni e consigli

12. Le consulenze sull'AIDS per persone a basso rischio, di *Gerd Gigerenzer, Ulrich Hoffrage, Axel Ebert* » 185

13. Giudizio clinico *vs.* giudizio statistico, di *Robyn M. Dawes, David Faust, Paul E. Meehl* » 205

14. Ricordi di una colonoscopia: uno studio randomizzato, di *Donald A. Redelmeier, Joel Katz, Daniel Kahneman* » 225

15. Effetto di incorniciamento e protezione solare: i vantaggi di una presentazione in termini di guadagni, di *Jerusha B. Detweiler, Brian T. Bedell, Peter Salovey, Emiliy Pronin, Alexander J. Rothman* » 239

16. Un approccio cognitivo alle decisioni sul cancro: implicazioni per la prevenzione e il trattamento, di *Baruch Fischhoff* » 257

Riferimenti bibliografici » 277

Gli autori » 301

Premessa

La maggior parte delle decisioni mediche sono prese in condizioni di incertezza, in cui la stima delle probabilità di un evento o degli effetti di un trattamento rappresenta l'elemento di riferimento. Per esempio: se una donna incinta accusa dolore e gonfiore alle gambe, il suo medico deve valutare la possibilità che sia affetta da tromboflebite (una complicazione relativamente comune in gravidanza), la probabilità che si verifichi un'embolia nel caso che non vengano somministrati anticoagulanti e i possibili rischi per il nascituro nel caso che essi vengano somministrati. In un paziente anziano con emicrania e febbre ricorrenti, è necessario considerare la diagnosi di arterite temporale e la possibilità di una terapia steroidea, con i suoi benefici (una riduzione del rischio di cecità) ed eventuali effetti collaterali (come la necrosi ossea). Di fronte a un paziente giovane con un neo sul torace la cui immagine dermoscopica risulta sospetta è necessario stabilire se il rischio dello sviluppo di una patologia grave (come un melanoma) giustifichi l'escissione chirurgica.

Le decisioni mediche quindi implicano la stima della probabilità dell'esistenza di una condizione patologica, della sua prognosi, dell'esistenza di altri elementi clinici, dell'utilità e degli effetti avversi di possibili trattamenti, e della loro accettabilità da parte del malato.

La qualità delle decisioni è evidentemente un fattore determinante per la qualità delle cure. Per questo motivo, nei curricula delle Facoltà di Medicina sono stati recentemente introdotti diversi insegnamenti volti a migliorare tali decisioni. In particolare, la diffusione di conoscenze mediche scientificamente aggiornate rientra nell'ambito di intervento della *evidence based medicine*, mentre per quanto riguarda il corretto impiego delle informazioni disponibili nel ragionamento, nel giudizio e nella decisione il livello ottimale è fissato dai principi formali della teoria della scelta razionale. Tali principi indicano le regole che gli individui *dovrebbero* seguire per elaborare correttamente le informazioni incerte che ricevono dall'ambiente e per individuare coerentemente i mezzi disponibili in grado di massimizzare le loro possibilità di suc-

cesso. Nonostante gli sforzi fatti, però, gli studi documentano che l'applicazione di tecniche formali alle decisioni mediche è rimasta sporadica e non ha consentito di ridurre in modo soddisfacente la diffusione degli errori.

Questo volume avanza una possibile spiegazione di tale insuccesso: essa risiede in larga misura nei processi cognitivi di tipo intuitivo a cui medici e pazienti (come tutti noi) si affidano nel fronteggiare l'incertezza.

A partire dagli anni Settanta, in una serie di importanti esperimenti sugli aspetti procedurali del giudizio e della decisione, Amos Tversky e Daniel Kahneman hanno documentato diversi casi in cui gli individui violano sistematicamente i fondamentali principi della razionalità. Tali violazioni (dette *bias*) non possono essere facilmente spiegate con una mancanza di attenzione o di impegno. Per il loro carattere sistematico, esse somigliano piuttosto ad altri tipi di errori ben noti, come le illusioni percettive. Tversky e Kahneman hanno quindi suggerito che in molte circostanze tanto gli esperti quanto le persone comuni semplificano problemi relativamente complessi adottando strategie cognitive, dette *euristiche*, che spesso conducono a risultati erronei, cioè significativamente differenti da quelli indicati dai corretti principi formali pertinenti. Si tratta di errori che conseguono da ragionamenti che mettiamo in atto comunemente, in modo istintivo e automatico, e ciò spiega perché essi siano prevedibili.

Per quanto giovane, la ricerca cognitiva sulla decisione è un'area di studi già consolidata, come ha recentemente riconosciuto l'Accademia svedese delle scienze conferendo il Nobel per l'economia a Kahneman «per aver integrato intuizioni della ricerca psicologica nella scienza economica, specialmente nel campo del giudizio e della decisione in condizioni di incertezza» (lo stesso Tversky era stato fra i favoriti per l'assegnazione del Nobel prima di morire prematuramente nel 1996). Lo studio cognitivo della decisione si è quindi fatto strada nella scienza economica, e le sue implicazioni si rivelano non meno significative nell'ambito della decisione medica. È tipico infatti degli agenti economici (consumatori, imprenditori, investitori) dover selezionare l'opzione da preferire fra corsi di azione alternativi, ponderando i loro relativi costi e benefici in condizioni di incertezza e di rischio. Ebbene, numerosi problemi clinici rivelano caratteristiche analoghe, inducendo medici e pazienti a servirsi di scorciatoie cognitive che producono errori sistematici.

Il programma di ricerca inaugurato da Tversky e Kahneman ha dato luogo a un genuino e inedito progresso nella comprensione della razionalità umana e dei suoi limiti, trovando nell'ambito della decisione medica un fertile campo di applicazione. A conferma di ciò, da più di vent'anni esistono una associazione professionale internazionale per lo studio della decisione in medicina, la *Society for Medical Decision Making*, e una rivista specifica sul tema, «Medical Decision Making» (Sage Publications), che dedicano ampia attenzione a questa prospettiva di indagine. Inoltre, con sempre maggior frequen-

za, contributi allo studio della decisione medica basati sull'indagine empirica, di tipo sperimentale o sul campo, sono ospitati da riviste autorevoli come il «New England Journal of Medicine», il «Journal of the American Medical Association», e il «British Medical Journal».

Questo volume raccoglie alcuni dei più significativi contributi allo studio della dimensione cognitiva dell'errore e della decisione in medicina. La scelta dei saggi è determinata dai limiti di spazio ed è necessariamente influenzata dalle preferenze dei curatori. L'ampio numero dei contributi rilevanti in questo ambito di ricerca ha reso inevitabile l'esclusione di diversi importanti lavori e di alcuni autori influenti. Ciononostante, gli articoli qui raccolti e tradotti sono tutti di primo piano, e la loro diffusione è particolarmente importante per la formazione dei medici, a tutti i livelli di attività.

Nel selezionare i saggi, abbiamo cercato di fornire un utile punto d'accesso a un'ampia letteratura specialistica, introducendo i principali aspetti cognitivi dell'indagine diagnostica e delle scelte terapeutiche, bilanciando la presenza di studi sperimentali e sul campo e documentando la varietà delle specialità mediche nelle quali la ricerca cognitiva ha trovato applicazione.

Crediamo che i risultati della ricerca cognitiva siano imprescindibili per qualunque progetto volto a migliorare le decisioni: ragionare con più impegno, infatti, non può far scomparire un errore cognitivo più di quanto guardare con più attenzione possa far scomparire un'illusione ottica. Per mettere a punto procedure che ci permettano di evitare le conseguenze di errori di questo tipo è piuttosto necessario prendere atto della loro presenza e imparare a riconoscere le condizioni nelle quali si manifestano.

Vincenzo Crupi, Gian Franco Gensini, Matteo Motterlini
Gennaio 2006

Introduzione
Errori e decisioni in medicina

di *Matteo Motterlini, Vincenzo Crupi*

Diagnosi dell'errore medico

«Se dovessimo tollerare di convivere con un livello di efficienza del 99,9% avremmo due atterraggi a rischio al giorno nel solo aeroporto O'Hare di Chicago, e ogni ora ci sarebbero 16.000 recapiti postali falliti e 32.000 assegni bancari prelevati dal conto sbagliato». Questa osservazione dello statistico ed esperto di organizzazioni William Deming (citato in Leape, 1994, p. 1851) mostra quanto sia ridotto il margine di errore accettabile in molti importanti settori di attività. In confronto, la pratica medica rappresenta una notevole anomalia.

Si consideri il problema della correttezza delle diagnosi. Un tradizionale banco di prova consiste nel misurare l'accordo fra le diagnosi cliniche e quelle ricavate attraverso le autopsie. Indagini su ampi campioni di casi hanno rivelato un'incidenza di errori diagnostici gravi intorno al 15-20%, la metà circa dei quali con un probabile impatto sulla prognosi (Podbregar *et al.*, 2001; Shojania *et al.*, 2004; Tai *et al.*, 2001).

Gli errori diagnostici risultano particolarmente rilevanti in ambienti ad alto rischio, come i reparti di emergenza (O'Connor, 1995). In uno studio recente si è calcolato che fra i pazienti con dolori al petto che si presentano in pronto soccorso, l'incidenza delle diagnosi mancate di infarto cardiaco acuto, di infarto cardiaco e di angina instabile sono, rispettivamente, del 2, del 2,1 e del 2,3% (Pope, 2000; si veda anche McCarthy, 1993). Per contro, il ricovero di pazienti con basse probabilità di patologie coronariche acute determina in molti casi un impiego inefficiente delle risorse (Keffer, 1996; Apple, Henderson, 1999, p. 1178). Negli Stati Uniti, la mancata diagnosi di infarto cardiaco acuto è la principale causa di vertenze legali contro reparti di emergenza ospedalieri e medici di base (Rusnak, 1989) e, più in generale, un quinto dei ricorsi legali in campo medico riguarda errori diagnostici (Bartlett, 1998; si veda anche Weeks *et al.*, 2001).

Gli studi rivelano inoltre una allarmante variabilità nell'interpretazione di riscontri clinici strumentali, come la mammografia (Beam *et al.*, 1996), e una incidenza di errori che può arrivare fino al 40% per problemi diagnostici relativamente difficili, come la diagnosi di artrite psioriatica da parte dei reumatologi (Gorter, 2002) o la precoce identificazione di neoplasie maligne (Burton, Troxclair, Newman, 1998). Fra le diagnosi radiologiche o patologiche si stima che la percentuale di errore si aggiri intorno al 5% (Foucar, Foucar, 2000; Fitzgerald, 2001). Sulla base dell'analisi di più di un anno di attività ordinaria nel reparto di emergenza di un grande ospedale americano, per esempio, si è osservata una media del 3% di radiografie erroneamente interpretate come negative, considerando *solo* quelle con un impatto sulla prognosi del paziente (Espinosa, Nolan, 2000).

Alla luce di questi dati, è allarmante che l'errore diagnostico *non* compaia fra le priorità delle istituzioni sanitarie impegnate nel miglioramento delle prestazioni mediche e della sicurezza dei pazienti (Graber, 2005). Per non dire del fatto che meno di un medico su tre riporti di essersi imbattuto in un *qualsiasi* errore *nell'ultimo anno* (Blendon *et al.*, 2002).

Che cosa ci può insegnare la ricerca cognitiva riguardo all'errore diagnostico? Per cominciare essa suggerisce che tanto gli esperti quanto i profani manifestano, in diversi contesti e indipendentemente dal loro dominio di competenza, una generale difficoltà a stimare adeguatamente la propria fallibilità – un fenomeno noto come *overconfidence*.

Un esempio classico proviene dall'ambito giuridico, dove si è stabilito che la correlazione fra la correttezza delle identificazioni fatte da testimoni oculari e il loro livello di fiducia nella validità della loro opinione è generalmente inferiore a 0,25 (Sporer *et al.*, 1995). Un'analoga discrepanza fra accuratezza e fiducia è stata osservata anche in ambito medico ed è in parte alimentata dal fatto che, mentre molti errori sono *potenzialmente* dannosi, solo in una piccola frazione dei casi le loro conseguenze risultano manifeste e riconoscibili.

Dawson *et al.* (1993) hanno per esempio chiesto a 198 medici alle prese con 864 pazienti in condizioni critiche di stimare alcuni parametri emodinamici clinicamente rilevanti (come l'indice cardiaco e la resistenza vascolare sistemica) nei diversi pazienti prima che fosse possibile misurare direttamente quegli stessi parametri attraverso l'introduzione di un catetere cardiaco. I medici esprimevano in termini probabilistici la loro fiducia nella correttezza delle proprie stime. Si è così potuto osservare che il livello di fiducia dei medici nelle loro stime *non* era significativamente associato al livello di accuratezza di quelle stesse stime; inoltre i medici più esperti si mostravano più convinti della correttezza dei propri giudizi, ma *non* più accurati. Analogamente, Podbregar *et al.* (2001) hanno rilevato, attraverso le autopsie di pazienti di un reparto di terapia intensiva, una percentuale di errori diagnostici con probabile impatto sulla prognosi (intorno al 10%) che *non* differiva signi-

ficativamente in tre diversi gruppi di pazienti le cui diagnosi erano state considerate dai medici, rispettivamente, «assolutamente certe», «minimamente incerte» e «notevolmente incerte».

Se l'entità e la rilevanza degli errori sono sottovalutate, non meno preoccupante è la modesta comprensione che si ha della loro natura e delle loro cause. In un recente studio, i medici hanno indicato le principali fonti di errore nel personale infermieristico sottodimensionato e nel sovraccarico di lavoro senza neppure menzionare possibili cause cognitive (Blendon *et al.*, 2000). Eppure gli studi indicano che uno su sei errori medici «si verifica nel sintetizzare le informazioni disponibili, o nel decidere e agire alla luce di quelle informazioni» (Wilson *et al.*, 1999).

La componente cognitiva è di importanza centrale nel compito diagnostico e le diagnosi mancate hanno un'incidenza notevole fra gli errori medici. Fra gli errori riportati da più di cento specializzandi americani in un questionario anonimo (errori che avevano prodotto complicazioni serie nel 90% dei casi, e la morte nel 31%), quelli diagnostici erano i più comuni (33%); solo uno su due fra i giovani medici, però, dichiarava di aver discusso i suoi errori con il medico responsabile (Wu *et al.*, 1991). È da notare, inoltre, che la metà di un ampio campione di cittadini americani ha indicato nella sospensione dall'attività dei medici che commettono errori una misura efficace per il miglioramento delle pratiche cliniche (Blendon *et al.*, 2000).

Questi ultimi dati riflettono una tendenza, che gli stessi medici non di rado condividono, a considerare l'errore medico come un problema in gran parte legato a incompetenza o negligenza del singolo (Deskin, Hoye, 2004). L'approccio cognitivo al ragionamento clinico suggerisce una prospettiva differente. L'individuazione di errori cognitivi comuni e diffusi emerge infatti dalla crescente comprensione della natura della razionalità umana e dei suoi limiti (Simon, 1957, 1997; Kahneman, 2002). Il tratto distintivo di questi fenomeni è il fatto di presentarsi in modo sistematico e prevedibile. Pertanto, se alla radice di una diagnosi mancata si trova un errore che consegue dall'impiego di una «scorciatoia» cognitiva, rimuovere un professionista dal suo posto costituirà un modesto contributo al miglioramento delle cure. Molto probabilmente il suo sostituto sarà altrettanto fallibile.

Nella formazione professionale è ormai pratica comune introdurre i medici ai principi che determinano la qualità delle informazioni (l'evidenza clinica) e la loro corretta interpretazione (la teoria della probabilità e la statistica). Nella presente introduzione e nei saggi raccolti in questo volume avremo modo di vedere come sia necessario affiancare a queste competenze la consapevolezza della presenza degli errori nel ragionamento clinico. Nella misura in cui gli errori cognitivi contribuiscono in modo specifico a indebolire la qualità del giudizio medico, per ridurne gli effetti occorre conoscerne le cause.

Il ragionamento diagnostico può essere rappresentato come un processo di valutazione di ipotesi incerte (le possibili diagnosi) alla luce dell'acquisizione di informazioni imperfette (i dati clinici). Idealmente, tale processo comporta innanzi tutto l'individuazione di una lista di ipotesi diagnostiche compatibili con i primi dati clinici raccolti. Già a questo stadio, le diverse ipotesi diagnostiche potranno essere considerate più o meno probabili in funzione della loro maggiore o minore diffusione nella popolazione cui il paziente appartiene per età, sesso e sintomatologia. A questo punto, il compito del medico consisterà nel raccogliere informazioni aggiuntive che la conoscenza clinica identifica come maggiormente indicative di alcune delle possibili diagnosi a discapito di altre. L'iniziale plausibilità delle patologie considerate e la rilevanza degli ulteriori dati raccolti permetteranno così di «aggiornare» la probabilità delle diverse ipotesi diagnostiche, fino al punto in cui qualcuna di esse avrà raggiunto un livello di affidabilità sufficiente per orientare l'azione.

Un importante principio del calcolo delle probabilità, il *teorema di Bayes*, indica in che modo precisamente sia possibile aggiornare la probabilità di una o più ipotesi alla luce di nuove informazioni. Per questo motivo, in un lavoro recente e ben documentato sulla decisione e l'errore in medicina, il teorema di Bayes è stato definito «la stele di Rosetta del ragionamento clinico e il Sacro Graal per sfuggire agli errori diagnostici» (Wachter, Shojania, 2004, p. 112).

A titolo illustrativo, immaginate che sia stata effettuata un'estrazione da un'urna di venti palline, sette delle quali contengono un premio monetario (per esempio, 10 euro). In queste condizioni, il rapporto di probabilità in favore dell'ipotesi che sia stata estratta una pallina vincente è di sette contro tredici, o del 35% contro il 65%. Immaginate ora di sapere che nell'urna vi sono in tutto cinque palline di colore rosso: quattro di esse appartengono all'insieme delle (sette) palline vincenti e una appartiene all'insieme delle (tredici) palline non vincenti. Vi viene detto che la pallina estratta è di colore rosso. Alla luce di questa nuova informazione, il rapporto di probabilità in favore dell'ipotesi che la pallina estratta sia vincente risulta più che invertito: poiché delle cinque palline rosse, fra cui si trova quella estratta, quattro sono vincenti, le probabilità di vincita sono a questo punto di quattro contro uno, cioè dell'80% contro il 20%. Il teorema di Bayes riproduce in termini formali questa linea di ragionamento nel modo seguente (l'espressione $p(a|b)$ indica una probabilità *condizionata* ed è da intendersi come «la probabilità di *a* ammesso che *b*»):

$$\frac{p(\text{vincente}|\text{rossa})}{p(\text{non-vincente}|\text{rossa})} = \frac{p(\text{rossa}|\text{vincente})}{p(\text{rossa}|\text{non-vincente})} \times \frac{p(\text{vincente})}{p(\text{non-vincente})} = \frac{4/7}{1/13} \times \frac{7/20}{13/20} = \frac{4}{1}$$

La prima frazione da sinistra indica il «punto di arrivo» del ragionamento, cioè il rapporto di probabilità in favore di una pallina vincente alla luce dell'informazione relativa al colore. La seconda frazione da sinistra indica lo specifico contributo di quest'ultima informazione sotto forma di rapporto fra la proporzione relativamente alta di «veri positivi» (sono rosse quattro delle sette palline vincenti) e la bassa proporzione di «falsi positivi» (è rossa una sola delle tredici palline non vincenti). Infine la terza frazione da sinistra indica il «punto di partenza» del ragionamento, ovvero il rapporto di probabilità in favore di una pallina vincente «prima» che l'informazione relativa al colore venga considerata.

Ovviamente, l'analogia con il ragionamento diagnostico è imperfetta per molti motivi, non ultimo il fatto che in questo esempio sono coinvolte due sole ipotesi fra loro complementari (la pallina è vincente oppure no), mentre le ipotesi diagnostiche rilevanti in un caso clinico individuale possono essere ben di più (e non necessariamente si escludono l'un l'altra). L'estensione del teorema di Bayes alla valutazione di una molteplicità di ipotesi incerte non presenta particolari problemi matematici, ma comporta una certa difficoltà cognitiva che consiste nel fare a mente i calcoli necessari per la valutazione delle ipotesi – come ora vedremo passando da una ricostruzione ideale del ragionamento diagnostico a un caso reale.

La disponibilità

> Il signor D., un uomo afroamericano di 65 anni, si presenta al pronto soccorso per dolori alla parte superiore della schiena e ad altre parti del corpo che perdurano da diversi giorni. Riporta anche mal di gola e la percezione soggettiva di uno stato febbrile. In base all'esame fisico emerge che il paziente non ha febbre, che è presente un leggero eritema orofaringeo e che i polmoni sono liberi. Gli viene diagnosticata un'infezione del tratto respiratorio superiore e, dopo il prelevamento di campioni per colture dal sangue e dalla gola, viene dimesso con l'indicazione di mantenere regolare l'assunzione di fluidi e di prendere ibuprofene per sedare i dolori.

Il caso del signor D. è descritto e analizzato da uno dei più importanti studiosi degli errori cognitivi in ambito medico, Donald Redelmeier, professore di Medicina presso l'Università di Toronto e Direttore dell'unità di ricerca di Epidemiologia Clinica del Sunnybroook and Women's College Health Sciences Centre (si veda Redelmeier, 2005). È la storia di una serie di errori diagnostici e di scelte di trattamento inefficaci. Una storia di cui ci serviremo liberamente e che costituirà il nostro filo rosso nella discussione delle diverse «trappole cognitive» del ragionamento clinico.

La diagnosi di infezione del tratto respiratorio superiore assegnata al signor D. (che come vedremo si rivelerà scorretta) è stata elaborata sulla base di un insieme di dati clinici piuttosto ridotto, che non permetteva di escludere diverse altre possibili cause (per esempio di tipo non infettivo) attraverso un'analisi articolata come quella suggerita dal teorema di Bayes. In che modo allora si è giunti a questa diagnosi?

Come si è detto, la spiegazione cognitiva si rivolge ai processi intuitivi che guidano gli individui nelle loro scelte. Che cosa ci può pertanto insegnare a questo riguardo? Una delle scorciatoie cognitive più note, detta *euristica della disponibilità*, suggerisce una possibile spiegazione della diagnosi inizialmente mancata nel caso del signor D.

L'euristica della disponibilità induce a giudicare più alta la probabilità di un evento se un maggior numero di eventi simili può essere facilmente richiamato alla memoria. In una classica illustrazione di questo fenomeno, Tversky e Kahneman hanno osservato che la maggior parte delle persone ritiene più probabile che una parola finisca in «ing» piuttosto che la sua penultima lettera sia una «n», per il fatto che solitamente si richiamano alla mente con più facilità esempi del primo tipo che del secondo (Tversky, Kahneman, 1983). In una situazione del genere, il giudizio basato sulla disponibilità alla memoria produce una palese violazione di una regola fondamentale della teoria della probabilità: è semplicemente impossibile che la prima probabilità sia maggiore della seconda, perché qualunque esempio della seconda classe di parole è *anche* un esempio della prima. L'euristica della disponibilità può quindi, in certe condizioni, dar luogo a errori sistematici e prevedibili. Come nel caso del medico che riconducesse immediatamente il dolore alla schiena e la mialgia riportati dal signor D. a un'infezione del sistema respiratorio, confortato dal fatto che si tratta di un'associazione di eventi familiare che ricorda di avere osservato frequentemente.

Gli studi compresi nella prima parte di questo volume documentano come, al di là del valore illustrativo del singolo caso, i trabocchetti del ragionamento clinico possano rivelarsi diffusi e insidiosi. In particolare, il saggio *I giudizi diagnostici in casi di sospetta batteriemia* di Poses e Anthony presenta uno studio sul campo relativo a più di 200 casi reali, evidenziando l'influenza specifica dell'euristica della disponibilità nel distorcere i giudizi probabilistici dei medici.

La rappresentatività

Come ci si può aspettare, la storia del signor D. non finisce qui.

> Il giorno seguente il medico curante viene informato dal laboratorio dell'ospedale che l'emocoltura del signor D. è risultata positiva per lo *staphylococcus aureus*. Il paziente è quindi richiamato con urgenza al pronto soccorso. Il signor D. riporta sintomi simili a quelli del giorno precedente, con un peggioramento del dolore alla schiena e al collo.
> La storia clinica del signor D. comprende un tumore di Hodgkin diagnosticato 10 anni prima e (in base al resoconto del suo oncologo) curato con una combinazione di radio- e chemioterapia. Oltre a ciò, il paziente ha una lunga storia di *lichen planus* moderatamente severo e trattato in modo intermittente.
> L'esame fisico mostra diverse lesioni cutanee (polsi squamosi con papule eritematose, iperpigmentazione e placche ipertropiche nella parte anteriore delle gambe, escoriazioni diffuse sulle braccia). Non vi sono segni periferici di endocardite né segni che indichino l'uso di droghe per via endovenosa. Non si rilevano altre condizioni anomale all'esame fisico o in test di laboratorio. In particolare, una radiografia del torace rivela un profilo cardiaco normale e l'assenza di infiltrazioni o effusioni polmonari. Radiografie della spina cervicale e lombare mostrano alcune modificazioni degenerative, ma nessuna frattura o lesione indicativa di una patologia maligna.
> Il signor D. viene ricoverato e la mattina successiva una nuova visita rivela un soffio olosistolico 2/6 più intenso all'apice. Viene ordinata un'ecocardiografia transtoracica. Il soffio in seguito scompare e l'ecocardiogramma transtoracico non mostra lesioni valvolari o vegetazioni. Il signor D. è quindi sottoposto a trattamento antibiotico per via endovenosa (nafcillina) per contrastare la batteriemia.

Nell'approfondire la situazione del signor D. i medici hanno raccolto nuove informazioni (attraverso la storia clinica, l'esame fisico, e alcuni test strumentali) in modo da sottoporre a controllo alcune ipotesi diagnostiche (come disturbi cardiaci, polmonari, o spinali) e ne hanno concluso che fosse opportuno concentrarsi sul trattamento della batteriemia. Ma un'analisi più attenta del caso rivela alcuni limiti notevoli in questa linea di ragionamento.

Si consideri l'ipotesi diagnostica dell'endocardite (un'infiammazione che interessa le pareti interne del cuore). Non è facile dire quale sia la probabilità che un paziente con un quadro clinico simile al signor D. soffra di questa condizione, ma prescrivendo l'esecuzione di un esame specifico per controllarla (l'ecocardiografia transtoracica) i medici coinvolti hanno mostrato di non considerarla trascurabile, specie a seguito della rilevazione di un soffio cardiaco. Supponiamo che una stima ragionevole del loro «sospetto clinico» sia il 20%.

Ai fini della diagnosi di endocardite, l'ecocardiografia transtoracica ha un tasso di veri negativi (detto *specificità* del test) molto alto, ma un tasso di veri

positivi (detto *sensibilità* del test) decisamente modesto: le cifre riportate in letteratura sono, rispettivamente, del 98,6% e del 28,3% (Daniel *et al.*, 1991). A seguito dell'esame, si pone quindi il problema dell'interpretazione appropriata dei risultati di un test diagnostico strumentale relativamente affidabile ma imperfetto. L'esito negativo dell'esame ha indotto i medici a escludere la diagnosi di endocardite. Ma, come ora vedremo, si tratta di una conclusione affrettata.

Immaginiamo 1.000 pazienti con il quadro clinico del signor D. e seguiamo un ragionamento conforme al teorema di Bayes. In base alle nostre assunzioni, circa 800 di questi pazienti *non* sono affetti da endocardite (come implica una probabilità a priori del 20% per questa specifica condizione clinica) e, fra questi, 789 (il 98,6%) avranno quindi una ecocardiografia *negativa*. Ma fra i 200 pazienti che *sono* affetti da endocardite, 143 (il 71,7% di *falsi* negativi, dovuto a una sensibilità del solo 28,3%) avranno comunque un risultato negativo. Dunque, a seguito di un esame negativo, la probabilità della diagnosi di endocardite è scesa solo in modo marginale: dal 20% al 15% (equivalente a 143 diviso per 143+789). Tenuto conto della serietà della patologia in questione, una stima del 15% resta tutt'altro che trascurabile e richiederebbe controlli ulteriori e più efficaci.

Per quanto riguarda una seconda ipotesi diagnostica rilevante, quella di osteomielite (un'infezione ossea), la situazione è per molti versi simile. La presenza di questa condizione clinica è stata esclusa in base alle radiografie spinali eseguite sul paziente. Tuttavia, la limitata sensibilità e specificità della radiografia spinale in relazione a questa possibile diagnosi implicano che un risultato negativo permetta di escluderla (facendo scendere la relativa probabilità, poniamo, al di sotto del 3%) solo per individui che hanno fin dall'inizio una probabilità molto ridotta di esserne affetti (si veda Koenker, DeLuca, 1989). Non, quindi, in un caso come quello del signor D., in cui l'improvviso insorgere di persistenti dolori alla schiena e al collo suggeriscono un concreto sospetto clinico di questa patologia.

L'interpretazione dei risultati di un test strumentale è un compito di routine nella medicina clinica. Eppure, come avremo modo di vedere, proprio una di queste due diagnosi rapidamente accantonate (precisamente, quella di osteomielite) si rivelerà infine quella corretta. È quindi il caso di chiedersi come le considerazioni appena riportate possano essere state trascurate. In questo caso, la responsabilità cade sull'*euristica della rappresentatività*.

L'euristica della rappresentatività consiste nel giudicare la probabilità di una condizione *a* alla luce di un'altra condizione *b* in base a quanto *b* è «tipica» o «rappresentativa» di *a*. Immaginate per esempio una paziente di cinquantacinque anni che ha avuto un'embolia polmonare documentata da un'angiografia svolta dieci giorni dopo una colecistectomia. Ritenete più probabile che la paziente in questione accusi (1) emiparesi o (2) dispnea ed emiparesi? La teoria della probabilità ci dice che non è possibile che (2) sia più

probabile di (1), semplicemente perché in qualunque caso in cui si verifichi (2) si verificherà anche (1). Tversky e Kahneman hanno tuttavia osservato sperimentalmente che nove medici su dieci giudicano effettivamente (2) più probabile di (1), cadendo nella cosiddetta «fallacia della congiunzione» (Tversky, Kahneman, 1983). L'emiparesi appare infatti come una condizione molto *a*tipica in una paziente come quella considerata, mentre la congiunzione di dispnea ed emiparesi presenta almeno una caratteristica (dispnea) che in una paziente di questo tipo è relativamente comune. L'euristica della rappresentatività ci ha giocato un brutto scherzo!

Nell'interpretazione del risultato di un test diagnostico si può cadere in un trabocchetto analogo: per esempio si osserva (correttamente) che una ecocardiografia negativa è un risultato tipico in un paziente che *non* è affetto da endocardite (il test è infatti molto specifico) e se ne conclude (erroneamente) che la probabilità che un paziente con un risultato negativo *non* sia affetto da endocardite è (pressappoco) altrettanto alta. In realtà, come illustra il teorema di Bayes, quest'ultima probabilità (nota come *valore predittivo negativo*) dipende non solo dalla specificità del test, ma anche dalla sua sensibilità, nonché dalla probabilità iniziale della diagnosi.

Alcune notevoli conseguenze del fenomeno appena descritto sono esposte nel saggio di Eddy *Il ragionamento probabilistico nella medicina clinica*. L'autore presenta un'analisi accurata del problema dell'interpretazione dei risultati di una mammografia in vista della decisione se eseguire una biopsia, e documenta una serie di errori nel ragionamento probabilistico compatibili con l'impiego dell'euristica della rappresentatività e con la tendenza a trascurare le probabilità iniziali.

L'àncora

Ma le lezioni che si possono trarre dalla storia del signor D. non sono ancora finite. La scarsa considerazione di alcune «note dissonanti» in questo specifico caso clinico può infatti essere in parte dovuta al fatto che, a seguito dei risultati delle emocolture, l'ipotesi che la batteriemia da *staphylococcus aureus* potesse rendere compiutamente conto dei disturbi del paziente ha svolto la funzione di un'«àncora» in una forma di ragionamento euristico nota come *ancoraggio e adattamento*.

Si tratta di un fenomeno psicologico comune e in molti casi efficace. Immaginate per esempio di voler stimare il valore commerciale della vostra auto. Ricordate il prezzo di un'auto dello stesso modello, anch'essa usata ma tenuta un po' meglio. Potete allora prendere questo prezzo come punto di riferimento (un'«àncora», appunto) e «assestarlo» sommariamente verso il basso tenendo conto che la vostra auto è in condizioni leggermente peggiori.

La ricerca sperimentale sulla psicologia del giudizio ha mostrato che in molte situazioni l'«adattamento» che segue all'ancoraggio tende a essere approssimativo e insufficiente (Tversky, Kahneman, 1974). L'impiego di questa euristica in un contesto clinico è stata quindi suggerita come possibile spiegazione dei risultati, alquanto sconcertanti, di un classico studio della associazione pediatrica nazionale americana (si vedano American Child Health Association, 1934; Bakwin, 1945; e gli analoghi risultati ottenuti più recentemente da Ayanian, Berwick, 1991).

A un gruppo di venti pediatri venne chiesto di visitare circa quattrocento bambini di New York dell'età di undici anni che non erano ancora stati sottoposti a tonsillectomia e di indicare per quali avrebbero suggerito l'intervento. L'operazione fu consigliata nel 45% dei casi. Ognuno dei bambini ai quali *non* era stata consigliata la tonsillectomia ricevette quindi una seconda visita pediatrica da un medico diverso da quello che lo aveva visto la prima volta. La tonsillectomia venne ora indicata nel 46% dei casi. Infine, i bambini per i quali l'operazione non era stata ritenuta necessaria in nessuno dei due consulti precedenti, furono visitati indipendentemente una terza volta. Al 44% di loro venne raccomandata la tonsillectomia. A quanto pare, i pediatri coinvolti si aspettavano che circa il cinquanta percento dei bambini di undici anni avessero bisogno di una tonsillectomia. Di fronte a una serie di pazienti che *non* ne avevano bisogno, i medici avranno probabilmente rivisto verso il basso le loro aspettative sulla base dei dati clinici. Ma, a causa di un «adattamento» troppo modesto, hanno finito per suggerire interventi che non sarebbero stati necessari.

L'euristica di ancoraggio e adattamento può quindi indurre a sottovalutare l'impatto di dati in contrasto con un'ipotesi iniziale di riferimento (e in alcuni casi persino a ricostruirli in maniera distorta come informazioni *favorevoli* a quell'ipotesi), trascurando così la necessità di mantenere l'attenzione su ipotesi alternative. Nel caso del signor D., per esempio, non si è sentita l'esigenza di approfondire la presenza inusuale di batteriemia in un paziente anziano senza diabete mellito (si veda ancora Redelmeier, 2005).

Molti errori diagnostici riflettono una tendenza a non tenere conto in modo opportuno di diverse ipotesi in competizione e a trascurare informazioni in grado di discriminarle. Alcuni dei saggi inclusi in questo volume sono specificamente dedicati ai fattori cognitivi alla base di errori di questo tipo. In *Giudizi di probabilità in medicina: trascurare le possibilità implicite*, Redelmeier, Koehler, Liberman e Tversky mostrano come il modo in cui uno stesso insieme di ipotesi alternative viene presentato produca stime sistematicamente discordanti riguardo alla probabilità delle diverse diagnosi e prognosi possibili. In *La raccolta e l'integrazione di informazioni come fonti di errore nelle decisioni diagnostiche*, Gruppen, Wolf e Billi analizzano sperimentalmente la difficoltà da parte dei medici nel valutare in modo appropriato la rilevanza di informazioni potenzialmente utili per la selezione di una diagnosi;

mentre *L'influenza di informazioni pseudodiagnostiche nella valutazione dell'ischemia miocardica* di Green e Yates estende lo studio di questo fenomeno a un importante problema diagnostico in medicina d'urgenza.

Il senno di poi

Tenere conto degli sviluppi di una situazione per valutare se la strada imboccata è quella giusta è un meccanismo fondamentale per l'acquisizione e la crescita della conoscenza, e in ambito clinico è spesso altamente raccomandabile. Tuttavia, anche in questo processo di apprendimento dall'esperienza apparentemente semplice, possono manifestarsi errori e distorsioni sistematiche del giudizio. Il seguito della storia del signor D. ne offre un esempio.

> Dopo quattro giorni di trattamento antibiotico per via endovenosa, i sintomi del signor D. migliorano e il paziente viene dimesso con la diagnosi di batteriemia, secondaria rispetto a escoriazioni dovute alla condizione dermatologica cronica del paziente.
> Tre settimane dopo il signor D. si presenta dal suo medico di base per una visita di controllo. Sebbene i suoi sintomi siano inizialmente migliorati, il suo generale stato di affaticamento e il dolore al collo e alla schiena si sono presentati in modo ricorrente. Inoltre ora avverte formicolii alle dita e ha difficoltà a orinare. Non viene ricoverato, ma rimandato a casa con la prescrizione di una serie di esami di laboratorio.
> Una fra due emocolture risulta positiva per lo *staphylococcus aureus*, e il signor D. è convocato in ospedale per un secondo immediato ricovero. Una risonanza magnetica urgente rivela a questo punto alterazioni compatibili con ascesso epidurale e osteomielite alla sesta e settima vertebra, con interessamento del midollo spinale.

Come si vede, la corretta identificazione della causa dei disturbi del signor D., vale a dire l'osteomielite, è stata notevolmente ritardata dall'osservazione che durante il primo ricovero la terapia antibiotica contro la batteriemia produceva un miglioramento clinico. È infatti probabile che tale miglioramento (rivelatosi poi solo momentaneo) abbia contribuito a ridimensionare il significato di diversi aspetti problematici del percorso diagnostico, suggerendo ai medici la conferma della diagnosi di batteriemia e il congedo ospedaliero del paziente. Con il rischio che la sua osteomielite, erroneamente esclusa, potesse così aggravarsi e dar luogo a complicanze.

Per illustrare la rilevanza della nostra ricostruzione di eventi passati come potenziale fonte di errore, consideriamo due casi di una stessa operazione chirurgica (come un taglio cesareo) nei quali viene effettuato un certo intervento anestesiologico (per esempio, l'anestesia spinale). Assumiamo che i due casi siano fra loro identici (condizioni cliniche della paziente, tipo di in-

tervento praticato, dosi impiegate ecc.), tranne che per un aspetto: in un caso si verifica una complicazione che produce un esito clinico avverso *temporaneo* (per esempio, un arresto cardiaco durante l'operazione, a seguito del quale la madre e il neonato si riprendono pienamente); nell'altro caso, invece, lo stesso tipo di complicazione produce un esito clinico avverso *permanente* (per esempio, l'arresto cardiaco causa la morte della partoriente e danni cerebrali al neonato).

Caplan, Posner e Cheney (1991) hanno costruito, a partire da episodi reali, ventuno coppie di casi simili a quella appena descritta e hanno suddiviso centododici anestesisti in due gruppi. Per ciascuna delle ventuno coppie di casi, uno dei due gruppi, selezionato casualmente, riceveva il caso con esito avverso temporaneo; l'altro gruppo riceveva invece il caso con esito avverso permanente. Il compito dei medici era esprimere una valutazione sul trattamento anestesiologico praticato in ognuno dei ventuno casi che si trovavano di fronte, classificandolo come «appropriato», «non appropriato», o «impossibile da giudicare». In ben quindici delle ventuno coppie di casi, la valutazione è stata prevalentemente positiva («intervento appropriato») se il danno descritto era temporaneo, e prevalentemente negativa («intervento inappropriato») se il danno descritto era permanente – a dispetto del fatto che, come abbiamo detto, i due casi di ogni coppia erano identici tranne che per l'estensione del danno procurato. Un esito particolarmente avverso rende quindi il giudizio con «il senno di poi» molto più severo del giudizio basato sugli *stessi elementi* qualora l'esito dell'intervento si sia rivelato più favorevole, distorcendo significativamente la valutazione dell'appropriatezza di una specifica decisione clinica.

Questa «distorsione retrospettiva» non tiene conto del fatto che, in condizioni di incertezza, è possibile che una situazione si sviluppi in un modo o in un altro *a dispetto* delle nostre capacità, o indipendentemente da esse, e non *a causa* di esse (Baron, Hershey, 1988a). Per contro, gli esseri umani sembrano avere una speciale attitudine a dare senso a un corso di eventi noto, ricostruendolo come uno sviluppo prevedibile in anticipo, coerente e persino inevitabile. Questa tendenza, nota agli psicologi cognitivi come *hindsight bias*, può produrre sistematici errori di valutazione, come ha messo originariamente in luce Fischhoff (1975) in una serie di ingegnosi lavori sperimentali. La rilevanza di questo fenomeno in ambito diagnostico è illustrata nel saggio *La distorsione retrospettiva nella valutazione della probabilità delle diagnosi* di Arkes, Saville, Wortmann e Harkness, che mostra sperimentalmente come il solo fatto di sapere a posteriori qual era la diagnosi corretta induca i medici a ritenerla sistematicamente più probabile alla luce dei dati clinici che erano disponibili fin dal principio.

Fra le notevoli conseguenze della distorsione retrospettiva del giudizio sono da segnalare quelle di tipo pedagogico. Si prenda il caso delle conferenze

clinicopatologiche, considerate in molte istituzioni un utile strumento per la formazione medica. In una conferenza clinicopatologica, si chiede a un medico di formulare una diagnosi riguardo a un certo paziente reale (anonimo) sulla base della storia clinica e dei risultati di alcuni esami preliminari. In seguito, si rivela all'uditorio la diagnosi corretta e la si discute. L'idea è che l'esposizione a questo tipo di esercizio «per casi concreti» possa incoraggiare una valutazione critica delle proprie capacità diagnostiche, e sollecitarne il miglioramento. La presenza di una sistematica distorsione retrospettiva del giudizio solleva perlomeno dei dubbi sull'efficacia di questo approccio. Infatti, quando il caso viene presentato, un partecipante potrebbe stimare la probabilità della diagnosi che si rivelerà corretta in una certa misura. Ma, se interrogato dopo che la soluzione è stata svelata, spesso egli sarà convinto che, dovendo affrontare un caso come quello discusso, avrebbe certamente associato a quella diagnosi una stima di probabilità più alta. In questo modo egli *si ingannerà* in modo sistematico sulla reale qualità del suo giudizio clinico.

Dawson *et al*. (1988) hanno documentato questo fenomeno in uno studio che coinvolgeva centosessanta partecipanti di conferenze clinicopatologiche presso il Cleveland Metropolitan General Hospital. Una parte di loro valutava la probabilità delle principali alternative di diagnosi *prima* che la diagnosi corretta venisse comunicata: solo nel 30% dei casi la diagnosi corretta era classificata come la più probabile. Un secondo gruppo di soggetti svolgeva lo stesso compito *dopo* che la diagnosi corretta era stata comunicata: questa volta il 50% era pronto a dichiarare che quella era la più probabile alla luce del quadro clinico inizialmente a disposizione. Si osservi che un giovane praticante che lavori a contatto con colleghi più esperti e ufficialmente responsabili si troverà spesso in una situazione analoga: avrà cioè accesso ai dati clinici, alle diagnosi e alle decisioni dei suoi colleghi e, soprattutto, agli esiti dei singoli casi. In queste condizioni, sarà probabilmente indotto a sovrastimare l'accuratezza del proprio giudizio diagnostico e le sue capacità decisionali, sottovalutando sensibilmente la differenza fra elaborare correttamente una diagnosi o prendere una decisione appropriata e rifletterci su una volta che la storia ha fatto il suo corso. In breve, in alcuni casi uno sguardo retrospettivo sugli eventi può incidere negativamente sulle nostre valutazioni, riducendo l'affidabilità del giudizio clinico basato sull'osservazione, l'esperienza e l'esercizio.

Decisioni (non) consequenziali

Prendendo ancora spunto dal caso del signor D., vedremo ora come la rilevanza dei fattori cognitivi si estenda al di là della psicologia del ragionamento diagnostico coinvolgendo il problema delle scelte cliniche di intervento.

Si è detto che durante il primo ricovero del signor D. era stata eseguita un'ecografia transtoracica (risultata negativa) per controllare l'ipotesi diagnostica di endocardite. Durante il secondo ricovero del paziente viene effettuata una *seconda* ecografia transtoracica, che risulta ancora una volta negativa. A seguito di ciò, i medici ritengono comunque utile un ulteriore approfondimento, raccomandando l'esecuzione di un esame più invasivo ma più affidabile per la stessa patologia, vale a dire l'ecografia transesofagea (si veda Daniel *et al.*, 1991).

È lecito assumere che tale ulteriore approfondimento sarebbe stato considerato altrettanto opportuno, e a maggior ragione, alla luce di un risultato *positivo* della seconda ecografia trastoracica, che avrebbe innalzato di molto il rischio di endocardite. Sembra pertanto che *qualunque* fosse il risultato della seconda ecocardiografia transtoracica, sarebbe comunque stato raccomandato un ulteriore esame più invasivo.

Si tratta di un caso analogo a quello in cui si ordina una mammografia per un nodulo sospetto al seno anche se poi, *qualunque sia il risultato*, si procederà comunque prescrivendo una biopsia: per escludere definitivamente la diagnosi di cancro (nel caso la mammografia sia risultata negativa); oppure per confermare definitivamente quella stessa diagnosi (nel caso la mammografia sia risultata positiva). Ora, l'utilità di un esame strumentale risiede nel fatto che i suoi risultati possono influenzare il successivo trattamento discriminando fra corsi d'azione alternativi. Se ciò non accade, l'esecuzione dell'esame in questione dà luogo a un inutile dispendio di risorse e, in alcuni casi, può impedire l'acquisizione tempestiva di informazioni maggiormente rilevanti (in proposito, si veda ancora il saggio di Eddy in questo volume).

Questo tipo di scelte «non consequenziali» ha precise radici cognitive, come mostra un semplice esperimento.

Immaginate di essere uno studente e di aver appena sostenuto un esame molto impegnativo. Se foste stati promossi, comprereste un biglietto in offerta per una vacanza in un posto rinomato? E se invece foste stati respinti? In due gruppi di studenti universitari Shafir e Tversky (1992a) hanno osservato che la percentuale di risposte positive nei due casi era pressoché identica (leggermente superiore al 50%). Ciò significa che l'esito dell'esame non influenzava la decisione in un senso o nell'altro. Eppure la maggioranza di un terzo gruppo di studenti, ai quali *non* veniva detto se erano stati promossi o respinti, avrebbe preferito subire una piccola penale pur di rinviare la scelta dell'acquisto del biglietto a dopo la comunicazione dei risultati. Pagando così di tasca propria pur di avere una «buona ragione» per partire: per esempio, il bisogno di riposo in vista della necessaria ripetizione dell'esame nel caso fossero stati respinti; oppure il legittimo desiderio di «premiarsi» nel caso fossero stati promossi.

Scelte conflittuali

L'esigenza psicologica di fondare le proprie scelte su «buone» ragioni può produrre risultati indesiderabili in diverse situazioni. Immaginate, per esempio, che siano stati programmati alcuni interventi di endarterectomia in pazienti affetti da una forma seria e asintomatica di ostruzione (stenosi) dell'arteria carotidea. La disponibilità temporaneamente ridotta di sale operatorie rende necessaria la decisione di operare per primo uno solo fra due pazienti: uno (paziente 1) di settantadue anni, senza problemi clinici concomitanti ma con una stenosi molto avanzata (del 90%); l'altro (paziente 2) di cinquantadue anni, con una stenosi avanzata (70%) e un passato di alcolismo. Immaginate ora una situazione del tutto identica, con una sola eccezione: questa volta i pazienti fra cui scegliere sono *tre* – i due precedenti con le caratteristiche indicate più il paziente 3, in condizioni paragonabili a quelle del paziente 2: cinquantacinque anni, stenosi avanzata (70%), quarant'anni di fumo regolare di sigarette.

La logica suggerisce che un medico che nel primo caso di scelta di fronte a due sole alternative ritenga opportuno escludere il paziente più anziano (paziente 1), dando la priorità a un paziente più giovane (paziente 2), sarà dello stesso avviso anche nel secondo caso in cui semplicemente si è aggiunto un altro paziente con caratteristiche simili al paziente 2. Questa aspettativa, perfettamente ragionevole, ha la sua controparte formale in un fondamentale principio della teoria della scelta razionale noto come «principio di regolarità». Il principio di regolarità implica che l'ordine di preferenza tra due opzioni A e B non deve mutare con l'aggiunta di un'ulteriore opzione C (Tversky, Simonson, 1993). Per contro, si è osservato che uno specialista su cinque sceglierebbe di operare un paziente più giovane (paziente 2) quando la scelta è fra il paziente 1 e il paziente 2, ma quello più anziano (paziente 1) quando la scelta è fra il paziente 1, il paziente 2 e il paziente 3 – in palese contraddizione con il principio di regolarità (Redelmeier, Shafir, 1995).

Come abbiamo anticipato, risultati come questo trovano una possibile spiegazione nei termini delle ragioni a cui gli individui ricorrono per giustificare a loro stessi o agli altri una data decisione (Shafir, Simonson, Tversky, 1993).

Una situazione decisionale può essere più o meno conflittuale. In una situazione conflittuale, il fatto che un'opzione sia attraente può non essere sufficiente perché essa venga scelta. È altrettanto importante che ci siano «buone» ragioni per considerarla superiore alle opzioni con le quali va confrontata. Nel problema dell'endarterectomia il paziente 3 ha molti tratti in comune con il paziente 2, e può essere difficile trovare «buone» ragioni per dare la priorità a uno di loro rispetto all'altro. Il conflitto decisionale può così essere «risolto» orientandosi sul paziente più anziano, le cui caratteristiche si diffe-

renziano più chiaramente. Questo meccanismo psicologico può quindi mettere in secondo piano fattori clinicamente rilevanti (come l'aspettativa di vita) che in altre condizioni avrebbero indotto il medico, senza indugio, a operare per primo un paziente più giovane.

Un altro importante esempio di violazione del principio di regolarità (noto come «effetto di attrazione») è analizzato in questo volume da Schwartz e Chapman nel saggio *È sempre preferibile avere più opzioni?*, incentrato sulle scelte dei medici riguardo a diverse alternative di trattamento farmacologico. Immaginate un medico che affronti la scelta conflittuale fra due farmaci A e B per il trattamento di una data patologia. L'esperimento di Schwartz e Chapman mostra che aggiungere all'insieme delle due opzioni disponibili un terzo farmaco C, palesemente inferiore ad A (per esempio perché più rischioso, ma non più efficace) farà crescere sensibilmente la probabilità che A venga scelta. Gli autori discutono nei dettagli come il mercato dei farmaci possa presentare ai medici problemi decisionali di questo tipo, influenzando sistematicamente le loro scelte.

Fallacia dell'omissione

Abbiamo lasciato il signor D. alle prese con la prescrizione di un esame alquanto invasivo, l'ecocardiografia transtoracica, volto a escludere la possibilità di un'endocardite. La nostra storia ci spinge ora ad analizzare il tema del *consenso informato* e a considerare come la partecipazione dei pazienti alle decisioni mediche relative alla propria salute assegni un ruolo centrale alle loro preferenze nell'impiego degli strumenti diagnostici e delle cure disponibili. Nel nostro caso, il signor D. rifiuta di sottoporsi all'ecocardiografia transesofagea che gli viene raccomandata, e i medici non insistono in considerazione del fatto che il paziente riceverà comunque una prolungata terapia antibiotica per la sua osteomielite.

Di fronte al rifiuto di un trattamento medico consigliato è lecito chiedersi se, e a quali condizioni, l'autonomia del paziente sia in accordo con la qualità delle cure; vale a dire se il riconoscimento dell'autonomia decisionale promuova di fatto le scelte migliori. Anche in questo ambito è necessario tenere conto di alcuni aspetti cognitivi che possono indurre gli individui ad accettare o a rifiutare gli interventi medici a seconda del modo in cui viene presentato loro il problema di scelta e quindi influenzata la percezione soggettiva del rischio associato a quegli interventi.

Si prenda un caso che è stato studiato nei dettagli: quello della vaccinazione. Supponete che un certo vaccino comporti un rischio di morte di 10 casi su 10.000. Presumibilmente, nessuno sarebbe disposto a utilizzare tale vaccino se la condizione che con esso si intende prevenire (per esempio, un certo tipo di influenza) comporta un rischio di morte *inferiore* a 10 su 10.000.

Semplicemente, i potenziali costi sono superiori ai benefici attesi. Ma ora immaginate che la situazione sia invertita: una certa forma influenzale uccide in media 10 persone su 10.000 ed è stato sviluppato un vaccino il cui impiego comporta esso stesso un rischio di morte, ma di entità inferiore. Per le stesse ragioni invocate nel caso precedente, in questa situazione il vaccino dovrebbe essere universalmente considerato vantaggioso. Ovvio, no?

Eppure molte persone trattano i due casi in modo diverso. Ritov e Baron (1990) hanno osservato che la maggior parte degli individui (circa sei su dieci) *non* accetterebbe di vaccinare un figlio di un anno contro una forma influenzale che comporta un rischio di morte di 10 su 10.000, anche se gli effetti collaterali potenzialmente fatali del vaccino fossero inferiori a questa cifra. Addirittura, circa una persona su quattro sarebbe disposta a usare il vaccino *solo* se il rischio di effetti collaterali potenzialmente fatali fosse uguale a 0 (si veda Asch *et al.*, 1994, per uno studio analogo relativo a un vaccino reale contro la pertosse).

Questo risultato illustra un fenomeno noto come «fallacia dell'omissione» (*omission bias*): la tendenza a voler evitare i rischi associati a un intervento attivo, anche se questo comporta l'esposizione a rischi ancora maggiori dovuti all'omissione dell'intervento. È da segnalare che la rilevanza cruciale di un diverso atteggiamento riguardo all'azione e all'inazione può estendersi a questioni particolarmente delicate ai confini fra medicina ed etica, come la morte medicalmente assistita (Baron, 1998).

Effetto di incorniciamento

Gli ultimi sviluppi del caso del signor D. ci hanno indotto a guardare all'«altra estremità dello stetoscopio», vale a dire al modo in cui le valutazioni dei medici interagiscono con le esigenze e le preferenze dei pazienti. La conclusione della storia ci fornirà a questo proposito alcuni ulteriori spunti.

> La diagnosi confermata di osteomielite suggerisce l'opportunità di un consulto neurochirurgico, dal quale emerge la raccomandazione, condivisa dal medico di base, di un intervento di tipo chirurgico. Il paziente esprime però la sua preferenza per un trattamento non chirurgico. Vengono così prescritti steroidi per la compressione del midollo spinale e un ciclo di sei settimane di cefazolina per via endovenosa.
> Alla fine di una serie di successivi controlli, il signor D. riporta ancora parestesia e una leggera debolezza delle braccia, ma nessun altro disturbo. Tre successive emocolture risultano negative.

Perché un processo decisionale condiviso fra medico e paziente dia luogo a scelte di trattamento appropriate è necessario che siano soddisfatte alcune

condizioni. Innanzi tutto, le valutazioni proposte dal medico devono rispecchiare la considerazione degli aspetti clinici pertinenti nel caso in esame. In secondo luogo, deve essere possibile identificare in modo affidabile le preferenze del paziente rilevanti per la decisione, ovvero il valore relativo che egli attribuisce agli esiti dei diversi possibili corsi d'azione. Alcuni celebri esempi tratti dalla ricerca sulla psicologia della decisione in condizioni di rischio e di incertezza suggeriscono però che il verificarsi di queste condizioni non è affatto ovvio. Non sempre le preferenze degli individui sono preordinate, prestabilite e facili da individuare; spesso sono costruite nel processo stesso della scelta, sono influenzate dal contesto e risentono in modo sproporzionato del modo in cui le alternative vengono presentate (Slovic, 1995).

Considerate i problemi seguenti.

Problema 1

L'Italia si sta preparando ad affrontare un'insolita malattia asiatica a causa della quale ci si aspetta che debbano morire 600 persone. Per combatterla sono stati proposti due programmi alternativi: *X* e *Y*.

Se verrà adottato il programma *X*, verranno salvate 200 persone. [72%]

Se verrà adottato il programma *Y*, c'è 1/3 di probabilità che si salvino 600 persone e 2/3 di probabilità che nessuno si salvi. [28%]

Quale programma sceglieresti?

Problema 2

L'Italia si sta preparando ad affrontare un'insolita malattia asiatica a causa della quale ci si aspetta che debbano morire 600 persone. Per combatterla sono stati proposti due programmi alternativi: *W* e *Z*.

Se verrà adottato il programma *W*, 400 persone moriranno. [22%]

Se verrà adottato il programma *Z*, c'è 1/3 di probabilità che nessuno muoia e 2/3 di probabilità che muoiano 600 persone. [78%]

Quale programma sceglieresti?

La maggior parte delle persone sceglierebbe *X* nel problema 1 e *Z* nel problema 2 – come si può constatare dalle percentuali effettivamente registrate (riportate tra parentesi quadre) in un noto esperimento di Tversky e Kahneman (1981). Questa coppia di scelte (*X* e *Z*) rappresenta una flagrante incoerenza. Il motivo è che il programma *X* è identico a *W*, e *Y* è identico a *Z*. *X* e *W* garantiscono con certezza uno stesso esito: rispettivamente, 200 vite salvate su 600 e 400 perse su 600. *Y* e *Z* offrono una stessa combinazione di possibili esiti: rispettivamente, 1/3 di probabilità di salvare 600 vite e 2/3 di probabilità di perdere 600 vite. Nei due problemi le scelte degli individui dovrebbero quindi distribuirsi in ugual misura su programmi fra loro identici. Quin-

di le opzioni *X* e *W* dovrebbero (contrariamente a quanto effettivamente accade) essere scelte con ugual frequenza, e lo stesso dovrebbe valere per *Y* e *Z*.

Il problema 1 e il problema 2 descrivono una stessa situazione ma la «incorniciano» in modo differente: nel primo caso in termini positivi di vite *salvate*, nel secondo in termini negativi di vite *perse*. L'inversione delle preferenze dal problema 1 (nel quale il 72% dei soggetti sceglie il programma *X*/*W*) al problema 2 (nel quale il 78% dei soggetti sceglie invece il programma *Y*/*Z*) è noto come *framing effect* o «effetto di incorniciamento». Tversky e Kahneman hanno individuato alcuni specifici fattori psicologici che possono produrre l'effetto di incorniciamento e la conseguente violazione di un fondamentale principio di razionalità detto di «invarianza descrittiva», secondo il quale uno stesso problema decisionale dovrebbe essere trattato nello stesso modo indipendentemente da come i termini della situazione vengono descritti (Arrow, 1982).

Tversky e Kahneman hanno suggerito che le persone percepiscono i risultati di una scelta in termini di *cambiamenti* relativi a un (non costante) livello di riferimento (e quindi in termini di *guadagni* o di *perdite*), piuttosto che in termini di stati assoluti (di ricchezza, di salute, o altro). Una importante differenza fra guadagni e perdite riguarda l'attitudine degli individui verso il rischio, come illustrato dai due seguenti problemi di scelta fra coppie di scommesse.

Scommesse 1

X: 90% di probabilità di vincere 3.000 dollari,
10% di probabilità di non vincere nulla [86%]

Y: 45% di probabilità di vincere 6.000 dollari,
55% di probabilità di non vincere nulla [14%]

Scommesse 2

W: 90% di probabilità di perdere 3.000 dollari,
10% di probabilità di non perdere nulla [8%]

Z: 45% di probabilità di perdere 6.000 dollari,
55% di probabilità di non perdere nulla [92%]

Come si vede dalle percentuali sperimentalmente osservate (riportate tra parentesi quadre), nel primo problema la maggior parte delle persone preferisce l'opzione con un esito positivo relativamente più sicuro (*X*), rinunciando a una scommessa più rischiosa che potrebbe dare una vincita doppia (*Y*). Nel secondo problema, viceversa, pur di evitare una perdita limitata ma molto probabile (*W*) la maggior parte delle persone è disposta a affrontare il rischio di subirne una doppia (*Z*). In breve, le persone si mostrano tendenzialmente avverse al rischio nell'ambito delle vincite, ma più propense al rischio nell'ambito delle perdite (Kahneman, Tversky, 1979).

Un secondo importante fenomeno evidenziato da Tversky e Kahneman è noto come «effetto certezza»: la stessa riduzione o lo stesso aumento della probabilità di un determinato esito ha un impatto maggiore quando implica una differenza fra certezza e incertezza che non quando comporta una semplice differenza fra diversi gradi di incertezza. Per esempio la maggior parte delle persone, immaginando di dover giocare alla roulette russa, sarebbe disposta a pagare di più per togliere dalla pistola un solo e unico proiettile che per toglierne uno su quattro. In entrambi i casi la probabilità di morte è ridotta della stessa quantità oggettiva (1/6) ma psicologicamente l'impatto da 1/6 a zero è differente da quello da 4/6 a 3/6 (Zeckhauser, 1991). In accordo con questo principio, si è osservato che le persone sono disposte a pagare una cifra significativamente più alta per una riduzione da 5 a 0 su 10.000 del rischio di inalare una sostanza nociva che per una riduzione da 15 a 5 su 10.000 (Viscusi, Magat, Huber, 1987).

La diversa attitudine al rischio fra guadagni e perdite e l'effetto certezza, incorporati da Tversky e Kahneman in una sofisticata teoria *descrittiva* della decisione detta *prospect theory* (Kahneman e Tversky, 1979, 1992), riflettono principi psicologici robusti che valgono anche per le decisioni mediche. Per esempio, sottoponendo a diverse centinaia di pazienti scelte fra coppie di ipotetiche terapie farmacologiche (per curare forti emicranie o dolori al petto) si è potuto osservare che, quando gli esiti descritti erano i *benefici* terapeutici dei trattamenti (come una riduzione del dolore o un aumento dell'aspettativa di vita), la maggior parte dei pazienti preferiva un effetto sicuro ma moderato alla possibilità che si verificasse un effetto molto favorevole oppure nessun effetto. Per contro, quando gli esiti da valutare erano gli effetti collaterali *dannosi* dei diversi trattamenti (per esempio, nausea o dispnea), piuttosto che accettare un effetto negativo sicuro ma moderato, i pazienti erano disposti a rischiarne uno più grave, pur di avere maggiori possibilità di non subire alcun effetto collaterale (Eraker, Sox, 1981) (è da notare che in questo studio i pazienti che soffrivano *effettivamente* dei disturbi descritti non hanno risposto in modo significativamente differente dagli altri).

In cerca delle preferenze

La tendenza a trattare in modo sistematicamente differente presentazioni distinte di uno stesso problema è uno dei fenomeni più significativi documentati dalla ricerca cognitiva, analizzato nei dettagli da diversi studi compresi nella seconda parte di questo volume.

Il saggio *L'elicitazione delle preferenze per terapie alternative* di McNeil, Pauker, Sox e Tversky illustra un'importante applicazione medica di questo fenomeno in riferimento alla scelta fra due diversi trattamenti in un caso ipo-

tetico di cancro al polmone: l'intervento chirurgico oppure la radioterapia. Una presentazione in termini di «guadagni» (probabilità di sopravvivenza) e una presentazione in termini di «perdite» (probabilità di morte) degli stessi dati relativi agli effetti dei due trattamenti inducono medici e pazienti a esprimere preferenze significativamente diverse.

È da notare che gran parte delle scelte mediche possono essere descritte nei termini dei loro potenziali benefici o dei loro potenziali costi: nel discutere lo svolgimento di mammografie periodiche si potrebbero, per esempio, sottolineare i possibili danni che conseguono dalla loro mancata esecuzione («non eseguendo regolarmente una mammografia si perde l'occasione di avvalersi del miglior metodo disponibile per individuare tempestivamente un tumore maligno al seno») o i possibili benefici associati allo loro esecuzione («eseguendo regolarmente una mammografia si approfitta del miglior metodo disponibile per individuare tempestivamente un tumore maligno al seno»). Il fatto che di fronte a presentazioni differenti di uno stesso problema le scelte degli individui possano divergere in modo sistematico solleva un problema immediatamente rilevante per le decisioni cliniche: quali sono le scelte che manifestano le «vere» preferenze dei medici e dei pazienti, quelle ottenute presentando il problema in termini di benefici o quelle ottenute presentando lo stesso problema in termini di perdite?

Un altro caso in cui incorniciamenti differenti possono indurre decisioni diverse riguarda alcuni strumenti statistici convenzionalmente impiegati per presentare e comunicare i rischi medici: il rischio *relativo* e il rischio *assoluto*. Supponete per esempio di sapere che su 100 uomini di cinquant'anni con livelli normali di colesterolo 4, in media, avranno un infarto nei prossimi dieci anni, mentre su 100 uomini della stessa età con livelli alti di colesterolo questo numero sale a 6. In questa popolazione un provvedimento in grado di tenere efficacemente sotto controllo i livelli di colesterolo (come un'opportuna dieta alimentare) produrrà una diminuzione del rischio *relativo* di infarto del 33% in dieci anni (evitandolo a 2 su 6 individui che altrimenti l'avrebbero avuto, cioè appunto al 33%) e una diminuzione del rischio *assoluto* del 2% in dieci anni (equivalente alla semplice differenza algebrica fra 6 e 4). I dati di *una stessa* ricerca clinica possono essere presentati in ciascuno di questi formati, ma dal punto di vista psicologico la scelta del formato da utilizzare non è affatto indifferente. Numerosi studi suggeriscono infatti che gli individui nelle loro scelte sono più sensibili alla descrizione di un effetto notevole in un ambito circoscritto che non a quella di un effetto comparabilmente modesto in relazione a un ambito più ampio, anche se i risultati nei due casi sono oggettivamente gli stessi (Fetherstonhaugh *et al*., 1997; Slovic, Fischhoff, Lichtenstein, 1982; Slovic *et al*., 2002; Thaler, 1999). In accordo con questa osservazione, i lavori di Forrow, Taylor e Arnold (*Assolutamente relativo: come la presentazione dei risultati della ricerca medica influenza le scelte di tratta-*

mento) e Fahey, Griffiths e Peters (*Finanziamenti basati sull'evidenza: la comprensione dei risultati della ricerca clinica*) mostrano come medici e responsabili di politiche sanitarie siano significativamente più inclini a promuovere un particolare trattamento quando i suoi benefici vengono loro descritti in termini di rischio relativo piuttosto che in termini di rischio assoluto.

Infine le decisioni mediche possono variare in modo sistematico, a parità di informazioni, per effetto del modo in cui tali informazioni vengono ottenute. In *L'ingannevole ricerca di più informazioni* Redelmeier, Shafir e Ajula presentano una serie di sofisticati esperimenti in cui i soggetti coinvolti (infermieri interpellati nella veste di potenziali donatori, e medici specialisti) manifestano scelte discrepanti di fronte a problemi identici, a seconda che un particolare elemento di informazione sia stato ricercato e acquisito nel tempo o fosse noto fin dall'inizio. Il saggio include un'attenta discussione delle conseguenze di questo fenomeno in relazione al delicato rapporto fra indagine clinica e scelte di intervento.

Si noti che risultati come quelli appena visti non mettono di per sé in dubbio la necessità di fondare le decisioni cliniche sulle valutazioni degli esperti e sulle preferenze dei pazienti. Essi piuttosto suggeriscono che la robustezza e la coerenza delle preferenze di medici e pazienti non dovrebbero essere date per scontate e dovrebbero essere valutate tenendo conto dei processi cognitivi che presiedono a ogni decisione.

La forza della rappresentazione

La presenza di errori e incoerenze nel giudizio e nella decisione mette in discussione la validità dei tradizionali principi formali del ragionamento in quanto modelli *descrittivi* del comportamento umano, ma non ne riduce la forza *normativa*, vale a dire la capacità di indicare le soluzioni corrette a cui legittimamente aspiriamo nel tentativo di giudicare e scegliere in modo ottimale. È però soltanto l'interazione fra la conoscenza dei principi normativi e la consapevolezza dei limiti cognitivi che stanno alla base del funzionamento «automatico» del nostro modo di giudicare e di scegliere a indicarci le cause, per così dire, strutturali dei nostri errori, aumentando la probabilità di evitarli e di prendere decisioni migliori.

Per renderci conto di come ciò sia possibile, consideriamo lo scenario seguente, che ricalca un problema clinico comune.

> La probabilità che una persona con più di cinquant'anni senza sintomi abbia un cancro colorettale è dello 0,3%. Se una persona ha un cancro colorettale, c'è una probabilità del 50% che abbia una copremia positiva; se non ha un cancro colorettale, c'è una probabilità del 3% che abbia comunque una copre-

mia positiva. Immaginate una persona sopra i cinquant'anni, asintomatica, sottoposta a *screening* e con una copremia positiva. Qual è la probabilità che abbia veramente un cancro colorettale?

Come ormai sappiamo, il teorema di Bayes permette di ottenere in modo rigoroso la soluzione corretta (4,7%), mentre la mente umana è spesso confusa da questo tipo di problemi. A riprova di ciò, Hoffrage e Gigerenzer hanno presentato il problema della copremia a ventiquattro medici tedeschi a tutti i livelli di esperienza clinica: solo uno di loro ha risposto alla domanda ragionando conformemente al teorema di Bayes, e la maggior parte ha fornito stime superiori al 40% (Hoffrage, Gigerenzer, 1998). Dall'analisi delle annotazioni dei medici e da colloqui successivi all'esperimento è stato possibile constatare che la strategia cognitiva più diffusa fra i partecipanti consisteva nello stimare il valore predittivo positivo del test sulla base della sua sensibilità (che è del 50%). Ma fornendo gli stessi dati *in una forma diversa* il ricorso a questa strategia (e gli errori che ne conseguono) si riduceva drasticamente, lasciando il posto a molte più risposte accurate.

Su 10.000 persone con più di cinquant'anni senza sintomi, 30 hanno un cancro colorettale. Fra queste 30 persone con un cancro colorettale, 15 avranno una copremia positiva. Fra i 9.970 restanti, senza cancro colorettale, 300 avranno ciononostante una copremia positiva. Immaginate un campione di individui sopra i cinquant'anni, asintomatici, sottoposti a *screening* e con una copremia positiva. Quanti di loro avranno avranno veramente un cancro colorettale? ___ su ___

Qui la risposta corretta è a portata di mano: i veri positivi sono 15 su un totale di 315 esami positivi (per una probabilità di 15 diviso 315, uguale appunto al 4,7%). Questa volta, ben due medici su tre (il 67%) hanno ragionato correttamente. Questa formulazione del problema (che fa uso di «frequenze naturali» anziché di valori di probabilità o di percentuali) induce a tenere in considerazione la probabilità iniziale, perché fin dall'inizio ripartisce una popolazione campione (10.000) nel piccolo gruppo di persone che ha il cancro (30) e nella grande maggioranza che non ce l'ha (9.970), rendendo così possibile la soluzione del problema in accordo con le regole del calcolo delle probabilità e in pochi semplici passaggi.

La decisione di eseguire test di controllo (come mammografia, test del PSA per il cancro alla prostata, TAC spirale per il tumore al polmone, test genetici ecc.) su individui senza sintomi né specifici fattori di rischio è spesso più delicata di quanto possa sembrare a prima vista a causa dell'eventualità di un risultato falsamente positivo e delle sue possibili conseguenze. Il saggio *Le consulenze sull'AIDS per persone a basso rischio* di Gigerenzer, Hoffrage e Ebert, con cui si apre la terza parte di questo volume, presenta una ricerca sul

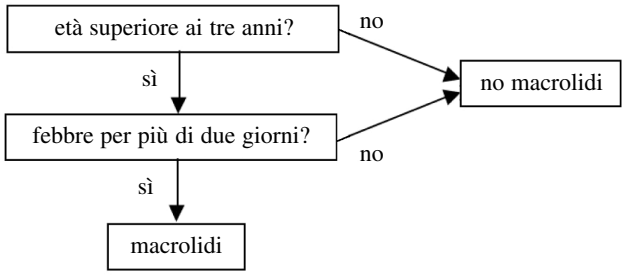
campo sul modo in cui i consulenti di venti istituzioni sanitarie tedesche (per la maggior parte medici) elaborano le informazioni relative al test dell'HIV e le comunicano a utenti con un basso rischio di infezione. Dai risultati dello studio emerge in modo inequivocabile quanto sia urgente l'introduzione di metodi efficaci per agevolare il ragionamento statistico nella comprensione dei problemi medici e nella comunicazione dei rischi ai pazienti.

Procedure efficienti

L'interpretazione di un test diagnostico non è l'unico compito in cui è possibile mettere alla prova le potenzialità di semplici procedure correttive per il miglioramento del giudizio e della decisione medica. Si consideri il problema seguente. Un bambino piccolo ha contratto la polmonite. Dovrebbe ricevere una terapia antibiotica che include macrolidi? A causa di possibili resistenze degli agenti patogeni coinvolti, l'efficacia di questo tipo di trattamento non è garantita, mentre in soggetti così fragili è necessario tenere ben presente il rischio di possibili effetti collaterali. In breve, si tratta di una decisione clinica alquanto delicata, che per di più deve essere presa tempestivamente.

L'«albero decisionale» in fig. 1 determina la risposta a questo specifico problema illustrando una regola estremamente semplice: *prescrivi macrolidi se e solo se l'età è maggiore di tre anni e si è avuta febbre per più di due giorni.*

Fig. 1 - Una procedura semplice e veloce per decidere se impiegare macrolidi nel trattamento di bambini piccoli affetti da polmonite (si veda Fischer et al.*, 2002)*



Ciò che colpisce è la facilità con la quale una procedura del genere può essere memorizzata e applicata. Tale facilità solleva il dubbio che si tratti di un metodo molto approssimativo e poco accurato. Sono pertanto notevoli i risultati ottenuti da Fischer *et al*. (2002) che hanno controllato questa semplice re-

gola decisionale in più di 250 casi reali osservando che, in confronto con le pratiche accettate, essa dava luogo a una riduzione del 68% del rischio relativo di una prescrizione inappropriata. In questo studio, tale margine di miglioramento è risultato paragonabile a quello che si sarebbe ottenuto basando ogni singola decisione clinica su un modello ben più sofisticato e complesso ricavato attraverso il metodo statistico della regressione logistica. Questo risultato si inscrive in un ampio programma di ricerca intrapreso da Gigerenzer e collaboratori (Gigerenzer, Todd, ABC Group, 1999) sulle potenzialità teoriche e pratiche di procedure decisionali semplici ed economiche (*fast and frugal heuristics*) nello studio della cognizione umana. Le applicazioni mediche già documentate comprendono, oltre a quella appena discussa, il trattamento di pazienti colpiti da patologie cardiache acute nei reparti di emergenza (Bremen *et al.*, 1993; Gigerenzer, 1996; Green, Mehr, 1997).

È da notare che la procedura illustrata in fig. 1 può anche essere rappresentata come un caso particolarmente semplice di «modello lineare», vale a dire un'equazione nella quale un aspetto che si intende valutare è espresso attraverso una funzione lineare di una o più variabili:

$$y = x_1 + x_2$$

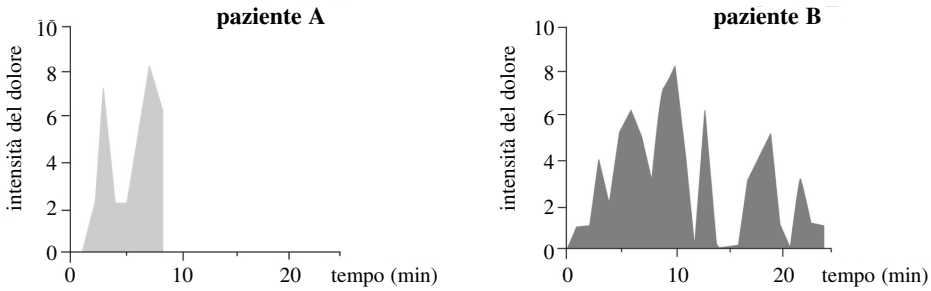
Si tratta di assumere che: x_1 abbia valore 1 se il paziente ha avuto febbre per più di due giorni e valore 0 altrimenti; x_2 abbia valore 1 se il paziente ha più di tre anni e valore 0 altrimenti. A questo punto, se $y = 2$, allora si prescrive la terapia antibiotica; se $y < 2$, non la si prescrive.

Nel saggio *Giudizio clinico* vs. *giudizio statistico*, compreso in questo volume, Dawes, Faust e Meehl presentano un'ampia analisi, basata sui risultati di numerosi studi, del rapporto fra il giudizio clinico intuitivo e quello fondato su procedure standardizzate di questo tipo (quali appunto i modelli lineari) a parità di informazioni disponibili. Le indagini su questo tema, inizialmente dedicate soprattutto alla psicologia clinica e alla psichiatria, si sono estese in seguito a numerose altre discipline, mettendo in evidenza in molti ambiti la superiorità delle previsioni ricavate attraverso opportune procedure standardizzate rispetto al giudizio intuitivo degli esperti. In medicina, studi ormai classici riguardano l'accuratezza delle prognosi in casi di cancro (Einhorn, 1972) o in pazienti con patologie coronariche (Lee *et al.*, 1986) e problemi diagnostici come la discriminazione fra itterizia intraepatica ed extraepatica (Boom *et al.*, 1988). L'importante saggio di Dawes, Faust e Meehl analizza nei dettagli il significato, le cause e le implicazioni di questi risultati (per una ricostruzione dei rapporti fra il programma di ricerca di Gigerenzer e collaboratori e quello che ispira il lavoro di Dawes, Faust e Meehl, si veda Katsikopoulos *et al.*, 2005).

Alcuni dei più importanti contributi costruttivi della psicologia cognitiva nell'ambito della decisione medica riguardano direttamente le preferenze e i comportamenti dei pazienti. In un'originale indagine sul campo, Redelmeier e Kahneman si sono posti una domanda che da questo punto di vista risulta immediatamente rilevante: quali sono i fattori che influenzano il ricordo del disagio di un paziente a seguito di un esame medico spiacevole, con possibili conseguenze sulla sua disponibilità verso successivi interventi?

Per stabilirlo hanno chiesto a un ampio campione di pazienti sottoposti a colonoscopia (un esame clinico moderatamente invasivo e solitamente doloroso della durata media di 20 minuti) di quantificare su una scala da 0 («nessun dolore») a 10 («dolore estremo») il grado di sofferenza e disagio ogni sessanta secondi durante la procedura. Ai partecipanti veniva quindi chiesta una *valutazione globale* del disagio provato una volta che l'esame era concluso (ancora su una scala da 0 a 10).

Fig. 2 - Rappresentazione grafica del dolore percepito da due pazienti durante una colonoscopia (adattata da Redelmeier, Kahneman, 1996)



Si considerino i grafici in fig. 2, che rappresentano i dati ottenuti da due pazienti durante l'esame. L'asse *x* rappresenta il tempo, l'asse *y* l'intensità del dolore (riportata dai soggetti ogni sessanta secondi). Dall'osservazione dei due grafici, risulta ovvio che il paziente B ha sofferto durante l'esame più di quanto abbia sofferto il paziente A. Il *dolore totale* (soggettivamente percepito) è rappresentato dall'ampiezza delle superfici colorate. Come si può facilmente constatare, esso è tanto maggiore quanto più l'esperienza dolorosa si protrae nel tempo (cioè: quanto più la linea spezzata si estende verso destra) e quanto più è intenso il dolore percepito momento per momento (cioè: quanto più, in media, la linea spezzata si mantiene in alto). I risultati dello studio di Redelmeier e Kahneman mostrano però che, nel giudicare a poste-

riori la propria esperienza, i pazienti *trascurano* la durata complessiva dell'episodio: le variazioni nella durata dell'esame – da 4 minuti fino a più di un'ora – hanno influenzato in misura modesta o nulla l'opinione dei pazienti su quanto l'esame stesso fosse stato complessivamente spiacevole (indice di correlazione: 0,03).

Il giudizio finale complessivo dei pazienti era invece predetto molto bene da una media fra la valutazione del dolore più intenso provato durante la procedura e quella degli ultimi istanti (con un indice di correlazione di 0,65). In particolare, la media fra picco del dolore e fine dell'episodio predice le valutazioni complessive a posteriori del disagio provato più efficacemente del dolore totale.

Redelmeier e Kahneman hanno chiamato «regola del picco e della fine» (*peak & end rule*) il meccanismo che guida questo tipo di giudizi. A quanto pare, le persone, anziché registrare fedelmente e integrare tutti i dettagli di una complessa esperienza dolorosa che si è protratta nel tempo, valutano retrospettivamente quell'esperienza sulla base del richiamo selettivo di alcuni momenti distinti che fungono da strumenti euristici per il giudizio. Il momento più doloroso e quello conclusivo forniscono così punti di riferimento utili per elaborare un giudizio complessivo.

È da notare, però, che tale giudizio può discostarsi significativamente da quello ricavato da una valutazione a priori, che tiene conto anche di altri fattori, come appunto la durata. Vi sono infatti pochi dubbi sulla risposta che si otterrebbe chiedendo a una persona se preferisce sottoporsi a una procedura medica dolorosa di 8 oppure di 25 minuti, come quelle illustrate in fig. 2. Ancora una volta, questa discrepanza solleva problemi delicati per l'applicazione di regole sul consenso informato nella pratica medica. Ecco, con le parole di Kahneman, un vero e proprio «dilemma».

> Immaginate una procedura medica dolorosa che dura un certo numero di minuti e si conclude bruscamente quando il dolore è al massimo. Come abbiamo visto, il medico potrebbe probabilmente ottenere che il paziente conservi un ricordo più favorevole dell'intervento aggiungendovi una fase clinicamente superflua in cui il dolore diminuisce. Tuttavia, il paziente probabilmente rifiuterebbe l'offerta del medico di ottenere un ricordo migliore al prezzo di un dolore maggiore. Il medico dovrebbe procedere comunque, in nome del sé futuro del paziente? (Kahneman, 2000a, pp. 771-772).

I termini del dilemma sono esplorati in questo volume da Redelmeier, Katz e Kahneman nel saggio *Ricordi di una colonoscopia: uno studio randomizzato*. Il loro lavoro mostra come la fallacia del ricordo possa essere sfruttata positivamente per mitigare gli effetti di un esame doloroso nella memoria dei pazienti. In particolare, l'aggiunta alla procedura colonoscopica convenzionale di una conclusione clinicamente superflua ma di disagio minimo (con

l'estremità del colonoscopio lasciata ferma nel retto per pochi minuti) fa diminuire significativamente la valutazione globale a posteriori del dolore percepito e influisce positivamente sulla frequenza di ulteriori esami colonoscopici di controllo.

Percezione e comunicazione del rischio

Il caso della colonoscopia illustra come la conoscenza dei principi cognitivi che sottendono all'elaborazione delle informazioni possa essere sfruttata positivamente per modificare i comportamenti delle persone in situazioni rilevanti per la loro salute. Un altro importante esempio riguarda la *prospect theory* di Tversky e Kahneman (già menzionata nei paragrafi precedenti). Rothman e Salovey hanno esplicitamente proposto l'applicazione dei principi della *prospect theory* allo studio delle decisioni dei pazienti, delineando un'importante distinzione fra *interventi di controllo* (come un test o una visita di controllo) e *azioni preventive* (come l'adozione di una dieta più equilibrata, o le pratiche di sesso sicuro) (Rothman, Salovey, 1997).

Gli interventi di controllo sono tipicamente presentati e percepiti come comportamenti che coinvolgono un significativo elemento di rischio imminente (è possibile che si scopra un problema di salute) e la cui funzione è quella di limitare una potenziale «perdita» (attraverso contromisure tempestive). Le azioni preventive, d'altra parte, sono tipicamente presentate e percepite come comportamenti che tendono a garantire un «guadagno» (la conservazione di uno stato di benessere) senza coinvolgere rischi aggiuntivi per la salute. Data la diversa attitudine al rischio degli individui nell'ambito delle perdite e in quello dei guadagni – postulata dalla *prospect theory* – dovremmo aspettarci che la considerazione delle perdite incoraggi particolarmente le persone a sottoporsi a interventi di controllo (che rappresentano opzioni «rischiose») e che la considerazione dei benefici le incoraggi particolarmente ad adottare azioni preventive (che rappresentano opzioni «sicure»).

I risultati di numerosi studi sono in accordo con questa ipotesi. In un esperimento, Rothman, Salovey e collaboratori hanno presentato a due gruppi di donne sui quarant'anni due video con informazioni identiche, ma modi di presentazione differenti, intitolati rispettivamente *I benefici della mammografia* e *I rischi per chi trascura la mammografia* (Banks *et al.*, 1995). La maggiore efficacia del video presentato in termini di perdite è risultata significativa e duratura: le donne che lo avevano visto avevano una maggior probabilità di sottoporsi a una mammografia fino a un anno dopo che erano state contattate (66,2% contro 51,5%). Un simile vantaggio delle cornici «negative» nel promuovere interventi di controllo è stato osservato per l'autoesame del seno (Meyerowitz, Chaiken, 1987), l'autoesame della cute (Block, Keller, 1995), il

controllo periodico dei livelli di colesterolo nel sangue (Maheswaran, Meyers-Levy, 1990) e il test dell'HIV (Kalichman, Coley, 1995).

In ulteriori studi, Rothman, Salovey e collaboratori hanno invece mostrato la maggiore efficacia di un incorniciamento «positivo» in relazione a comportamenti preventivi, come l'utilizzo regolare di creme solari con un adeguato fattore di protezione solare per la prevenzione di patologie cutanee (Rothman *et al.*, 1993). Nel saggio *Effetto di incorniciamento e protezione solare*, Detweiler e collaboratori estendono questo risultato in una ricerca sul campo con più di 200 bagnanti.

Le indagini appena considerate mostrano che un'efficace comunicazione dei rischi deve avvalersi della conoscenza del modo in cui tali rischi sono percepiti dagli individui. In questo senso, una delle fonti di difficoltà risiede nel fatto che molti rischi per la salute emergono dall'esposizione *ripetuta* a condizioni potenzialmente dannose. Per fare un esempio, fra le persone che intraprendono un singolo viaggio in auto senza cinture di sicurezza, solo una su 100.000, in media, incorrerà in un danno permanente a causa di un incidente stradale. Tuttavia l'entità del rischio sale significativamente con la ripetizione del comportamento nel tempo: basta considerare un periodo di un anno (800 viaggi circa senza cinture di sicurezza) per calcolare che, in media, una persona su 125 (lo 0,8%) riporterà un danno permanente alla salute (da un punto di vista matematico, infatti, la probabilità di *non* riportare alcun danno permanente in 800 viaggi senza cinture equivale a 99.999/100.000 alla potenza di 800, cioè al 99,2%). Le ricerche mostrano però che le persone tendono a sottovalutare sistematicamente in che misura il rischio cresce per effetto di scelte di comportamento ripetute (per esempio, nell'ambito della contraccezione; si veda Shaklee, Fischhoff, 1990).

Fischhoff e collaboratori hanno identificato questa e altre discrepanze fra il modo in cui gli adolescenti si rappresentano il problema della prevenzione di malattie a trasmissione sessuale e un'analisi formale dello stesso problema basata sulle conoscenze degli esperti (Fischhoff, Downs, Bruine de Bruin, 1998). L'identificazione di questi specifici errori di valutazione ha quindi guidato la creazione di un DVD interattivo volto a migliorare la consapevolezza delle scelte degli adolescenti in relazione alla condotta sessuale. Un successivo studio randomizzato su 300 ragazze sessualmente attive fra i 14 e i 18 anni ha mostrato che, fra le partecipanti che venivano informate sui rischi con l'ausilio di questo DVD, l'incidenza di malattie sessualmente trasmesse diminuiva dal 23% al 12% circa in sei mesi (Downs *et al.*, 2004). Nel saggio che conclude la terza parte (*Un approccio cognitivo alle decisioni sul cancro: implicazioni per la prevenzione e il trattamento*), Fischhoff si serve di questo e di altri esempi per un'articolata discussione delle potenzialità della ricerca cognitiva sulle decisioni, con particolare riferimento alla prevenzione e al trattamento del cancro.

La rilevanza teorica e pratica degli errori cognitivi può essere apprezzata attraverso un'analogia con altri fenomeni ben noti. Per esempio, la forma fisica di un oggetto nel nostro campo visivo potrebbe essere determinata rigorosamente applicando i principi della geometria e dell'ottica; ma non è così che il nostro apparato percettivo affronta questo compito. Lo sappiamo perché, in certe condizioni, siamo vittima di «illusioni» sistematiche che altrimenti non avrebbero luogo. Basta osservare come un bastone parzialmente immerso nell'acqua ci appaia spezzato anche se è perfettamente dritto.

In che modo, quindi, riusciamo a evitare di ingannarci in casi come questo? È chiaro che guardare con più impegno non ci è di alcun aiuto: continuiamo a vedere quel che vediamo. Servirci della geometria e dell'ottica per valutare la forma di tutti gli oggetti che abbiamo intorno sarebbe un metodo sicuro, ma produrrebbe la paralisi. Sarebbe come voler giocare a scacchi calcolando a ogni passo tutte le varianti possibili: non andremmo mai oltre le mosse di apertura. Ma gli errori indagati in questo volume (non diversamente dall'effetto di rifrazione del bastone immerso nell'acqua) sono sistematici e quindi, in certe condizioni, *prevedibili*. Approfittando di ciò, quando rischiamo di cadere in un trabocchetto, possiamo prendere contromisure «mirate» che intervengono sui nostri processi mentali – modificando nella direzione corretta un giudizio distorto –, oppure sull'ambiente – mettendoci in condizione di osservare l'oggetto che ci interessa da un punto di vista differente.

Ebbene, in modo analogo, la conoscenza degli errori cognitivi ci permette di «aggiustare il tiro» proprio laddove il ragionamento euristico rischia di trarci in inganno, senza necessariamente dovervi rinunciare quando esso rappresenta un economico strumento per affrontare un'ampia classe di problemi. Anche in questo caso, gli interventi correttivi (detti *prescrittivi*) possono implicare l'apprendimento di strategie mirate per correggere valutazioni distorte oppure una ingegnerizzazione dell'ambiente in modo da rendere trasparente l'informazione rilevante per una data decisione (si veda Motterlini, Piattelli Palmarini, 2005).

L'apprendimento di accorgimenti correttivi *ad hoc* può rivelarsi utile in molte situazioni. È possibile, per esempio, che un medico e il suo paziente siano inclini a svolgere per primo un esame di facile esecuzione, decidendo solo in seguito per un controllo definitivo ma più invasivo. In un caso del genere, tenendo conto del rischio di un comportamento «non consequenziale», ci si dovrebbe domandare se l'esito del primo test richiede in ogni caso una conferma attraverso il secondo. Se si ritiene sia così, non c'è motivo di esitare: è bene eseguire fin da subito quest'ultimo. Un altro esempio è dato dalla tendenza a preferire un'opzione A che si differenzia chiaramente da altre due alternative fra loro simili, B e C, per evitare il confronto conflittuale fra que-

ste ultime. Sapendo che in una situazione del genere le nostre capacità decisionali intuitive possono tradirci (per esempio, nella selezione del paziente che è più urgente operare, o nella scelta di un farmaco) è possibile prendere delle contromisure: per esempio, controllare la robustezza delle nostre valutazioni verificando se l'opzione A ci appare chiaramente preferibile sia a B sia a C *prese ognuna singolarmente*. Se non è così, è probabile che la nostra avversione al conflitto decisionale ci stia inducendo in errore (Redelmeier, Shafir, 1995).

Per quanto riguarda il formato e la presentazione delle informazioni, abbiamo già visto come, nell'interpretazione di un test diagnostico, certe specifiche rappresentazioni del problema (per esempio, in termini frequentistici) rendano più agevole l'applicazione delle regole razionali pertinenti (il teorema di Bayes), inibendo la tendenza ad affidarsi a euristiche. Si consideri ora il problema delle scelte incoerenti prodotte da incorniciamenti diversi degli stessi dati (per esempio, in termini di perdite e in termini di guadagni). In questo caso, una semplice strategia correttiva consiste nell'impiegare *contemporaneamente* le diverse presentazioni possibili (Fischhoff, Slovic, Lichtenstein, 1980; Tversky, Kahneman, 1981). Il metodo del «doppio incorniciamento» è stato indagato sperimentalmente con risultati incoraggianti. In un problema di scelta fra due ipotetici trattamenti farmacologici per una patologia grave, si è osservato un tipico effetto di incorniciamento: la presentazione in termini di benefici induceva la maggioranza dei soggetti a preferire un certo trattamento (61,2%); la presentazione in termini di perdite a preferire l'altro (73,4%). Ma in un gruppo di partecipanti che riceveva le informazioni contemporaneamente *in entrambi i modi*, le preferenze fra le due alternative si sono distribuite intorno al 50%-50% (Bernstein, Chapman, Elstein, 1999). La distribuzione delle scelte ottenuta con la doppia presentazione era quindi meno estrema di quelle ricavate con i due «incorniciamenti» singoli, e si collocava a metà strada fra le due.

Gli esempi precedenti illustrano le potenzialità di interventi prescrittivi di tipo *non persuasivo*, vale a dire fondati sull'assunzione che gli individui siano in grado di giudicare e decidere efficacemente in modo autonomo – purché, tenendo conto dei limiti della cognizione umana, sia possibile aiutarli ad affrontare i problemi senza cader vittima di errori sistematici (si veda in proposito il saggio di Fischhoff in questo volume). Gli interventi *persuasivi*, per contro, fanno leva sulla conoscenza dei processi decisionali individuali per orientare le persone verso specifici corsi d'azione considerati benèfici. Esempi di quest'ultimo tipo sono offerti dalle indagini – citate nel paragrafo precedente – di Rothman, Salovey e collaboratori, nelle quali, dato un comportamento che si ritiene nell'interesse della salute degli individui, ci si propone di individuare la presentazione delle informazioni che sistematicamente lo favorisce. L'opportunità di un approccio persuasivo o non persuasivo dipende da-

gli obiettivi e dai risultati attesi nelle diverse circostanze. In ambito medico, un intervento non persuasivo volto a promuovere l'autonomia decisionale attraverso una rappresentazione più equilibrata ed efficace del problema è opportuno soprattutto quando i benefici di un trattamento sono oggetto di dibattito (per esempio, nel caso dello screening mammografico o del PSA in certe fasce di età, o del test dell'HIV in individui a basso rischio), mentre un intervento persuasivo può essere giustificato quando i benefici sono ben noti e particolarmente rilevanti (per esempio, nel caso della protezione dall'esposizione solare o della prevenzione di malattie a trasmissione sessuale).

Ricapitolando: l'analisi normativa dei problemi di giudizio e di scelta definisce i vincoli necessari per decidere e agire razionalmente, mentre lo studio empirico del comportamento degli individui mostra che nelle decisioni di medici e pazienti la conformità con i principi normativi della razionalità è tutt'altro che scontata.

Questo volume mostra come le conseguenze dello scarto tra comportamento ottimale e comportamento osservato siano particolarmente rilevanti in compiti quali l'elaborazione di una diagnosi, l'identificazione delle preferenze dei pazienti e la scelta fra diversi possibili corsi d'azione in condizioni di rischio e di incertezza. Riconoscere le cause di tale scarto attraverso l'indagine cognitiva diventa pertanto una condizione necessaria per promuovere il miglioramento delle decisioni attraverso la progettazione di interventi mirati.

Desideriamo ringraziare Marco Bobbio, Paolo Carli, Massimo Egidi, Baruch Fischhoff, Michel Gonzalez, Antonella Graiff, Francesco Guala, Jean-Pierre Meersseman, Massimo Piattelli Palmarini, Katya Tentori e Paolo Vineis per aver contribuito al nostro lavoro. Siamo particolarmente grati a Vittorio Girotto per aver commentato con critiche e suggerimenti varie stesure dell'Introduzione.

Ringraziamo inoltre gli autori dei diversi saggi e le istituzioni che hanno sostenuto la nostra attività durante la preparazione del libro: il Laboratorio di Scienze Cognitive dell'Università di Trento, il Laboratoire de Psychologie Cognitive dell'Università di Aix-Marseille I, il Centro di ricerca in epistemologia sperimentale e applicata (CRESA) dell'Università Vita-Salute San Raffaele di Milano e l'Università di Firenze. Infine, questo libro non sarebbe stato possibile senza l'interesse e il sostegno della Fondazione Smith Kline, e in particolare di Paolo Rizzini e Maurizio Bassi, il cui impegno ha permesso di portare a compimento il progetto.

I. Giudizi e probabilità

*1. Il ragionamento probabilistico nella medicina clinica: problemi e opportunità**

di *David M. Eddy*

La qualità e i costi delle cure sono in gran parte determinati dalle decisioni dei medici, il cui obiettivo ultimo è quello di elaborare e prescrivere programmi di trattamento in grado di migliorare le condizioni dei pazienti. Nella maggior parte di queste decisioni sono coinvolti numerosi fattori, una notevole incertezza e difficili questioni di valore.

Questo capitolo esamina un aspetto del modo in cui vengono prese le decisioni mediche, studiando l'uso del ragionamento probabilistico al fine di analizzare un particolare problema: la scelta di eseguire una biopsia su una donna con un nodulo al seno che potrebbe essere maligno. Ci occuperemo, in particolare, di come i medici elaborano le informazioni relative ai risultati di una mammografia. I dati che presenteremo mostrano che i medici incontrano serie difficoltà nel fronteggiare l'incertezza, che nel ragionamento probabilistico molti di loro incorrono in alcuni importanti errori e che tali errori possono influire negativamente sulla qualità delle cure.

1. Il problema

La biopsia al seno è una procedura medica piuttosto impegnativa. Nella sua forma più diffusa (circa l'80% dei casi) una massa sospetta individuata nel seno viene rimossa e quindi sottoposta a esame microscopico e a diagnosi istologica da parte di un patologo. Normalmente si ricovera la paziente in ospedale e si effettua una serie di test diagnostici preoperatori. La biopsia avviene quasi sempre in anestesia generale (con una probabilità di circa due su

* Traduzione di «Probabilistic reasoning in clinical medicine: problems and opportunities», in Kahneman, D., Slovic, P., Tversky, A. (a cura di), *Judgment under uncertainty: heuristics and biases*, Cambridge University Press, New York, 1982, pp. 249-267. Per gentile concessione di IRSA.

diecimila di morte associata al trattamento anestesiologico). Si effettua una piccola incisione (da 2,5 a 5 cm) e si rimuove un tessuto all'incirca delle dimensioni di una noce. In molti casi (forse uno su due) la perdita di tessuto si nota appena. Altre volte resta una rientranza visibile nella zona dell'intervento. In rari casi (forse uno su duecento) si verificano un'infezione o un drenaggio che possono protrarsi per diverse settimane. L'entità della spesa è di circa settecento dollari. La procedura può anche essere svolta in ambulatorio e sotto anestesia locale. In alternativa alla biopsia con incisione, i chirurghi preferiscono in alcuni casi prelevare il tessuto servendosi di un ago. Questo tipo di intervento può essere eseguito in ambulatorio, non lascia cicatrici né altri effetti residui ed è molto meno costoso. Tuttavia molti medici ritengono che sia meno affidabile, nel senso che è possibile che una lesione effettivamente maligna non venga identificata.

Un importante fattore che influenza l'opportunità di eseguire una biopsia è la possibilità che la massa rilevata sia un cancro. Per valutare questa possibilità un medico può elencare le possibili patologie, tenere conto della frequenza con la quale vari segni e sintomi clinici si manifestano in ciascuna della patologie considerate, confrontare queste informazioni con le condizioni della paziente, stimare la probabilità che soffra di ciascuna delle patologie comprese nella lista, ed eseguire una biopsia se la probabilità di un cancro o di un'altra lesione curabile è sufficientemente alta. Per venire in aiuto dei medici, molti libri di testo descrivono in che modo le patologie non maligne possono essere distinte dal cancro. Per esempio il passaggio seguente descrive una di queste condizioni benigne – la malattia cistica cronica del seno (*chronic cystic disease*).

> La malattia cistica cronica viene spesso confusa con il carcinoma del seno. *Solitamente* si trova in donne con un seno di piccole dimensioni. Si presenta *prevalentemente* nel quadrante superiore esterno ma *può* talvolta trovarsi in altre zone e anche arrivare a coinvolgere l'intero seno. È *spesso* dolorosa, particolarmente nel periodo premestruale, e sono *comuni* disturbi mestruali concomitanti. Nel *15% circa* dei casi si hanno perdite dal capezzolo, *solitamente* serie, ma non modificazioni del capezzolo stesso. La lesione *è* diffusa, senza una demarcazione netta, e non è fissata alla superficie epidermica. Le cisti multiple sono solide, rotonde e fluttuanti e *possono* essere transilluminate *se* contengono fluidi trasparenti. Una cisti estesa in un'area affetta da malattia cistica cronica *somiglia* a un tumore, ma *di solito* è più uniforme e ben delimitata. I linfonodi dell'ascella *solitamente* non sono ingrossati. *In rari casi* la malattia cistica cronica si manifesta con cisti estese di colore bluastro. Più frequentemente le cisti sono numerose e di piccole dimensioni (del Regato, 1970, pp. 860-861)[1].

1. In questa e in tutte le successive citazioni nel presente capitolo i corsivi sono stati aggiunti.

Si trovano comunemente descrizioni simili per fibroadenoma, necrosi del tessuto adiposo, trauma, e per una mezza dozzina di altre condizioni cliniche del seno, così come per il cancro stesso.

Informazioni probabilistiche di questo genere possono aiutare il medico ad analizzare le possibili cause della presenza di una massa nel seno della paziente. Associando dei valori ai possibili esiti (per esempio: diagnosi corretta di cancro; esecuzione di una biopsia non necessaria per una lesione non maligna; mancata esecuzione della biopsia e mancata identificazione di una lesione maligna; decisione appropriata di non eseguire una biopsia su una lesione benigna), il medico può valutare la probabilità che la paziente, con i suoi particolari segni e sintomi, abbia il cancro e selezionare un corso d'azione.

2. Il caso della mammografia

Vi sono altri strumenti diagnostici che possono aiutare il medico a stimare la probabilità che la lesione al seno di una particolare paziente sia maligna. Il più importante e comunemente usato è forse la mammografia. Il valore di questo test risiede nel fatto che le componenti delle cellule maligne assorbono i raggi X in modo differente dalle componenti di cellule non maligne. Attraverso lo studio dei mammogrammi, un radiologo può essere in grado di individuare alcuni segni che si presentano con diverse frequenze in differenti tipi di lesioni e formulare su questa base un giudizio riguardo alla natura della lesione in questione. Solitamente i mammogrammi sono classificati come positivi o negativi (in riferimento al cancro). Occasionalmente viene impiegato uno schema di classificazione più esteso, che può per esempio comprendere tre classi: massa maligna, sospetta o benigna.

Il test non è perfetto: alcune lesioni maligne sono erroneamente classificate come benigne e alcune lesioni benigne sono identificate come maligne. Il grado di accuratezza del test è quindi un fattore di grande importanza per il medico.

2.1. *Il ragionamento probabilistico*

Sviluppiamo ora con più precisione questo punto. L'obiettivo di un test diagnostico è quello di fornire al medico delle informazioni riguardo alle condizioni di un paziente. Il medico usa queste informazioni per rivedere le sue precedenti stime sulle condizioni del paziente e per selezionare un corso d'azione in base alle sue nuove stime. L'azione selezionata può essere la prescrizione di ulteriori test diagnostici o, se il medico è sufficientemente fiducioso riguardo alle condizioni del paziente, un intervento terapeutico. Il punto

essenziale è che il medico può avere diversi gradi di certezza riguardo alle condizioni del paziente. Il medico raccoglierà informazioni per raffinare, per esempio, la certezza che il paziente abbia o non abbia il cancro. Quando tale certezza sarà diventata sufficientemente forte, egli agirà in considerazione della severità della malattia e del cambiamento della prognosi che si aspetta a seguito del trattamento.

Possiamo associare un valore di probabilità a un certo grado di certezza, ossia la probabilità soggettiva del medico riguardo all'ipotesi che il paziente abbia il cancro. L'impatto di un test diagnostico come la mammografia sulla pratica clinica risiede quindi nella sua capacità di modificare la probabilità soggettiva del medico riguardo al fatto che una paziente abbia il cancro.

La nozione di probabilità soggettiva o di grado di certezza compare nel linguaggio medico in molte forme differenti. Per esempio, un autore scrive che «poiché il gruppo di età più avanzata ha la più alta proporzione di lesioni maligne, c'è un innalzamento del *livello di sospetto* di cancro nella mente di un medico di fronte a una paziente più anziana» (Gold, 1969, p. 162). Un altro autore sostiene che la mammografia può ridurre il numero di biopsie al seno «in molti casi in cui un'*opinione alquanto solida* del medico riguardo alla natura benigna di una patologia è sostenuta dalla *solida diagnosi mammografica* di una lesione benigna» (Wolfe, 1964, p. 253). Altri autori descrivono il concetto nel modo seguente: «se l'*impressione soggettiva* del medico dà ragioni sufficienti per sospettare un carcinoma, egli sarà spinto a raccomandare una biopsia nonostante un mammogramma negativo» (Clark *et al.*, 1965, p. 133). Ulteriori espressioni che riflettono questa stessa nozione includono «*livello di fiducia*» (Byrne, 1974, p. 37), «l'*impressione* di una condizione maligna» (Wolfe, 1967, p. 138), «una diagnosi *più sicura*» (Egan, 1972, p. 392) e altre ancora. Queste espressioni sono imprecise perché pochi medici hanno familiarità con gli aspetti formali del concetto di probabilità soggettiva e dell'analisi delle decisioni. Sembra comunque chiaro che i medici trovano naturale la nozione di «grado di certezza» e che ne fanno uso nel tentativo di selezionare un corso d'azione appropriato.

2.2. *Interpretare l'accuratezza di una mammografia*

Consideriamo una paziente con un nodulo al seno che il medico ritiene sia probabilmente benigno, supponiamo con una probabilità del 99%. Possiamo interpretare la frase «che il medico ritiene sia probabilmente benigno (con una probabilità del 99%)» come segue. Supponiamo che il medico si sia imbattuto in un certo numero di donne simili a questa particolare paziente sotto tutti gli aspetti importanti (come età, sintomi, storia familiare e riscontri fisici) e quindi sappia dalla sua esperienza clinica che la frequenza del cancro in

questo gruppo è, diciamo, di uno su cento. In mancanza di altre informazioni, il medico assegnerà pertanto (forse inconsciamente) una probabilità soggettiva dell'1% all'ipotesi che questa paziente abbia il cancro.

Supponiamo ora che il medico ordini l'esecuzione di un mammogramma e riceva il resoconto del radiologo secondo il quale la lesione è maligna. Questo dato rappresenta una nuova informazione e i successivi interventi dipendono ovviamente dalla nuova stima del medico riguardo alla probabilità che la paziente abbia il cancro. Un medico che si rivolga alla letteratura scientifica può trovare moltissime indicazioni utili, come le seguenti: «l'accuratezza della mammografia è approssimativamente del 90%» (Wolfe, 1966, p. 214); «in [una paziente con un nodulo al seno] un risultato [mammografico] positivo per il carcinoma è altamente accurato» (Rosato, Thomas, Rosato, 1973, p. 491); «l'accuratezza della mammografia nel diagnosticare correttamente lesioni maligne del seno si aggira in media fra l'80 e l'85%» (Cohn, 1972, p. 98). Se desidera informazioni più dettagliate, il medico può trovare asserzioni come: «i risultati mostrano che il 79,2% di 475 lesioni maligne sono state diagnosticate correttamente e il 90,4% di 1.105 lesioni benigne sono state diagnosticate correttamente, per un tasso di accuratezza complessivo dell'87%» (Snyder, 1966, p. 217).

A questo punto potete approfondire la vostra comprensione del problema che il medico deve affrontare stimando voi stessi la nuova probabilità che questa paziente abbia il cancro. Il medico ritiene che il nodulo sia probabilmente (99%) *benigno*, ma il radiologo ha fornito un risultato *positivo* il cui livello di accuratezza è quello appena visto.

Per determinare questo valore di probabilità è possibile applicare il teorema di Bayes. Tale teorema ci dice che

$$p(\text{ca}|\text{pos}) = \frac{p(\text{pos}|\text{ca}) \times p(\text{ca})}{[p(\text{pos}|\text{ca}) \times p(\text{ca})] + [p(\text{pos}|\text{ben}) \times p(\text{ben})]}$$

dove:

$p(\text{ca}|\text{pos})$ è la probabilità che la paziente abbia il cancro dato che l'esito dell'esame radiografico a cui è stata sottoposta è positivo (tale probabilità è anche detta *probabilità a posteriori*);

$p(\text{pos}|\text{ca})$ è la probabilità che, nel caso in cui la paziente abbia il cancro, il radiologo lo diagnostichi correttamente (tale probabilità è anche detta *tasso di vere positività*, o *sensibilità* del test);

$p(\text{ca})$ è la probabilità che la paziente abbia il cancro (anche detta *probabilità a priori*);

$p(\text{ben})$ è la probabilità a priori che la paziente sia affetta da una condizione benigna (si noti che $p(\text{ben}) = 1 - p(\text{ca})$);

p(pos|ben) è la probabilità che, nel caso in cui la paziente sia affetta da una condizione benigna, il radiologo diagnostichi erroneamente un cancro (il *tasso di false positività*).

La tab. 1 riassume le cifre fornite da Snyder (1966). I numeri all'interno delle diverse celle rappresentano i valori di probabilità (per es., p(pos|ca) = 0,792).

Tab. 1 - Accuratezza della mammografia nella diagnosi di lesioni benigne e maligne (dati tratti da Snyder, 1966)

	Lesione maligna (cancro)	*Lesione benigna (no cancro)*
Risultato radiografico positivo	0,792	0,096
Risultato radiografico negativo	0,208	0,904

Usando l'1% come stima del medico della probabilità a priori che il nodulo sia maligno, e tenendo conto della nuova informazione rappresentata dal risultato del test, abbiamo:

$$p(\text{ca}|\text{pos}) = \frac{0{,}792 \times 0{,}01}{(0{,}792 \times 0{,}01) + (0{,}096 \times 0{,}99)} = 0{,}077$$

Quindi il medico dovrebbe stimare che la probabilità che la paziente abbia il cancro sia approssimativamente dell'8%.

2.3. *Errori del ragionamento probabilistico*

Sfortunatamente, la maggior parte dei medici (circa 95 su 100 in un campione informale studiato dall'autore) fraintende le asserzioni relative all'accuratezza del test e stima p(ca|pos) al 75% circa. Altri studiosi hanno ottenuto risultati simili (Casscells, Schoenberger, Grayboys, 1978)[2]. Se interrogati, i medici che hanno risposto in modo erroneo solitamente dichiarano di avere assunto che la probabilità del cancro dato che la paziente aveva avuto un test positivo – cioè p(ca|pos) – fosse approssimativamente uguale alla probabilità di un test positivo in una paziente con il cancro – cioè p(pos|ca). Questa se-

2. Si veda anche, in questo volume, il cap. 12 [N.d.C.].

conda probabilità è quella che viene misurata nelle ricerche cliniche ed è una nozione molto familiare ai medici, ma è la prima probabilità che risulta necessaria per la decisione clinica. A quanto pare, molti medici, e forse la maggior parte di loro, confondono l'una con l'altra.

Per ogni test diagnostico utile nel determinare la presenza di una specifica patologia vi sono in realtà due tipi di accuratezza. L'*accuratezza retrospettiva* riguarda p(pos|ca) e p(neg|no ca). (L'abbreviazione «no ca» si riferisce al caso in cui il paziente non ha il cancro. Questo può accadere sia perché è affetta da una patologia benigna sia perché non è affetta da alcuna patologia.) Questa accuratezza, quella solitamente considerata nella letteratura sulla mammografia, è determinata analizzando le diagnosi ottenute attraverso la mammografia dopo che le diagnosi corrette (istologiche) sono ormai note. Usiamo invece il termine *accuratezza predittiva* per indicare p(ca|pos) e p(ben|neg), cioè il tipo di accuratezza utile per il medico che dispone del risultato di un esame in una paziente senza una diagnosi certa e intende prevedere quale sia la condizione clinica della paziente.

La confusione fra accuratezza retrospettiva e predittiva. Una rassegna della letteratura medica sulla mammografia rivela una forte tendenza a equiparare l'accuratezza predittiva di un risultato positivo con l'accuratezza retrospettiva del test, cioè una forte tendenza ad assumere che valga p(ca|pos) = p(pos|ca). Ci sono molte ragioni per sospettare la presenza di questo errore. Innanzi tutto, la formulazione di molte asserzioni che si trovano in letteratura suggerisce con forza che gli autori ritengono che l'accuratezza predittiva equivalga all'accuratezza retrospettiva riportata nei loro studi. Per esempio, in un articolo apparso su «Radiology» nel 1964 si dichiarava: «la *correttezza totale* delle diagnosi ottenute attraverso l'esame radiografico è stata di 674 casi su 759, o dell'89%» (vol. 84, p. 254). L'autore di un contributo a «Clinical Obstetrics & Gynecology» del 1966 ha scritto: «Asch ha trovato una *correlazione* del 90% fra la mammografia e le condizioni patologiche di 500 pazienti» (vol. 9, p. 217). Ancora: «L'*accordo* fra diagnosi radiografiche e diagnosi patologiche è stato del 91,6%» (Egan, 1972, p. 379). Tutte queste asserzioni sembrano suggerire che, se la paziente ha un test positivo, il test sarà corretto e la paziente avrà il cancro il 90% delle volte. Questo non è vero.

In secondo luogo, alcuni autori commettono l'errore in modo del tutto esplicito. Il passo seguente è apparso su un numero del 1972 di «Surgery, Gynecology and Obstetrics» in un articolo intitolato *La mammografia in una corretta prospettiva* che, nelle intenzioni dell'autore, avrebbe dovuto rettificare alcune confusioni che si trovano in letteratura: «Fra le donne in cui la presenza di un carcinoma è provata e sulle quali vengono eseguiti esami mammografici, in circa un caso su cinque il test radiografico non individua una lesione maligna. Quindi se, sulla base di un mammogramma negativo, riman-

diamo una biopsia su una lesione al seno, c'è una probabilità di uno su cinque che stiamo rimandando la biopsia su di una lesione maligna» (vol. 134, p. 98). L'autore ha erroneamente asserito che p(neg|ca) = 0,2 implica p(ca|neg) = 0,2. L'errore risulta particolarmente serio, in quanto lo spinge a concludere che «rimandare la biopsia su una lesione che appare *benigna su basi cliniche* e che risulta *benigna* in base a una mammografia significa fare un passo indietro nell'eradicazione del carcinoma del seno nella popolazione femminile». La probabilità che una paziente in questa situazione abbia il cancro dipende dalla probabilità a priori, ma è comunque inferiore all'1%. L'analisi dell'autore dell'articolo mostra quindi un errore di un fattore 20.

«Surgery, Gynecology and Obstetrics» ha pubblicato nel 1970 (vol. 131, pp. 93-98) i risultati ottenuti da un gruppo di ricerca che ha calcolato «la correlazione fra le diagnosi radiologiche e quelle patologiche» nel modo seguente. Gli autori hanno preso tutte le pazienti con diagnosi istologiche provate e le hanno suddivise in tre gruppi sulla base delle diagnosi radiografiche: «lesione benigna», «carcinoma» e «sospetto carcinoma». Nel gruppo relativo alla categoria «lesione benigna» (che nella nostra terminologia corrisponde a «risultato negativo») i calcoli hanno mostrato che l'84% delle pazienti aveva in effetti lesioni benigne. Si è anche osservato che l'87,5% delle pazienti nel gruppo relativo alla categoria «carcinoma» (o «risultato positivo») aveva lesioni maligne istologicamente provate. Quindi p(ca|pos) = 87,5% e p(ben|neg) = 84%. Ma gli autori hanno confuso questa accuratezza (predittiva) con l'accuratezza retrospettiva, asserendo che «si è ottenuta una diagnosi mammografia corretta nell'84% *delle pazienti con lesioni benigne* e nell'87,5% *di quelle con carcinoma*». In realtà, in questo studio il tasso di veri positivi – cioè p(pos|ca) – è stato del 66% e quello di veri negativi – cioè p(neg|ben) – del 54%.

In una lettera al direttore comparsa nel numero dell'11 settembre del 1976 del «National Observer» un medico ha presentato una serie di cinque «osservazioni e dati di fatto» per sostenere la sua opinione secondo la quale «lo screening mammografico periodico non è nel vero interesse della popolazione generale in nessuna fascia di età». Ecco il primo insieme di osservazioni.

> L'accuratezza dell'esame mammografico è stimata fra l'80% e il 90% a seconda di fattori come l'età della paziente, la presenza concomitante di una malattia fibrocistica, il tipo di equipaggiamento radiografico impiegato, l'esperienza del radiologo e la stessa definizione del termine «accurato» […] Anche ammettendo che l'accuratezza sia dell'85% (e sono sicuro che non ogni radiologo nel paese è in grado di raggiungere questo livello), ciò significa che *il 15% delle donne sottoposte a esame radiografico otterranno una interpretazione erronea dei risultati o, più probabilmente, che i loro mammogrammi non riveleranno la presenza della malattia. Il 15% delle donne verranno così erroneamente rassicurate dalla notizia che i loro esami radiografici sono normali,*

quando in realtà esse hanno già il cancro. È difficile valutare i costi per questo gruppo, perché queste persone sarebbero ovviamente in una condizione migliore senza alcuna informazione piuttosto che in possesso di informazioni scorrette. Una volta che le si è detto che il suo mammogramma è normale e che non è necessario che torni a farsi controllare per un anno, una donna con il cancro al seno potrebbe ignorare un nodulo che, altrimenti, l'avrebbe spinta a recarsi immediatamente dal suo dottore.

Vi sono diversi errori nel ragionamento di questo autore. Innanzi tutto, l'«accuratezza» della mammografia non può essere espressa in una singola cifra. Assumeremo comunque che l'autore intenda che sia il tasso di veri positivi sia il tasso di veri negativi equivalgano all'85%.

In secondo luogo, queste probabilità (85%) sono quelle che si osservano quando la mammografia è usata per una diagnosi differenziale in condizioni nelle quali sono noti segni e sintomi clinici. In questi casi si ha a che fare con lesioni generalmente più avanzate di quelle rilevate negli esami di screening, cioè nella situazione alla quale l'autore dell'articolo si riferisce. Stime più ragionevoli per il tasso di veri positivi e il tasso di veri negativi nel contesto dei programmi di screening sono rispettivamente del 60% e del 98%.

Terzo, anche usando la cifra dell'85%, il ragionamento è scorretto in diversi punti. Si consideri la seconda frase. Un'interpretazione erronea dei risultati del test può aver luogo in due modi: la paziente può avere il cancro ma avere un risultato negativo – la probabilità congiunta dei due eventi si indica solitamente con $p(\text{ca} \,\&\, \text{neg})$; oppure la paziente può avere un risultato positivo ma non avere il cancro – la probabilità che si verifichi questa situazione corrisponde, in termini formali, a $p(\text{no ca} \,\&\, \text{pos})$. Dalla teoria della probabilità sappiamo che

$$p(\text{ca} \,\&\, \text{neg}) = p(\text{neg}|\text{ca}) \times p(\text{ca})$$

$p(\text{neg}|\text{ca})$ è complementare a $p(\text{pos}|\text{ca})$ ed è quindi, in questo caso, uguale a 0,15. La probabilità $p(\text{ca})$ non è nota con precisione, ma per una popolazione sottoposta a screening è ragionevolmente certo che sia inferiore a 0,005. Ciò significa che meno di 5 donne su 1.000 sono affette da un cancro al seno asintomatico ma rilevabile attraverso un mammogramma.

Quindi:

$$p(\text{ca} \,\&\, \text{neg}) \leq 0{,}15 \times 0{,}005 = 0{,}00075$$

Inoltre:

$$p(\text{no ca} \,\&\, \text{pos}) = p(\text{pos}|\text{no ca}) \times p(\text{no ca}) \geq 0{,}15 \times 0{,}995 = 0{,}14925$$

La probabilità totale di una interpretazione erronea del test da parte del radiologo è la somma di queste due cifre, cioè p(ca & neg) + p(no ca & pos), ed equivale al 15%, come l'autore dell'articolo effettivamente asserisce. Tuttavia, questo non significa affatto che «più probabilmente, i loro mammogrammi non riveleranno la presenza della malattia». Infatti, un evento la cui probabilità è p(ca & neg) = 0,00075 *non è* più probabile di uno la cui probabilità equivale a p(na ca & pos) = 0,14925. Il primo evento è, piuttosto, circa duecento volte *meno* probabile del secondo.

Un ulteriore problema è che un'«accuratezza» dell'85% non significa che «il 15% delle donne verranno erroneamente rassicurate dalla notizia che i loro esami radiografici sono normali». A quanto pare, l'autore sta qui cercando di stimare p(ca|neg). Ora, in base al teorema di Bayes:

$$p(\text{ca}|\text{neg}) = \frac{p(\text{neg}|\text{ca}) \times p(\text{ca})}{[p(\text{neg}|\text{ca}) \times p(\text{ca})] + [p(\text{neg}|\text{no ca}) \times p(\text{no ca})]} =$$

$$= \frac{0{,}15 \times 0{,}005}{(0{,}15 \times 0{,}005) + (0{,}85 \times 0{,}995)} = 0{,}00089$$

Questo significa che, se 10.000 donne asintomatiche vengono sottoposte a screening, e se utilizziamo la stima (inesatta) dell'autore riguardo all'accuratezza del test, 8.458 di loro avranno un risultato negativo. L'autore ritiene che circa 1.269 di queste riceveranno una falsa rassicurazione. In realtà solo circa 9 di loro si troveranno in questa situazione. Questa cifra è stata quindi sovrastimata di un fattore di circa 150.

Infine, l'aggiunta della frase «quando in realtà esse hanno già il cancro» crea ulteriore confusione. Le frasi «verranno erroneamente rassicurate», «dalla notizia che [cioè: dato che] i loro esami radiografici sono normali» e «quando in realtà esse hanno già il cancro» vanno tradotte simbolicamente in p(ca|neg & ca). Questa probabilità equivale a 1, non a 0,15.

L'importanza di p*(ca)*. Oltre a confondere i due tipi di accuratezza, molti autori non sembrano rendersi conto che, per un test per il quale si assume un valore costante dell'accuratezza retrospettiva, il significato dell'esito del test per il medico (l'accuratezza predittiva) dipende dal rischio iniziale di cancro nella paziente sottoposta a mammografia. Anche assumendo che il tasso di veri positivi e di veri negativi siano costanti in tutti gli studi, l'interpretazione appropriata dell'esito dell'esame – la probabilità che una paziente con un mammogramma positivo (o negativo) abbia il cancro – dipenderà dalla prevalenza del cancro nella popolazione alla quale la paziente appartiene, cioè dalla probabilità che la paziente abbia il cancro prima dell'esecuzione del test. Questo può essere di estrema importanza nel momento in cui si confronta l'uso del test in un ambito clinico, nel quale le pazienti mostrano segni e sin-

tomi indicativi di una patologia al seno, con l'uso del test in un programma di screening rivolto a donne asintomatiche.

L'importanza di questo punto può essere illustrata da un esempio. Supponiamo che la pratica usuale di un medico sia quella di far eseguire una mammografia su pazienti in cui l'esame fisico dia esiti anomali. In uno studio pubblicato, in questa popolazione di donne è stata osservata una frequenza di casi di cancro al seno dell'8% circa (Wolfe, 1964). In una serie di mammogrammi in questa popolazione si è ottenuto un tasso di veri positivi del 92% e un tasso di veri negativi dell'88% (*ibid.*). Ammettiamo ora che il medico si trovi di fronte una paziente che egli ritiene sia rappresentativa di questa specifica popolazione – assumiamo, cioè, che valga $p(\text{ca}) = 8\%$. Supponiamo che egli prescriva l'esecuzione di una mammografia e riceva dal radiologo un esito positivo. La decisione di raccomandare l'esecuzione di una biopsia dovrebbe essere basata sulla probabilità aggiornata che la paziente abbia il cancro. Tale probabilità può essere calcolata ed equivale al 40% (si veda tab. 2). Ora chiediamoci: un esito negativo avrebbe escluso la diagnosi di cancro? La probabilità che questa paziente, dato un esito negativo, abbia ciononostante il cancro è di poco inferiore all'1%. La logica che conduce a questa stima è illustrata in tab. 2.

Tab. 2 - Presenza del cancro e risultati degli esami radiografici in 1.000 donne con risultati anomali dell'esame fisico (Wolfe, 1964)

	Donne con cancro	*Donne senza cancro*	*Totale*
Donne con esami radiografici positivi	74	110	184*
Donne con esami radiografici negativi	6	810	816
Totale	80	920	1.000

* Si noti che un tasso di veri positivi di 0,92 [cioè $p(\text{pos}|\text{ca}) = 0{,}92$] implica che, su 80 donne che hanno il cancro, 74 avranno un esame radiografico positivo e 6 avranno un esame radiografico negativo. Fra tutte le donne con esami radiografici positivi, 74/184 hanno il cancro, vale a dire che $p(\text{ca}|\text{pos}) = 74/184 = 40\%$.

Supponiamo invece che un medico prescriva il test mammografico a fini di controllo in una paziente che non ha sintomi né un esame fisico anormale. Fra le donne in questa situazione la prevalenza di un cancro che possa essere individuato dalla mammografia è dello 0,10% circa (si veda, per esempio, Shapiro, Strax, Venet, 1967). A scopo illustrativo, supponiamo che l'accuratezza retrospettiva del risultato radiografico sia immutata rispetto a prima – cioè, supponiamo che anche in questa popolazione il tasso di veri positivi sia del 92% e il tasso di veri negativi sia dell'88%[3]. La letteratura fornisce sol-

3. Si tratta di un'assunzione inadeguata, perché l'«accuratezza» cambia a seconda della popolazione sottoposta all'esame. Per esempio, il tasso di veri positivi è inferiore in una popola-

tanto dati relativi all'accuratezza retrospettiva del test in donne con il cancro o con una patologia benigna. Tuttavia in uno studio pubblicato si è osservato che il 60% circa di un campione di donne asintomatiche non avevano alcuna patologia (Wolfe, 1965). Nel caso che stiamo discutendo, per il teorema di Bayes, vale che:

$$p(\text{ca}|\text{pos}) = \frac{p(\text{pos}|\text{ca})p(\text{ca})}{p(\text{pos}|\text{ca})p(\text{ca}) + p(\text{pos}|\text{ben})p(\text{ben}) + p(\text{pos}|\text{no pat})p(\text{no pat})}$$

p(ben), p(no pat) e p(pos|no pat) non sono esplicitamente discusse nella letteratura. Si tratta di un fatto istruttivo, che induce a sospettare che la loro importanza nell'analisi di questi problemi non sia ben compresa. Per il nostro esempio utilizzeremo i dati forniti da Wolfe (1965) e assumeremo che p(no pat) sia del 60% circa e che p(ben) sia del 40% circa. Faremo anche l'assunzione, favorevole alla mammografia, che valga p(pos|no pat) = 0%.

Proseguendo con questo esempio, supponiamo che il radiologo riporti che la mammografia in questa donna asintomatica è positiva. Dato il risultato positivo dell'esame, la probabilità che la paziente abbia il cancro p(ca|pos) è approssimativamente di 1 su 49, cioè il 2% circa (si veda tab. 3). Nell'esempio precedente, che riguardava una paziente con sintomi, p(ca|pos) era del 40%. Quindi, a seconda della particolare paziente che viene sottoposta al test, ci può essere una differenza *di venti volte* nella probabilità che una donna con una mammografia positiva abbia il cancro.

Tab. 3 - Presenza del cancro e risultati degli esami radiografici in 1.000 donne senza sintomi

	Donne con cancro	*Donne con lesioni benigne*	*Donne senza alcuna patologia*	*Totale*
Donne con esami radiografici positivi	1	48	0	49*
Donne con esami radiografici negativi	0	352	599	951
Totale	1	400	599	1.000

* Un tasso di veri positivi di 0,92 indica che l'esame radiografico identificherà il cancro in una paziente che ne è affetta. Un tasso di veri negativi di 0,88 in riferimento a una lesione benigna implica che, su 400 donne con una lesione benigna, 352 avranno un esame radiografico negativo, mentre in 48 casi l'esame radiografico sarà positivo. Quindi, in tutto 49 donne avranno esami radiografici positivi, ma solo una di esse avrà il cancro, vale a dire che p(ca|pos) = 1/49 = 2%.

zione asintomatica perché le masse maligne tendono a essere più piccole e più difficili da identificare. Manteniamo questa assunzione nel presente contesto solo per illustrare l'importanza della probabilità a priori p(ca).

Tutto ciò solleva un'importante questione riguardo al ragionamento medico: nel valutare i segni e i sintomi di una paziente, in che modo il medico dovrebbe servirsi delle informazioni relative alla frequenza di base delle diverse possibili patologie considerate nella popolazione di riferimento? La professione sembra essere confusa su questo problema. Da una parte, i medici dichiarano che la diffusione relativa di una patologia *non* dovrebbe influire sulla stima della probabilità che una particolare paziente abbia quella patologia. Questa nozione compare in varie massime come «ogni paziente è un caso a sé» e «le statistiche sono mute riguardo agli individui» (*statistics are dead for men*). Nella discussione di problemi specifici, questa stessa idea è talvolta espressa in una forma più raffinata, per esempio nell'asserzione «le donne più giovani hanno ovviamente un minor numero di lesioni maligne; questo, tuttavia, dovrebbe avere un'influenza minima nella valutazione di un caso individuale» (Wolfe, 1967, p. 138). Lo stesso punto può trovarsi talvolta espresso esplicitamente come una regola cui obbedire, per esempio nel passo seguente, tratto da un libro di testo sulla diagnosi clinica: «Quando un paziente con una patologia non diagnosticata consulta un medico, né il primo né il secondo sa se si tratta di una condizione rara fintanto che non è stata elaborata la diagnosi definitiva. I metodi statistici possono essere applicati soltanto a popolazioni di migliaia di individui. Il singolo paziente o ha una malattia rara, oppure non ce l'ha; l'incidenza relativa delle condizioni cliniche in questione è completamente irrilevante nell'elaborazione di una diagnosi per un particolare paziente» (DeGowin, DeGowin, 1969, p. 6).

D'altra parte, queste asserzioni sono spesso in contrasto con il comportamento dei medici, che tentano, per quanto in modo imperfetto, di usare questo tipo di informazioni diagnostiche. Lo testimoniano le seguenti massime, anch'esse diffuse nelle Facoltà di Medicina: «se sentite dei nitriti, pensate a un cavallo e non a una zebra», «gli eventi più comuni accadono più comunemente», ecc. Sembra quindi che molti medici percepiscano il valore delle probabilità iniziale di una malattia, ma che gli aspetti formali della teoria della probabilità non siano ben compresi. Senza un approccio formale, i medici tendono a fare gli stessi errori del ragionamento probabilistico che sono stati osservati in altri contesti (Kahneman, Tversky, 1973; Lyon, Slovic, 1976).

2.4. *Implicazioni: mammografia e biopsia*

I problemi discussi fin qui possono avere importanti implicazioni pratiche. Basti pensare agli esempi che abbiamo citato, in cui due autori hanno fondato le loro conclusioni su ragionamenti probabilistici scorretti. Uno dei due ha erroneamente sostenuto che una donna con una massa che appare benigna sulla base dell'esame fisico e di una mammografia ha ancora il 20% di pro-

babilità di avere il cancro e ha quindi raccomandato che una paziente come questa sia comunque sottoposta a biopsia. L'altro autore ha basato la sua avversione allo screening mammografico su un grossolano errore di valutazione riguardo alla frequenza con la quale una donna riceverebbe una falsa rassicurazione (cioè la frequenza con la quale potrebbe avere un cancro non identificato attraverso una mammografia). Entrambi gli autori sarebbero forse giunti alle stesse conclusioni se avessero ragionato correttamente, ma è anche possibile che avrebbero tratto conclusioni opposte.

Il valore dell'informazione diagnostica. Il valore della mammografia in donne con segni e sintomi di una patologia al seno risiede nella sua capacità di fornire informazioni diagnostiche che potranno influire sulla decisione del medico di effettuare o non effettuare una biopsia. Più precisamente, l'esito del test dovrebbe modificare la stima di probabilità del medico riguardo all'ipotesi che la paziente abbia il cancro. Con le parole di altri due studiosi.

> La mammografia può assistere il medico nel distinguere fra lesioni benigne e maligne […] Alcune lesioni, specialmente di piccole dimensioni, possono non presentare le caratteristiche che suscitano nel medico un livello di sospetto sufficientemente alto per giustificare l'esecuzione di una biopsia. È in questa situazione che […] un mammogramma può fornire ulteriori informazioni oggettive. Nel caso di una lesione al seno di un tipo indeterminato, la mammografia può quindi aiutare il medico a decidere se approfondire l'indagine attraverso una biopsia (Clark, Robbins, 1965, p. 125).

Perché un qualsiasi test diagnostico sia utile, esso deve fornire informazioni che possono modificare le decisioni sul modo in cui trattare il paziente – nel nostro caso, sia suggerendo di eseguire una biopsia in pazienti alle quali altrimenti non sarebbe stata prescritta, sia (è auspicabile) permettendo di evitare l'esecuzione di una biopsia in alcune pazienti alle quali altrimenti sarebbe stata raccomandata. Questo approccio è stato sviluppato in termini formali nella teoria statistica della decisione, ed è stato impiegato per analizzare alcuni problemi medici in lavori di ricerca (Lusted *et al*., 1977).

Molti medici riconoscono che i risultati di esami radiografici contengono informazioni utili per la cura delle pazienti, ma solitamente non viene detto in che modo precisamente tali informazioni dovrebbero essere usate. Le spiegazioni offerte dalla maggior parte degli studiosi contengono ben poche indicazioni specifiche. «La mammografia non obbliga a particolari procedure di intervento ma, in certi casi, può fornire qualche precisa informazione aggiuntiva, in modo da evitare sviluppi indesiderabili» (Egan, 1972, p. 392). «La mammografia è un valido strumento aggiuntivo per il chirurgo nella diagnosi e nel trattamento di lesioni al seno» (Lyons, 1975, p. 231). «La mammografia può essere d'aiuto nel chiarire il significato di osservazioni che risultano con-

fuse attraverso la sola palpazione» (Egan, 1969, p. 146). Essa «svolge un ruolo ausiliario e di sostegno» (Block, Reynolds, 1974, p. 589). La precisa natura ed entità di tale sostegno è solitamente lasciata al giudizio del medico.

Mammografia e biopsia: le pratiche accettate. A quanto pare, il ruolo della mammografia in casi come quelli discussi è compreso solo in parte. Per rendercene conto, esaminiamo quale impatto questo esame dovrebbe avere, secondo gli studiosi, sull'esecuzione di biopsie in pazienti con patologie al seno. Mentre i passi citati in precedenza implicano che l'uso di esami radiografici dovrebbe aiutare a selezionare le pazienti per le quali la biopsia è appropriata, un ugual numero di asserzioni suggerisce che la mammografia non può, e persino non deve, svolgere questa funzione. «Qualsiasi lesione individuata attraverso la palpazione richiede una verifica tramite intervento chirurgico e biopsia, indipendentemente dai risultati di esami radiografici» (Lesnick, 1966, p. 2007). «Per quanto la mammografia sia solitamente definitiva, essa non sostituisce la biopsia» (Egan, 1969, p. 148). «Questa procedura non dovrebbe in nessun modo togliere importanza alla biopsia. In effetti, l'impiego regolare della mammografia riaffermerà l'importanza della biopsia, perché l'indicazione di una lesione maligna sulla base di un esame radiografico richiede una conferma attraverso la biopsia, […] e non diminuisce affatto l'importanza della biopsia […] La biopsia è tanto necessaria per la conferma degli esiti di esami radiografici quanto lo è per la conferma dei segni fisici» (Gershon-Cohen, Borden, 1964, pp. 2753, 2754). «È evidente che la mammografia non è un sostituto per il trattamento chirurgico» (DeLuca, 1974, p. 318). «Sottolineiamo con forza che la mammografia non è un sostituto della biopsia» (McClow, Williams, 1973, p. 618).

Una delle formulazioni più precise sul modo in cui la mammografia dovrebbe essere impiegata per aiutare a selezionare le pazienti che necessitano di una biopsia è apparsa sugli «Archives of Surgery» (1966, vol. 93, pp. 853-856). Un attento esame di questo insieme di indicazioni rivela che solo una parte delle potenzialità del test viene effettivamente sfruttata. Lo schema per l'uso della mammografia «al fine di determinare il trattamento o la classificazione di ciascuna paziente» prevede l'identificazione di tre categorie di pazienti.

Categoria A: «Le pazienti con un nodulo o una "lesione dominante" al seno sono casi di tipo principalmente chirurgico e l'esecuzione di una biopsia non dovrebbe essere rimandata. La mammografia, in questo caso, è strettamente *accessoria* […] Essa può rivelare tumori nascosti» (p. 854).

Categoria B: «Le pazienti che hanno sintomi riferibili al seno ma non masse discrete o "lesioni dominanti". […] In questa categoria, il chirurgo e il medico potranno trarre la maggiore utilità dall'impiego della mammografia, perché qui la modalità è di tipo *confermatorio*». In questi casi la mammogra-

fia fornirà conferme e incoraggiamento «se l'impressione clinica indica una lesione benigna. La mammografia non dovrebbe, comunque, distogliere il medico da una precedente opinione favorevole all'esecuzione di una biopsia» (p. 855).

Categoria C: Le pazienti in questa categoria non hanno segni né sintomi. Non ci sono quindi indicazioni cliniche per l'esecuzione di una biopsia, e un mammogramma può solo incrementare il numero di biopsie.

In questo modo viene delineato un programma di intervento che annulla il valore delle informazioni ricavate dalla mammografia nell'individuare le pazienti per le quali la biopsia può essere evitata. Solo l'elemento di informazione che implica la necessità della biopsia viene impiegato; le informazioni che potrebbero escludere l'esigenza della biopsia sono invece del tutto ignorate.

Mammografia e biopsia: le potenzialità. Per apprezzare in che modo i problemi del ragionamento probabilistico possono influenzare la reale offerta di cure mediche esaminiamo ora il ruolo che la mammografia potrebbe svolgere nella diagnosi differenziale e nella selezione delle pazienti che richiedono una biopsia. Come abbiamo visto, l'obiettivo del test è modificare la stima di chi decide riguardo alla probabilità che la paziente abbia il cancro. Se tale probabilità è sufficientemente alta (dal punto di vista del medico e della paziente), la biopsia deve essere raccomandata. Chiamiamo questo valore di probabilità la *soglia della biopsia*[4]. Ora consideriamo l'impatto del test sul trattamento di due gruppi di pazienti.

Il primo gruppo comprende pazienti riguardo alle quali il medico ritiene, sulla base della storia clinica e dell'esame fisico, che siano senza dubbio affette da cancro al seno. Servendosi di dati pubblicati da Friedman *et al.* (1966), si può stabilire che la probabilità iniziale (la frequenza) del cancro in questo gruppo è del 90%. Se venisse eseguita una mammografia su una paziente di questo tipo, un risultato positivo incrementerebbe forse la probabilità del cancro $p(\text{ca}|\text{pos})$ fino al 95%. Un mammogramma negativo lascerebbe comunque la probabilità del cancro al 71%. Tale elevata probabilità giustifica asserzioni come «se l'impressione soggettiva del medico dà sufficienti ragioni per sospettare un cancro, il medico dovrà comunque raccomandare l'ese-

4. Chiunque non sia persuaso dell'esistenza di una soglia della biopsia potrebbe, per convincersene, ragionare come segue. Possiamo essere d'accordo che nessuno è disposto a sottoporsi a una biopsia se la probabilità della presenza di un cancro è di uno su trenta trilioni? E possiamo essere d'accordo che chiunque vorrebbe confermare la diagnosi e ricevere il relativo trattamento se la probabilità è di 98 su 100? Se è così, allora in qualche valore di probabilità compreso fra uno su trenta trilioni e 98 su 100 chiunque ha una soglia della biopsia. Ovviamente, nel caso che una paziente rifiuti la biopsia e il trattamento anche quando la presenza di un cancro è certa, si può dire che non ha alcuna soglia della biopsia.

cuzione della biopsia» (Clark *et al.*, 1965, p. 133). Una probabilità del 71% di un tumore maligno è ancora alta a sufficienza perché praticamente chiunque preferisca sottoporsi a una biopsia.

Ora consideriamo un secondo gruppo di pazienti, vale a dire pazienti con una massa dominante al seno tale che non è comunque ovvio che si tratti di un cancro. In uno studio pubblicato, la probabilità che una massa di questo tipo fosse maligna è stata calcolata al 14% (Friedman *et al.*, 1966). In assenza di ulteriori informazioni, le linee-guida prevalenti prevedono in questi casi l'esecuzione di una biopsia: «se si sviluppa un nodulo dominante, dovrebbe essere rimosso e sottoposto ad analisi microscopica» (del Regato, 1970, p. 861). Partendo da questa indicazione, supponiamo che la soglia della biopsia per la paziente sia del 10%. Assumiamo cioè che se, in base alle migliori stime del medico, la probabilità che vi sia un cancro è superiore al 10% allora medico e paziente sono d'accordo che si dovrebbe effettuare una biopsia[5].

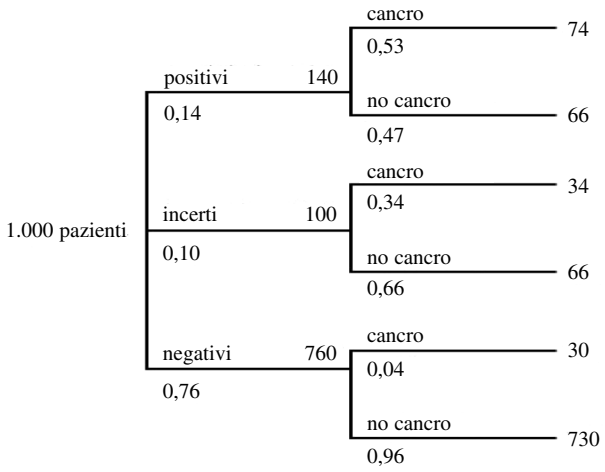
Usando una soglia del 10%, possiamo determinare l'impatto di un mammogramma sul trattamento di 1.000 pazienti appartenenti alla categoria considerata. Senza il test, tutte le pazienti verrebbero sottoposte a biopsia, 860 di loro senza conseguenze cliniche benefiche. Il destino approssimativamente

5. La soglia della biopsia è un numero molto importante e affascinante. Modificare tale soglia ha un impatto notevole sul numero di donne da sottoporre a biopsia, sulla frequenza di biopsie prive di benefici clinici, sui costi del trattamento delle pazienti e sulle stesse prognosi delle pazienti. A causa di una generale avversione al rischio, e per il fatto che stanno prendendo una decisione che riguarda un'altra persona, i medici generalmente fissano la soglia della biopsia a un livello piuttosto basso. Un'asserzione tipica è: «se c'è una *qualsiasi* possibilità che la lesione sia maligna, allora si dovrebbe fare una biopsia». «Se il medico non è pienamente convinto che la lesione sia benigna, allora si dovrebbe immediatamente eseguire una biopsia» (Allen, 1965, p. 640). L'evidenza disponibile suggerisce però che le pazienti fissano la soglia della biopsia a un livello generalmente più alto di quanto facciano i medici (anche se la variabilità è notevole).

A titolo di esempio, si possono esaminare i dati di un ampio studio clinico nel quale la mammografia e l'esame fisico sono stati impiegati per controllare la presenza di cancro al seno in donne asintomatiche (Shapiro, Strax, Venet, 1971). A seconda del modo in cui la lesione al seno era stata individuata (attraverso quale tipo di controllo o quale combinazione di diversi controlli), la probabilità che la paziente soffrisse di cancro al seno variava fra il 15% e il 54%. In base a un esame fisico positivo, i medici hanno raccomandato che su 545 donne con un mammogramma negativo venisse comunque eseguita una biopsia. Sebbene la frequenza del cancro in questo gruppo fosse del 15%, il 31% delle donne interessate ha declinato il suggerimento di sottoporsi a biopsia. La frequenza del cancro fra le donne con una mammografia positiva e un esame fisico negativo era del 20%, ma il 29% delle donne in questo gruppo ha declinato il suggerimento di sottoporsi a biopsia. Fra le donne con un risultato positivo in entrambi i controlli, la frequenza del cancro era del 54% e solo il 5% di queste donne ha preferito non sottoporsi a biopsia al momento in cui era stato raccomandato loro di farlo. Quindi, in base a queste cifre grezze, sembra che circa il 31% delle donne coinvolte avesse una soglia della biopsia superiore al 15%, che il 29% avesse una soglia della biopsia superiore al 20% e che il 5% avesse una soglia che superava il 54%.

prevedibile delle 1.000 pazienti con una lesione dominante nel caso che venga impiegata la mammografia è presentato in fig. 1[6].

Fig. 1 - Probabilità di cancro in donne con lesioni dominanti



6. Per delineare l'impatto della mammografia su queste pazienti (e su pazienti con altri segni e sintomi clinici) sono necessarie molte altre informazioni non direttamente disponibili in letteratura. È una fortuna che in uno degli studi pubblicati (Friedman *et al*., 1966) i dati relativi alla frequenza del cancro e all'accuratezza retrospettiva della mammografia siano presentati separatamente per questi tre distinti gruppi di pazienti – quelle con una forma palese di carcinoma, quelle con una massa dominante e quelle con altri segni e/o sintomi clinici di una patologia al seno. I dati pubblicati sono comunque incompleti e quelli relativi alla frequenza di una diagnosi radiografica incerta in casi di lesioni maligne e benigne non sono stati riportati. Per questo esempio ci siamo serviti dei dati dello studio di Friedman *et al*. (1966) facendo le seguenti assunzioni: (1) le lesioni non sottoposte a biopsia erano di fatto benigne; (2) le lesioni non sottoposte a biopsia sono state classificate come negative; (3) metà delle lesioni benigne che non sono state classificate come negative sono state classificate come positive (e le restanti sono state classificate come incerte); (4) metà delle lesioni maligne che non sono state classificate come positive sono state classificate come negative. I primi due punti rappresentano le assunzioni più ottimistiche riguardo all'accuratezza della mammografia. La terza e la quarta assunzione sono molto importanti. Infatti, al tendere a zero del tasso di falsi positivi (o di falsi negativi) la capacità di un risultato radiografico positivo (negativo) di confermare (escludere) la presenza del cancro aumenta. Analogamente, al crescere del tasso di falsi positivi o di falsi negativi, il test perde la sua capacità predittiva. L'interpretazione dei dati riportati da Friedman *et al*. (1966) è resa ancora più difficile dalla loro presentazione in termini di seni anziché di pazienti. Ciononostante, in questo studio si trovano molte informazioni utili ed è ragionevole utilizzarle in questo esempio, purché il lettore sia consapevole che si tratta di un'illustrazione e non di una vera e propria analisi formale. Un'analisi formale di questo problema richiederebbe dati di migliore qualità. Le cifre relative all'accuratezza usate nel testo per la valutazione delle pazienti nel secondo gruppo sono le seguenti: p(pos|ca) = 0,52; p(incerto|ca) = 0,24; p(neg|ca) = 0,24; p(pos|ben) = 0,075; p(incerto|ben) = 0,075; e p(neg|ben) = 0,85.

Le pazienti con un mammogramma positivo hanno una probabilità del 53% di avere il cancro e, poiché abbiamo fissato una soglia del 10%, esse dovrebbero essere sottoposte a biopsia. Poiché la probabilità che una paziente con un mammogramma dall'esito incerto abbia comunque il cancro è del 34%, anche queste pazienti dovrebbero ricevere una biopsia. Le pazienti con un mammogramma negativo hanno invece una probabilità del 4% di avere il cancro e, poiché tale cifra è inferiore alla soglia della biopsia che in base alle nostre assunzioni abbiamo attribuito loro (10%), esse non desidererebbero essere sottoposte a biopsia, ma preferirebbero che la loro situazione venisse seguita da vicino. Il numero totale di biopsie eseguite immediatamente risulta così ridotto da 1.000 a 240. Almeno altre 30 biopsie dovranno essere svolte in seguito, perché è previsto che 30 delle restanti 760 pazienti abbiano il cancro.

In questo modo, i benefici attesi della mammografia (come la riduzione da circa 86% al 13% del rischio di una biopsia non necessaria) possono essere confrontati con i costi (per esempio, il rischio dovuto alle radiazioni e una spesa di circa 75 dollari) e con la leggera diminuzione dell'aspettativa di sopravvivenza (c'è un 3% di probabilità che la diagnosi di una lesione maligna venga ritardata di un mese o di un periodo di tempo simile). Usando la nozione di «soglia della biopsia» e un po' di teoria della probabilità, a molte pazienti di questo gruppo con un mammogramma negativo verrebbe risparmiata una biopsia. Per contro, in assenza di questo tipo di analisi, «il consenso da un punto di vista chirurgico è che tutte la pazienti [di questo gruppo] dovrebbero essere sottoposte a biopsia, *indipendentemente dai risultati mammografici*» (Friedman *et al*., 1966, p. 889).

In questo esempio è opportuno sottolineare l'importanza della soglia della biopsia. Se il medico e la sua paziente avessero fissato tale soglia all'1% – cioè, se una probabilità di 1 su 100 che la paziente abbia il cancro fosse considerata sufficiente a giustificare l'esecuzione di una biopsia –, allora un mammogramma negativo non avrebbe eliminato la necessità della biopsia (perché una probabilità del 4% supererebbe quella soglia). Il mammogramma avrebbe fornito al medico delle informazioni, ma tali informazioni non avrebbero contribuito alla decisione di effettuare o non effettuare la biopsia. In questo caso, l'impiego della mammografia avrebbe dovuto essere giustificato su altre basi.

La revisione delle pratiche accettate. Questo tipo di analisi aiuta a chiarire le potenzialità della mammografia nella diagnosi differenziale di vari tipi di lesione e permette inoltre di valutare le seguenti indicazioni presenti in letteratura.

1. «La mammografia aggiunge ben poco al trattamento di noduli al seno rilevabili attraverso la palpazione, che, per le loro caratteristiche, richiedono una biopsia» (da «Archives of Surgery», 1974, 108, p. 589). Nel gruppo delle

pazienti con una massa dominante, abbiamo supposto che la biopsia venisse richiesta su basi puramente cliniche. L'uso della mammografia suddividerebbe questo gruppo in sottogruppi nei quali la frequenza del cancro varia dal 53% al 4%. Nel gruppo relativo a quest'ultima cifra la biopsia può essere evitata e il numero totale di biopsie potrebbe essere ridotto del 73% (da 1.000 su 1.000 a 270 su 1.000).

2. «Per gli scopi clinici, la mammografia dovrebbe offrire un'accuratezza del 100% prima che essa da sola possa orientare il trattamento» (da «Archives of Surgery», 1974, 108, p. 589). In una popolazione come quella del secondo gruppo descritto in precedenza, potrebbe essere decisamente razionale lasciare che siano i risultati della mammografia a guidare la selezione delle pazienti da sottoporre a biopsia. Si noti che il tasso di veri positivi usato in questo esempio era del 52%. Test più accurati potrebbero essere ancora più utili.

3. «La mammografia non sostituisce la biopsia» (da «Oncology», 23, p. 148). Il fine della mammografia, come della biopsia, è fornire informazioni riguardo alle condizioni di una paziente. Per alcune pazienti, in assenza di una mammografia, è richiesta l'esecuzione della biopsia. Ma per alcune di queste stesse pazienti un mammogramma negativo permetterebbe di evitare la biopsia. In questi casi, la mammografia sostituirebbe la biopsia.

4. «Ogni decisione di eseguire una biopsia dovrebbe essere preceduta da una mammografia» (da «Oncology», 1969, 23, p. 146). Si considerino i casi di cancro che appaiono ovvi su basi cliniche. La probabilità di cancro in questi casi supera la soglia praticamente di chiunque, non importa quale sia l'esito della mammografia. Un'indicazione come quella citata potrebbe essere giustificata soltanto facendo riferimento alla possibilità che la massa diagnosticata su basi cliniche sia in realtà benigna (altrimenti la paziente dovrebbe subire in ogni caso un intervento di asportazione) *e* alla possibilità che vi sia una lesione maligna nascosta e non rilevabile con la palpazione. La probabilità di questo evento congiunto è il prodotto delle probabilità dei due eventi distinti, ed è estremamente bassa (dell'ordine di 1 su 5.000).

5. «Rinviare la biopsia su una lesione che appare *benigna su basi cliniche* e che risulta *benigna in base a una mammografia* significa fare un passo indietro nell'eradicazione del carcinoma del seno» (da «Surgery, Gynecology and Obstetrics», 1972, 134, p. 98). Poniamo che «benigno su basi cliniche» significhi che *p*(ca) è uguale al 5% circa. Dopo una mammografia negativa, la probabilità che una paziente in questa situazione abbia il cancro è approssimativamente dell'1%. Su 100 biopsie, 99 non produrrebbero alcun beneficio clinico. In queste condizioni, il rinvio della biopsia è un passo indietro o un passo avanti? Inoltre, se venisse adottata la politica raccomandata nel testo citato, allora qualunque lesione – a partire da quelle «che appaiono benigne su basi cliniche» fino a quelle che su basi cliniche appaiono ovvi casi di can-

cro – richiederebbe una biopsia, quale che sia l'eventuale risultato del test mammografico. Il che sembra contraddire l'asserzione dello stesso autore secondo cui «se impiegata in una corretta prospettiva, la mammografia è un eccellente strumento aggiuntivo per il medico nel trattamento del carcinoma del seno» (*ibidem*).

6. «La mammografia non deve mai essere impiegata al posto della biopsia nel trattamento di una "lesione dominante" del seno e non dovrebbe mai modificare l'approccio fondamentale alle patologie del seno; in altre parole: "un nodulo è un nodulo" e deve essere sottoposto a biopsia o attraverso un'incisione o tramite aspirazione» (da «Archives of Surgery», 1966, 93, p. 854). Le pazienti con una lesione dominante e una soglia della biopsia superiore al 5% sarebbero in disaccordo con questa asserzione.

7. «La fallacia dipende dal fatto di affidarsi [alla mammografia] in casi dubbi. È essenziale, dopo lo svolgimento dell'esame fisico e della palpazione, decidere se si raccomanderebbe o meno una biopsia qualora non fossero disponibili esami radiografici. Se a questo punto sareste dell'idea di effettuare una biopsia, allora fatelo. Se siete sicuri che non ci sono le basi per effettuare una biopsia a seguito dell'esame fisico, allora prescrivete una mammografia. Non appena si dice a se stessi, e alla paziente, "non sono sicuro – facciamo un esame radiografico", allora, almeno inconsciamente, ci si impegna a fidarsi degli eventuali risultati negativi della mammografia, quando ci si dovrebbe basare soltanto sui risultati positivi. Si tratta di una trappola psicologica nella quale tutti tendono a cadere ed è un problema molto più serio di quello rappresentato da un certo numero di diagnosi falsamente positive ricavate dalla mammografia» (Rhoads, 1969, p. 1182). Neanche una singola biopsia verrebbe evitata applicando queste indicazioni. E sarebbe un peccato perché, con le parole dello stesso autore, «ci sono poche aree nelle quali, più che nel seno, risultano necessari così tanti interventi chirurgici che potrebbero essere evitati con metodi diagnostici più efficaci».

Siamo ora nella posizione di apprezzare la storia seguente, apparsa sul «Chronicle» di San Francisco (Kushner, 1976). Una reporter aveva appena scoperto una massa al seno e ha descritto un consulto con il suo medico.

> «Vorrei che si facesse fare una xeromammografia. È un nuovo modo per effettuare mammogrammi – immagini del seno.»
> «È accurato?»
> Il medico si stringe nelle spalle: «Probabilmente tanto accurato quanto può esserlo un'immagine. Sa» avverte «anche se il risultato è negativo – che significa che il nodulo non è maligno – l'unico modo di essere sicuri è tagliare quello che c'è, e guardarlo al miscoscopio.»

La paziente a questo punto discute la questione con il marito.

«Che cosa ha detto il dottore?»
«Vuole che faccia uno xeromammografia. Poi, quale che sia il risultato, bisognerà togliere il nodulo.»
«Ma allora perché fare l'esame radiografico?»
«Immagino che sia da fare. E il dottore dice che è esatto nell'85% circa dei casi… Quindi, per prima cosa ho fissato un appuntamento per fare un termogramma. Che sia positivo o negativo, se è in accordo con l'immagine xeromammografica le statistiche dicono che la diagnosi è affidabile al 95%.»

Riassumendo, se l'obiettivo della mammografia è quello di aiutare il medico a distinguere patologie benigne e maligne al seno, risparmiando con ciò ad alcune pazienti una procedura più invasiva e traumatica come la biopsia, allora dovremmo lasciare che il test svolga questa funzione. Se d'altra parte si ritiene che il medico debba sempre mantenersi fedele a una decisione precedente riguardo alla biopsia e attenersi a tale decisione di fronte a qualunque esito mammografico, allora non si dovrebbe sostenere che il fine del test è di contribuire a distinguere fra patologie benigne e maligne, perché sarà sempre e solo la biopsia a determinare tale distinzione in modo definitivo. Infine, se l'obiettivo del test è quello di scoprire carcinomi nascosti in altre aree del seno e non identificabili su basi cliniche (distinti quindi da una massa evidente alla palpazione che richiederebbe comunque una biopsia), si dovrebbe allora riconoscere esplicitamente che la probabilità di un evento del genere è estremamente bassa e che questo uso del test equivale a un intervento di screening.

Con l'analisi qui presentata non si intende argomentare in favore di una specifica linea di intervento nell'uso della mammografia o della biopsia – per fare ciò sarebbero necessari dati migliori e una migliore conoscenza dei valori delle pazienti. L'obiettivo è piuttosto suggerire che non è stato sviluppato un approccio formale al ragionamento probabilistico su questo tipo di problemi, che il giudizio clinico può essere fallace e che le attuali politiche di intervento potrebbero essere incoerenti o scorrette.

3. Conclusioni

Gli esempi precedenti sono stati presentati per illustrare la complessità delle decisioni mediche e per mostrare come alcuni medici affrontano un aspetto di tale complessità: la valutazione delle probabilità. Il caso che abbiamo discusso è relativamente semplice, dal momento che consiste nell'uso di un singolo test diagnostico al fine di classificare le lesioni al seno in due gruppi, quelle benigne e quelle maligne. La base di dati a disposizione su questo problema è piuttosto buona. L'accuratezza e il valore diagnostico del test sono stati studiati e analizzati in numerose istituzioni e per molti anni. Con le pa-

role di un ricercatore interessato al problema, «non conosco alcuna procedura medica che è stata controllata e ricontrollata più a fondo della mammografia» (Egan, 1971, p. 1555).

Gli strumenti probabilistici discussi in questo capitolo sono disponibili da secoli. Negli ultimi decenni sono stati applicati in misura crescente ai problemi medici (Lusted, 1968) e l'impiego di metodi formali per affrontare l'incertezza si è recentemente diffuso nei curricula delle Facoltà di Medicina, negli articoli specialistici e nei programmi di formazione postuniversitaria. A oggi, però, l'applicazione di queste tecniche è rimasta sporadica e non è penetrata nelle abitudini di pensiero di molti medici. Come si è mostrato nel presente studio, i problemi medici sono complessi e il ragionamento probabilistico con strumenti formali offre grandi opportunità per migliorare la qualità e l'efficacia delle cure mediche.

2. *Giudizi di probabilità in medicina: trascurare le possibilità implicite**

di *Donald A. Redelmeier, Derek J. Koehler, Varda Liberman, Amos Tversky*

1. Introduzione

Le decisioni mediche sono spesso prese in condizioni di incertezza. Nel valutare il caso di un paziente con dolori al petto, per esempio, un medico deve considerare la possibilità che il paziente abbia un infarto cardiaco, il rischio di una grave emorragia a seguito dell'eventuale somministrazione di trombolitici, e le possibili conseguenze nel caso che i trombolitici non vengano somministrati. L'incertezza può talvolta essere ridotta raccogliendo dati aggiuntivi, consultando la letteratura scientifica e rivolgendosi ad altri esperti. Ma non sempre può essere eliminata entro limiti di tempo ristretti (Pauker, Kopelman, 1992). Pertanto l'azione spesso dipende da giudizi intuitivi sulla probabilità di diversi eventi possibili.

La ricerca sul giudizio in condizioni di incertezza ha mostrato che tanto le persone comuni quanto gli esperti non sempre rispettano i principi della teoria della probabilità (Kahneman, Slovic, Tversky, 1982; von Winterfeldt, Edwards, 1986)[1]. In particolare, rappresentazioni alternative di uno stesso stato di cose possono dar luogo a giudizi di probabilità differenti (Tversky, Koehler, 1994). Per rendere conto di queste osservazioni, Tversky e Koehler hanno sviluppato un approccio in cui le probabilità non vengono associate a eventi – come accade in altri modelli – ma piuttosto a descrizioni di eventi, dette «ipotesi» (Tversky, Koehler, 1994). Questo approccio, chiamato *teoria del supporto* (*support theory*), prevede che ogni ipotesi (in questo senso tecnico) si riferisca a un unico evento, ma che un dato evento possa essere descritto da più di un'ipotesi. Per esempio, l'ipotesi esplicita «morte dovuta a incidente stradale, annegamento, scossa elettrica, o qualsiasi altro incidente»

* Traduzione di *Probability judgment in medicine: discounting unspecified possibilities*, «Medical decision making», 15 (1995), pp. 227-230. Per gentile concessione dell'editore.

1. Si vedano anche, in questo volume, i capp. 1, 3 e 12 [N.d.C.].

e l'ipotesi implicita «morte dovuta a un incidente» rappresentano descrizioni differenti dello stesso evento.

L'assunto centrale della teoria del supporto è il *principio di spacchettamento* (*unpacking principle*): fornire una descrizione più dettagliata di un'ipotesi implicita generalmente fa aumentare il corrispondente giudizio di probabilità. Per esempio, la probabilità stimata della descrizione esplicita che elenca vari tipi di incidente solitamente supera la stima di probabilità della descrizione implicita che non fa menzione di specifici tipi di incidente. Come la lunghezza di una costa, che aumenta al crescere del livello di dettaglio della mappa, la probabilità percepita di un evento aumenta se la sua descrizione si fa più specifica. Sia la memoria sia l'attenzione contribuiscono a questo fenomeno: lo «spacchettamento» di un'ipotesi può ricordare alle persone possibilità che altrimenti avrebbero potuto trascurare e la menzione esplicita di una specifica possibilità può renderla più saliente e quindi aumentare la sua probabilità percepita.

In accordo con la teoria classica della probabilità, la teoria del supporto assume che la probabilità stimata di un'ipotesi e quella dell'ipotesi complementare sommino all'unità. Per esempio, la probabilità stimata dell'ipotesi «morte dovuta a cause naturali» e quella dell'ipotesi «morte dovuta a cause non naturali» dovrebbero sommare a 1 (secondo un principio detto di *complementarità binaria*). Tuttavia, ciascuna delle due stime di probabilità può aumentare nel caso in cui si «spacchetti» la rispettiva categoria di eventi. L'effetto di spacchettamento, così come la complementarità binaria, è stato osservato in diversi esperimenti relativi a problemi non medici (Tversky, Koehler, 1994). Il presente lavoro esplora questi principi nel contesto dei giudizi medici. A questo scopo, abbiamo presentato a diversi gruppi di medici alcuni brevi scenari che descrivevano casi individuali e abbiamo chiesto loro di giudicare la probabilità di diversi eventi clinicamente rilevanti.

2. Spacchettare una categoria residua

In un'indagine svolta presso l'Università di Stanford con 59 specializzandi fino a cinque anni dalla laurea in medicina, si chiedeva ai partecipanti di leggere il seguente scenario.

> Una nota attrice ventiduenne si presenta al pronto soccorso con un dolore al quadrante inferiore destro dell'addome che perdura da dodici ore. L'ultimo periodo mestruale è stato quattro settimane fa.

A metà dei medici, selezionati casualmente, si chiedeva di stimare la probabilità di due distinte diagnosi (gastroenterite e gravidanza ectopica) e di

una categoria residua («nessuna delle precedenti»). Agli altri medici veniva chiesto di stimare la probabilità delle seguenti ipotesi diagnostiche: le due diagnosi appena viste (gastroenterite e gravidanza ectopica), tre ulteriori diagnosi specifiche (appendicite, pielonefrite e infiammazione pelvica) e una categoria residua («nessuna delle precedenti»). I due compiti differivano soltanto per il fatto che la categoria residua nella prima lista (più breve) risultava parzialmente spacchettata nella seconda lista (più lunga). A tutti i medici veniva detto che la paziente soffriva di una sola patologia e che quindi le stime di probabilità avrebbero dovuto sommare a 100%.

Da un punto di vista logico, la probabilità della categoria residua «nessuna delle precedenti» nella lista breve dovrebbe essere equivalente alla somma delle probabilità delle possibilità corrispondenti nella lista estesa. Al contrario, in accordo con il principio di spacchettamento, abbiamo osservato che in media la probabilità assegnata alla categoria residua nella lista breve era inferiore alla somma delle probabilità corrispondenti nella lista estesa (50% *vs.* 69%; $p < 0{,}005$; test Mann-Whitney). Lo spacchettamento della categoria residua ha quindi alterato le probabilità assegnate a specifiche diagnosi. Per esempio, la probabilità assegnata in media alla diagnosi di gastroenterite è stata notevolmente più alta nella lista breve che non in quella estesa (31% *vs.* 16%; $p < 0{,}005$; test Mann-Whitney). Evidentemente, spacchettare un'ipotesi residuale ha ricordato ai medici possibilità diagnostiche che altrimenti avrebbero potuto trascurare o ha reso più salienti diagnosi alle quali altrimenti avrebbero dato minor peso.

3. Evidenziare una possibilità

Nell'esempio appena discusso ai medici veniva chiesto di assegnare delle probabilità a un insieme di possibilità alternative. Spesso, però, i medici si concentrano su una singola possibilità. In questi casi, essi potrebbero sovrastimare la probabilità di quella particolare possibilità per il fatto che le alternative restano implicite. Per illustrare questo punto, abbiamo presentato a un gruppo di 52 medici esperti dell'Università di Tel Aviv il seguente scenario.

> R.G. è un contadino in pensione di 67 anni, che si presenta al pronto soccorso con un dolore al petto che perdura da quattro ore. La diagnosi è infarto cardiaco acuto. L'esame fisico non mostra segni di edema polmonare, ipotensione o alterazione dello stato mentale. L'elettrocardiagramma mostra sovraslivellamento del tratto ST nelle derivazioni anteriori, ma nessuna disritmia o arresto cardiaco. La storia clinica non presenta elementi di rilievo. Il paziente viene ammesso in ospedale e sottoposto al trattamento di routine. Considera i possibili esiti.

A ognuno dei medici è stato assegnato casualmente il compito di valutare una sola delle seguenti quattro prognosi per questo paziente: «morte durante questo ricovero», «sopravvivenza a questo ricovero, ma morte entro un anno», «sopravvivenza per più di un anno, ma meno di dieci» o «sopravvivenza per più di dieci anni». Le probabilità medie assegnate a queste prognosi sono state rispettivamente 14%, 26%, 55% e 69%. Secondo la teoria classica, le probabilità assegnate a questi esiti fra loro alternativi dovrebbero sommare al 100%. Al contrario, le stime medie danno come somma 164% (intervallo di confidenza al 95%: 134%-194%). Come implicato dal principio di spacchettamento, i medici in ciascuno dei quattro gruppi hanno sovrastimato la possibilità che veniva esplicitamente menzionata in confronto alle alternative che non venivano specificate. Va notato, inoltre, che in tutti e quattro i gruppi si sono ottenute stime medie che superavano le frequenze dei rispettivi esiti riportate nella letteratura scientifica (Goldberg *et al.*, 1991). Si osservi che, mentre il risultato del problema precedente può essere interpretato come un effetto relativo alla memoria (ai medici vengono ricordate possibilità aggiuntive), questo risultato rappresenta un effetto relativo all'attenzione (un particolare intervallo all'interno di un continuum risulta evidenziato).

4. Complementarità binaria

Abbiamo attribuito i risultati precedenti al principio di spacchettamento. Una interpretazione alternativa è che le persone sovrastimino l'ipotesi focale che viene loro chiesto di valutare. Se questa interpretazione fosse corretta, la somma delle probabilità stimate di una coppia di ipotesi complementari dovrebbe eccedere l'unità. Per controllare questa previsione, abbiamo presentato lo scenario precedente a 149 studenti di medicina dell'Università di Toronto al quarto anno di corso. A metà dei partecipanti, selezionati casualmente, veniva chiesto di valutare la probabilità che il paziente sarebbe «morto durante questo ricovero». All'altra metà veniva chiesto di valutare la probabilità che il paziente sarebbe «sopravvissuto a questo ricovero». Abbiamo osservato che le stime di probabilità medie nei due gruppi erano, rispettivamente, 78% e 21%, la cui somma dà 99% (intervallo di confidenza al 95%: 94%-104%). Come la teoria del supporto prevede, le stime di probabilità sommano a 100% con due sole possibilità, ma superano il 100% in presenza di più di due possibilità. Questa osservazione mostra che le persone sovrastimano ciò che viene specificato più dettagliatamente, non ciò che si chiede loro di valutare.

5. Scelte di trattamento

Un ultimo esempio mostra che l'effetto di spacchettamento non è limitato a giudizi di probabilità, ma può estendersi alle scelte di trattamento. Abbiamo chiesto a 148 studenti di medicina dell'Università di Toronto al quarto anno di corso di considerare il seguente scenario.

> M.S. è una giornalista di quarantatre anni che si presenta al pronto soccorso con febbre e mal di testa che durano da due giorni. Nella storia clinica l'unico dato rilevante è la presenza per quindici anni di *lupus erythematosus*, controllato con Tylenol e trattamento steroideo cronico (prednisone, 10 mg al giorno). La paziente non appare malata. I segni vitali sono normali. L'esame fisico rivela irritabilità del seno paranasale frontale ed eritema faringeo. Non vi sono rigidità del collo, arrossamento della membrana timpanica o adenopatia cervicale. Gli altri esiti dell'esame obiettivo non sono significativi, a parte alcune modificazioni degenerative delle giunture più piccole di entrambe le mani.

Per metà degli studenti, selezionati casualmente, lo scenario era seguito dalla frase: «Chiaramente, molte diagnosi sono possibili sulla base di questo limitato insieme di informazioni, comprese vasculite, *lupus cerebritis*, infezioni opportunistiche intracraniche, sinusite, ed ematoma subdurale». All'altra metà dei partecipanti veniva invece presentata una frase più breve: «Chiaramente, molte diagnosi sono possibili sulla base di questo limitato insieme di informazioni, compresa quella di sinusite». Ai soggetti di entrambi i gruppi si chiedeva di indicare se avrebbero raccomandato l'esecuzione di una TAC alla testa.

Da un punto di vista logico, non dovrebbe esserci alcuna differenza fra le risposte nelle due versioni, perché entrambe descrivono la stessa situazione. Per contro, la teoria del supporto suggerisce che la possibilità della sinusite influirà maggiormente quando è l'unica possibilità diagnostica che viene specificata. Ci aspetteremmo pertanto che un minor numero di medici ordini una TAC nella versione più breve, perché per una diagnosi di sinusite non è solitamente richiesto questo esame (Hourihane, 1994). In effetti, abbiamo osservato che un numero minore di partecipanti ha suggerito l'esecuzione della TAC in risposta alla versione più breve che in risposta a quella più estesa (20% *vs.* 32%; $p < 0{,}05$; test Mann-Whitney). Il principio di spacchettamento si applica quindi anche all'indicazione di interventi clinici, e non solo ai giudizi di probabilità.

6. Conclusioni

La valutazione soggettiva di eventi incerti è talvolta necessaria e spesso è fallibile. In questo studio ci siamo concentrati su una fonte di errore partico-

larmente rilevante, cioè la tendenza a trascurare possibilità implicite. Nel primo problema abbiamo documentato l'effetto di spacchettamento in un compito diagnostico ricordando ai medici possibilità che avrebbero potuto trascurare. Nel secondo problema abbiamo ottenuto lo stesso effetto in un compito prognostico mettendo in evidenza uno specifico intervallo all'interno di un continuum. Nel terzo problema abbiamo mostrato che l'effetto di spacchettamento non può essere spiegato facendo riferimento a una sovrastima dell'ipotesi focale. E nell'ultimo problema abbiamo mostrato l'effetto di spacchettamento in un compito di decisione. Nel loro insieme, questi risultati confermano la principale predizione qualitativa della teoria del supporto nel contesto del giudizio medico.

Si potrebbe ipotizzare che i soggetti coinvolti in questo studio abbiano ritenuto che la richiesta di valutare una specifica ipotesi trasmettesse informazioni rilevanti e suggerisse, in particolare, che l'ipotesi in questione non era improbabile. Sebbene una credenza di questo tipo possa contribuire all'effetto di spacchettamento, essa non spiega pienamente i risultati ottenuti. Innanzi tutto, tale spiegazione implica la sopravvalutazione dell'ipotesi focale, in contrasto con l'osservazione che la complementarità binaria è rispettata. Inoltre, l'effetto di spacchettamento è stato notevole nell'esempio dell'infarto cardiaco, nel quale gli esperti erano informati del fatto che altri medici stavano valutando ipotesi differenti. Infine, l'effetto di spacchettamento è stato osservato anche in contesti non medici nei quali i soggetti erano al corrente che l'ipotesi focale era stata selezionata casualmente (Tversky, Koehler, 1994).

Sebbene non vi sia alcun metodo semplice e immediato per eliminare l'effetto di spacchettamento, vorremmo richiamare l'attenzione sulla sua presenza e suggerire alcune possibili procedure correttive. In primo luogo, è opportuno che i medici prendano atto che i giudizi in condizioni di incertezza sono soggetti a errori; in particolare, descrizioni alternative della stessa situazione possono condurre a giudizi differenti. In secondo luogo, i medici dovrebbero essere incoraggiati a spacchettare ampie categorie diagnostiche e a confrontare fra loro possibilità diverse ad analoghi livelli di specificità piuttosto che confrontare una singola possibilità specifica con un insieme di alternative non specificate. Lo spacchettamento del complemento implicito dell'ipotesi focale può in effetti rappresentare un metodo efficace per ridurre il rischio di sopravvalutare la probabilità della propria ipotesi corrente (*overconfidence*). Più in generale, una migliore comprensione della dimensione cognitiva sottesa al giudizio medico può aiutare a identificare comuni fonti di errore e suggerire strategie correttive.

3. *La raccolta e l'integrazione di informazioni come fonti di errore nelle decisioni diagnostiche**

di *Larry D. Gruppen, Fredric M. Wolf, John E. Billi*

1. Introduzione

Le ipotesi diagnostiche svolgono un ruolo centrale in molti compiti di «soluzione di problemi» in ambito medico. Spesso i medici considerano diverse ipotesi diagnostiche come possibili spiegazioni dei disturbi di un paziente; quindi si mettono alla ricerca di informazioni che permettano di identificare una particolare diagnosi come corretta o come un'utile ipotesi di lavoro da cui partire. Un aspetto caratteristico di queste ipotesi è l'incertezza. Nella maggior parte dei problemi diagnostici, i medici hanno a che fare con un margine di incertezza riguardo alla validità e all'accuratezza delle ipotesi considerate. L'incertezza propria delle diverse ipotesi aumenta o diminuisce via via che il medico raccoglie informazioni che sostengono una specifica ipotesi a discapito di altre. L'obiettivo del medico di fronte a un problema diagnostico può essere visto come il raggiungimento di un livello accettabile di fiducia riguardo alla validità di una particolare ipotesi nell'insieme di quelle pertinenti (Allman *et al*., 1985; Pauker, Kassirer, 1980).

Solitamente i medici giungono a questo livello di fiducia raccogliendo dati clinici al fine di valutare le diverse ipotesi considerate. Questo processo di raccolta delle informazioni distingue i problemi medici da quelli coinvolti in molti altri contesti, come la fisica, la matematica o gli scacchi. In questi ambiti, chi è chiamato a risolvere un problema di solito ha già a disposizione le informazioni necessarie e rilevanti e il suo compito è quello di integrare tali informazioni in modo appropriato per determinare la soluzione. In medicina è raro che le informazioni necessarie e rilevanti siano disponibili fin dall'inizio. Al contrario, è comune che i medici debbano avviare una ricerca di

* Traduzione di *Information gathering and integration as sources of error in diagnostic decisions*, «Medical decision making», 11 (1991), pp. 233-239. Per gentile concessione dell'editore.

informazioni il più possibile tempestiva ed efficiente, dal momento che a tale raccolta di informazioni sono normalmente associati dei costi.

Numerosi studi hanno esaminato il modo in cui i medici e le persone non esperte trattano le informazioni in compiti di questo genere e hanno identificato una serie di problemi. Fra i limiti e gli errori che sono stati rilevati rientrano: la tendenza a sopravvalutare la qualità delle proprie conoscenze e decisioni (Einhorn, Hogarth, 1978; Voytovich, Rippey, Suffredini, 1985; Arkes *et al.*, 1987); il «conservatorismo» nell'aggiornamento delle stime di probabilità (Casscells, Shoenberger, Graboys, 1978; Christensen-Szalanski, Bushyhead, 1983); la tendenza a ignorare informazioni relative al tasso di base (*base-rate*), o prevalenza, di una data condizione (Christensen-Szalanski, Beach, 1982; Bar-Hillel, 1980 e 1983; Baron, Hershey, 1988b)[1]; stime di probabilità inaccurate (Casscells, Shoenberger, Graboys, 1978; Dawson, Arkes, 1987; Hogarth, 1975; Jones, Howard, Haley, 1984; Tversky, Kahneman, 1981)[2]; la tendenza a selezionare informazioni che confermano le proprie aspettative a discapito di informazioni che si trovano in contrasto con esse (Einhorn, Hogarth, 1978; Doherty *et al.*, 1979; Baron, Beattie, Hershey, 1988); e la tendenza a giudicare in modo distorto il valore o il significato delle informazioni (Doherty *et al.*, 1979; Doherty *et al.*, 1981; Kern, Doherty, 1982; Wolf, Gruppen, Billi, 1985)[3].

Sebbene questi fenomeni non conducano necessariamente a decisioni erronee, è comunque importante approfondire la conoscenza della loro diffusione e della loro natura. Uno dei problemi aperti in tale tentativo di comprensione consiste nell'individuare *in quale fase* del processo decisionale si manifesti una tendenza sistematica all'errore. Il processo decisionale può infatti essere suddiviso in diverse componenti: la generazione di possibili ipotesi; la raccolta di informazioni; l'organizzazione e l'integrazione delle informazioni raccolte; la valutazione di diverse alternative alla luce delle informazioni disponibili. Queste diverse componenti, generalmente, contribuiscono al processo decisionale in maniera sequenziale e ciclica, in modo che l'esito di ogni passo rappresenti il punto di partenza di quello successivo. Per esempio, è tipico che le prime ipotesi generate guidino la raccolta e l'interpretazione delle informazioni (Elstein, Shulman, Sprafka, 1978); le informazioni raccolte rappresentano la base per un successivo processo di integrazione e di impiego delle informazioni stesse; il modo in cui le informazioni sono integrate e organizzate influenza il modo in cui verranno valutate, e gli esiti di tale valutazione possono modificare le ipotesi di partenza, orientando una nuova raccolta di informazioni in un ulteriore ciclo del processo.

1. Si vedano anche, in questo volume, i capp. 1 e 12 [N.d.C.].
2. Si veda anche, in questo volume, il cap. 4 [N.d.C.].
3. Si veda anche, in questo volume, il cap. 5 [N.d.C.].

Dal momento che ogni passo dipende dal compimento di quello precedente, la «catena» del processo decisionale può essere al massimo tanto forte quanto lo è il suo anello più debole. È pertanto importante distinguere le possibili fonti di errore, nel tentativo di migliorare le decisioni mediche. I provvedimenti più opportuni per il miglioramento delle decisioni attraverso la formazione o l'impiego di strumenti di sostegno possono risultare notevolmente diversi a seconda che gli errori decisionali siano da attribuire prevalentemente a una delle componenti del processo (come la raccolta di informazioni) oppure a un'altra (come l'integrazione delle informazioni raccolte) (Beyth-Marom, Fischhoff, 1983).

Nei due esperimenti descritti in questo articolo abbiamo tentato di distinguere fra loro due possibili fonti di errore nella decisioni diagnostiche: la raccolta delle informazioni e l'integrazione delle informazioni disponibili. Se le informazioni vengono *raccolte* in modo manchevole o inefficace, anche una *integrazione* ottimale di quelle informazioni non potrà produrre decisioni appropriate. Inoltre, a una raccolta ottimale delle informazioni può seguire un processo di integrazione insufficiente o distorto, con risultati altrettanto insoddisfacenti. La decisione finale può quindi essere inadeguata in conseguenza di limiti propri di una o dell'altra di queste componenti (o di entrambe).

Negli esperimenti qui presentati abbiamo studiato due condizioni distinte: una condizione di *informazione parziale*, nella quale i medici coinvolti raccoglievano un insieme limitato di informazioni e in seguito le integravano per giungere a una decisione; una condizione di *informazione completa*, nella quale i medici ricevevano l'insieme completo delle informazioni diagnostiche ed erano chiamati a integrare i dati disponibili in vista della decisione. Le domande di nostro interesse riguardavano il modo in cui le due diverse condizioni influivano 1) sulle informazioni che i medici consideravano più importanti in confronto a quelle indicate come tali dai pertinenti principi normativi di ragionamento (in particolare, il teorema di Bayes), 2) sulla fiducia dei medici riguardo all'appropriatezza della raccolta di informazioni da loro svolta, e 3) sull'accuratezza delle diagnosi dei medici.

2. Esperimento I

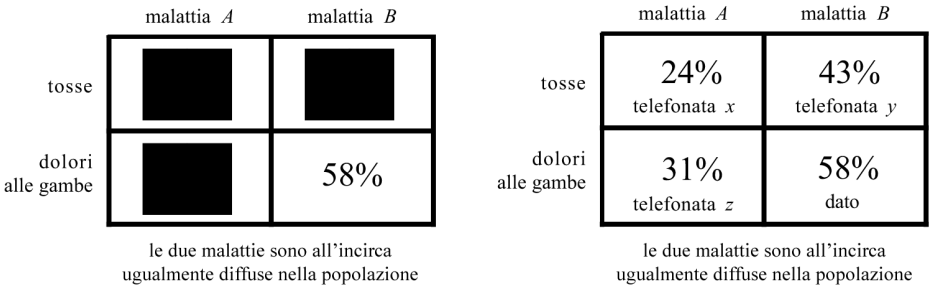
2.1. *Metodi*

Il problema sperimentale. Il problema impiegato in questo studio consiste in un compito diagnostico semplificato in cui si tratta di decidere fra due ipotesi alternative sulla base delle relazioni probabilistiche fra un sintomo e ciascuna delle due condizioni cliniche considerate. La struttura del problema è modellata su quella impiegata da Doherty *et al.* (1979) e (1981) in contesti

non medici. Un paragrafo introduttivo chiedeva ai partecipanti di immaginare di praticare la professione in un'area geografica remota. Veniva quindi descritto il caso di un paziente che manifestava due sintomi, entrambi compatibili con due diverse ipotesi diagnostiche relative a due malattie ugualmente diffuse nella popolazione (la prevalenza, o probabilità a priori, delle due condizioni era pertanto la stessa). Ai partecipanti veniva quindi detto che avrebbero potuto ottenere ulteriori informazioni attraverso una sola «telefonata» a una clinica specialistica lontana dalla loro zona di attività.

Le informazioni cliniche venivano presentate in una matrice 2 × 2 (fig. 1). Ogni cella della matrice conteneva la probabilità condizionata della presenza di uno specifico sintomo data la presenza di una particolare condizione patologica: la malattia *A* (M_A) oppure la malattia *B* (M_B). Le due diagnosi sono state presentate in termini astratti (*A* e *B*) per minimizzare l'influenza delle particolari conoscenze dei soggetti e della loro precedente esperienza clinica. Una delle probabilità condizionate (la probabilità che un paziente soffra di dolore alle gambe se affetto dalla malattia *B*) veniva fornita fin dall'inizio. Ai partecipanti si chiedeva quindi di indicare *un* elemento aggiuntivo di informazione che avrebbero voluto ricevere attraverso la loro «telefonata».

Fig. 1 - Le due versioni del problema presentato nell'Esperimento I. La matrice a sinistra illustra il problema nella condizione di informazione parziale, nella quale si chiedeva ai medici di selezionare una sola delle celle con i valori di probabilità nascosti rimuovendo un adesivo opaco. La matrice a destra illustra il problema nella condizione di informazione completa, nella quale si chiedeva ai medici di indicare quale «telefonata» avrebbero fatto, dovendo scegliere quella più importante



La soluzione ottimale per questo tipo di problema è data dall'applicazione del teorema di Bayes (eq. 1), che permette di calcolare la probabilità a posteriori delle due diagnosi sulla base del prodotto fra le loro probabilità a priori (la prevalenza di ciascuna delle due) e le probabilità condizionate di un particolare sintomo data la presenza di ciascuna delle malattie.

$$\frac{p(M_A|S)}{p(M_B|S)} = \frac{p(S|M_A)}{p(S|M_B)} \times \frac{p(M_A)}{p(M_B)} \qquad (Eq.\ 1)$$

Più precisamente, il risultato (a sinistra) della eq. 1, $p(M_A|S)/p(M_B|S)$, rappresenta il rapporto fra le probabilità a posteriori delle due malattie alla luce della presenza di un sintomo S (per esempio: dolori alle gambe). Questo rapporto equivale al prodotto del rapporto fra le probabilità a priori delle due diagnosi, $p(M_A)/p(M_B)$ (cioè il rapporto fra la prevalenza dell'una e dell'altra malattia), e del cosiddetto «rapporto fra le verisimiglianze», $p(S|M_A)/p(S|M_B)$ (cioè il rapporto fra le probabilità che si dia il sintomo quando, rispettivamente, l'una o l'altra fra le due malattie è presente).

Il teorema di Bayes ha due implicazioni rilevanti per la soluzione di problemi come quello impiegato in questo studio: la prima riguarda la raccolta delle informazioni, la seconda l'integrazione delle informazioni. In base al teorema di Bayes, nella fig. 1 l'informazione più utile da ottenere è la probabilità che si manifestino dolori alle gambe in un paziente che soffre della malattia A (la cella in basso a sinistra nella matrice). Infatti, dal momento che è già disponibile l'informazione relativa alla probabilità che si manifestino dolori alle gambe in un paziente che soffre della malattia B, la cifra nella casella in basso a sinistra è precisamente l'elemento di informazione richiesto per poter applicare il teorema di Bayes. Nella formulazione del problema sperimentale veniva specificato che la prevalenza delle due malattie era la stessa; perciò il termine $p(M_A)/p(M_B)$ equivale a 1, e non influisce in alcun modo sul rapporto fra le probabilità a posteriori. Un elemento di informazione relativo al sintomo della tosse (collocato quindi in una delle due celle in alto) non potrebbe confluire nell'eq. 1 e pertanto, dati i vincoli imposti dalla definizione del problema sperimentale, non avrebbe alcun valore diagnostico. Informazioni di questo tipo sono dette «pseudodiagnostiche» (Doherty *et al.*, 1979 e 1981; Kern, Doherty, 1982; Wolf, Gruppen, Billi, 1985 e 1988).

Per quanto riguarda l'*integrazione* delle informazioni, le probabilità condizionate rilevanti devono essere combinate con le informazioni relative alla diffusione delle due malattie (le probabilità a priori) in modo da determinare le probabilità a posteriori delle due diagnosi. Se il valore di probabilità necessario, cioè p(dolori alle gambe|malattia A), non è disponibile, allora non è possibile calcolare il rapporto fra le probabilità a posteriori attraverso il teorema di Bayes. In questo caso, la decisione diagnostica più appropriata dovrebbe essere basata unicamente sulla diffusione delle due malattie. Si noti però che, anche quando l'elemento di informazione necessario per applicare il teorema di Bayes è disponibile, è ancora possibile che il processo di integrazione delle informazioni si discosti dal teorema di Bayes (come accade nel caso in cui la diffusione delle due malattie venga ignorata, o in cui le pro-

babilità a posteriori vengano determinate impiegando in modo inappropriato le probabilità condizionate note).

Partecipanti e procedura. È stato impiegato un schema sperimentale con misurazioni ripetute delle variabili di interesse *entro* i soggetti (*within subjects*). I compiti sperimentali sono stati assegnati a 120 medici che frequentavano un corso-base obbligatorio di soccorso in condizioni di emergenza durante il loro primo anno di specializzazione presso la Facoltà di Medicina dell'Università del Michigan. I partecipanti ricevevano una versione del problema nella condizione di informazione parziale e una versione del problema nella condizione di informazione completa (oltre ad altri due problemi estranei al presente studio). Nella condizione di informazione parziale la probabilità che un paziente accusi dolori alle gambe se affetto dalla malattia *B* (58%) era disponibile fin dall'inizio (si veda il lato sinistro in fig. 1). Il primo compito assegnato ai medici era quello di fare la loro «telefonata» alla clinica specialistica (sollevando un adesivo opaco) per ottenere *uno solo* dei restanti valori di probabilità. Le tre alternative fra cui scegliere consistevano nella probabilità che si diano dolori alle gambe in presenza della malattia *A* (la scelta ottimale), la probabilità che si dia tosse in presenza della malattia *A* (una scelta non ottimale) e la probabilità che si dia tosse in presenza della malattia *B* (un'altra scelta non ottimale).

Il problema nella condizione di informazione completa era identico al primo per quanto riguarda la formulazione e i valori probabilistici. Tuttavia, in questa versione del problema tutti valori di probabilità erano scoperti e visibili fin dall'inizio (si veda il lato destro in fig. 1). Come nel primo problema, la descrizione del caso informava i partecipanti della probabilità che si manifestino dolori alle gambe in presenza della malattia *B*. Ai medici veniva quindi chiesto di indicare quale delle restanti probabilità condizionate avrebbero cercato di ottenere contattando la clinica specialistica. Tuttavia, diversamente da quanto accadeva nella prima versione del problema, le informazioni di cui i partecipanti disponevano per la successiva integrazione erano complete, indipendentemente da ciò che avevano scelto di chiedere con la loro «telefonata».

In entrambi i problemi, dopo aver fatto la «telefonata», i partecipanti esprimevano quanto si sentivano convinti di aver scelto l'informazione più appropriata su una scala che variava fra 1 (= «sto soltanto tirando a indovinare») e 10 (= «sono assolutamente certo»). Si chiedeva infine ai medici di indicare una diagnosi (*A* oppure *B*) per il paziente descritto.

Per i dati relativi alla selezione dell'informazione e all'elaborazione della diagnosi da parte dei medici sono stati impiegati metodi di analisi statistica delle contingenze (il test di simmetria McNemar e il test χ^2 di Pearson). Per confrontare i dati relativi alla fiducia dei medici nelle due condizioni è stata utilizzata l'analisi della varianza.

2.2. Risultati

In totale, 105 medici hanno completato i compiti relativi ai due problemi sperimentali. Non si è trovata nessuna differenza significativa nell'appropriatezza delle informazioni selezionate dai medici nelle due condizioni (informazione parziale *vs*. totale): test McNemar, χ^2 (1, 103) = 0,03, *n.s.* (si veda tab. 1). In entrambe le versioni, il 58,7% dei partecipanti ha selezionato l'informazione ottimale. Si tratta di una percentuale superiore a quella attesa con scelte completamente casuali (33%), ma ben lontana da un livello ideale. Sono stati di più i medici che hanno scelto l'informazione ottimale in entrambi i problemi (40,4%) rispetto a quelli che hanno scelto informazioni non ottimali in entrambi i problemi (23,1%). Circa il 37% dei medici ha scelto in modo difforme nei due problemi (questo gruppo di soggetti è equamente diviso fra coloro che hanno scelto in modo ottimale nel primo problema ma non nel secondo e coloro che hanno scelto in modo ottimale nel secondo problema ma non nel primo). La fiducia dei medici nel valore delle informazioni da loro scelte era significativamente più bassa nella condizione di informazione parziale: media (deviazione standard) = rispettivamente, 5,2 (5,1) *vs*. 6,3 (5,0), F (1, 104) = 25,50, $p < 0{,}0001$.

Tab. 1 - Confronto delle scelte dei dati da parte dei 104 medici coinvolti nella condizione di informazione parziale e nella condizione di informazione completa – McNemar, χ^2 (1, 103) = 0,03, non significativo

	Scelta dei dati (informazione parziale)		
	Ottimale	Non ottimale	*Totale*
Scelta dei dati (informazione completa)			
Ottimale	42 (40,4%)	19 (18,3%)	61 (58,7%)
Non ottimale	19 (18,3%)	24 (23,1%)	43 (41,3%)
Totale	61 (58,7%)	43 (41,3%)	104*

* Uno dei partecipanti è stato escluso da questa analisi per un errore nella compilazione del questionario.

Una volta selezionata l'informazione, i medici dovevano integrare i dati che avevano a disposizione per scegliere una diagnosi. La qualità delle loro diagnosi, così come la qualità della loro selezione fra le informazioni, può essere valutata servendosi del teorema di Bayes. Si tratta in questo caso di mettere a confronto la diagnosi scelta da ciascun medico con quella identificata come la più probabile applicando il teorema di Bayes alle informazioni che quel medico aveva a disposizione. Nella condizione di informazione parziale, se il medico ha selezionato l'informazione ottimale, la diagnosi più probabile

è quella relativa alla malattia *B*. Tuttavia, se il medico ha compiuto una scelta non ottimale, allora le probabilità condizionate disponibili non forniscono un rapporto di verisimiglianza e non è quindi possibile combinare fra loro tutti i dati considerati nel teorema di Bayes. In questo caso, la decisione ottimale dal punto di vista normativo dovrebbe basarsi sulla diffusione delle due malattie nella popolazione. Dal momento che la diffusione delle due malattie è la stessa e la scelta è binaria (malattia *A* oppure malattia *B*), nella condizione di informazione parziale con dati non ottimali ci si dovrebbe aspettare che le diagnosi formulate dai medici siano ugualmente distribuite fra le due opzioni. Nella condizione di informazione completa, tutte le informazioni erano disponibili a tutti i partecipanti, indipendentemente dallo svolgimento del compito iniziale di selezione. In questo problema, il teorema di Bayes indica che la malattia *B* rappresenta la diagnosi più probabile.

Nella condizione di informazione parziale, i medici che hanno selezionato l'informazione in modo ottimale sono stati in gran parte capaci di individuare correttamente la diagnosi (92,5%, tab. 2). Tuttavia, i medici che avevano richiesto informazioni non ottimali attraverso il loro «consulto telefonico» hanno diagnosticato la malattia *B* più spesso di quanto ci si sarebbe dovuto aspettare alla luce delle informazioni di cui disponevano (72,1% osservato *vs*. 50% atteso, $p < 0{,}002$, tab. 2). Nella condizione di informazione completa, il 98% dei medici ha individuato la diagnosi più probabile (malattia *B*), anche se parte di loro aveva scelto un dato non ottimale per la «telefonata». Questo risultato non è significativamente differente da quello atteso in base al teorema di Bayes.

Tab. 2 - Percentuale delle decisioni dei 104 medici coinvolti in favore della diagnosi B *nella condizione di informazione parziale*

	n	*Percentuale di diagnosi* B		*z*	*Significatività statistica*
		Osservata	*Attesa (Bayes)*		
Selezione ottimale	62	95,2% (59)	100%	1,77	$p < 0{,}08$
Selezione non ottimale	43	72,1% (31)	50%	3,23	$p < 0{,}002$

2.3. *Discussione*

Riassumendo, non vi è stata una differenza significativa nella percentuale dei medici che hanno selezionato il dato necessario per l'applicazione del teorema di Bayes nel problema in cui disponevano di una parte soltanto delle

informazioni e nel problema in cui tutte le informazioni erano disponibili fin dall'inizio. In entrambi i problemi, più della metà dei medici ha scelto il dato ottimale, mentre circa il 40% ha scelto informazioni non ottimali. La disponibilità di tutte le informazioni non ha quindi influito sulla selezione del dato da ricercare. Ma i medici che avevano scelto in modo non ottimale erano significativamente più fiduciosi dell'appropriatezza delle loro scelte quando avevano a disposizione tutte le informazioni. Sebbene non vi sia una spiegazione certa per questa differenza, sembra probabile che la possibilità di vedere tutti i dati (compreso quello ottimale) abbia permesso a questi soggetti di trovare una giustificazione alle loro scelte sulla base del contrasto fra i diversi valori delle probabilità condizionate.

In confronto alla considerevole incidenza di errori nella raccolta delle informazioni, i medici sono stati più capaci nell'integrare le informazioni disponibili in vista della diagnosi. Quando l'insieme delle informazioni disponibili dipendeva dal precedente compito di selezione (cioè nella condizione di informazione parziale), i medici che avevano selezionato l'informazione ottimale hanno poi impiegato in modo appropriato le informazioni disponibili, formulando diagnosi conformi a quelle indicate dal teorema di Bayes. Tuttavia, i medici che avevano selezionato dati non ottimali hanno mostrato una discrepanza rispetto a un comportamento bayesiano, diagnosticando la malattia *B* più spesso di quanto avrebbero «dovuto» da un punto di vista normativo. Quando tutte le informazioni erano disponibili (cioè nella condizione di informazione completa) il 98% dei medici ha diagnosticato correttamente la condizione del paziente descritto, indipendentemente dalla scelta iniziale (ottimale o non ottimale) del dato più importante da ricercare.

La spiegazione e interpretazione di questi risultati solleva alcune questioni. La proporzione significativamente più alta di scelte per la diagnosi *B* potrebbe essere in parte dovuta al semplice confronto fra i valori di probabilità che i medici avevano di fronte dopo aver svolto una selezione non ottimale delle informazioni: il valore di probabilità (dato) del 58% associato alla malattia *B* era più alto di entrambi i valori di probabilità nelle due celle non ottimali (24% e 43%). Questo semplice contrasto potrebbe aver indotto i medici a preferire la diagnosi *B*, in quanto associata a un valore più alto, anche se era loro possibile interpretare tale valore in termini normativamente corretti. Inoltre l'alto livello di accuratezza delle risposte nella condizione di informazione completa potrebbe essere in parte dovuto al fatto che i partecipanti si erano già impegnati sul problema nella condizione di informazione parziale. D'altra parte, la plausibilità di questa spiegazione è indebolita dal fatto che i medici non si sono mostrati più capaci di identificare il dato ottimale nella condizione di informazione completa, sebbene avessero precedentemente svolto il compito nella condizione di informazione parziale.

L'incertezza nell'interpretazione dei risultati riflette diversi limiti di questo studio. Il fatto che ogni medico abbia affrontato il problema in entrambe le versioni potrebbe suscitare dubbi riguardo all'indipendenza delle risposte nelle due condizioni. Inoltre, i due problemi erano identici nella formulazione e sono stati presentati a tutti i soggetti nello stesso ordine. Questo induce a chiedersi se i medici possano aver appreso o individuato il procedimento di selezione ottimale dell'informazione nello svolgimento del primo problema e averlo poi trasferito nello svolgimento del secondo problema. Va detto, comunque, che i dati non sostengono questa ipotesi, nel senso che, come abbiamo indicato, non vi è stato nel complesso alcun incremento nella qualità della selezione dei dati dal primo al secondo problema. Inoltre, i medici che sono passati da una selezione ottimale nel primo problema a una non ottimale nel secondo e quelli che hanno mostrato la sequenza opposta sono in numero uguale.

È anche possibile che il compito sperimentale sia stato eccessivamente semplificato dalla scelta di presentare due malattie la cui diffusione nella popolazione di riferimento era presentata come identica e non influenzava quindi in alcun modo il rapporto fra le probabilità delle due diagnosi alternative. In queste condizioni, non è possibile stabilire se i soggetti abbiano o non abbiano considerato la diffusione delle due malattie nelle loro decisioni diagnostiche.

Nel tentativo di affrontare alcuni dei problemi appena descritti e di chiarire ulteriormente l'importanza relativa della raccolta e dell'integrazione delle informazioni nelle decisioni mediche, abbiamo ideato un secondo studio con alcune importanti differenze: 1) il secondo esperimento prevede innanzi tutto un confronto fra informazione completa e informazione parziale in cui la condizione varia fra due distinti gruppi di soggetti (*between-subjects*); 2) comprende una serie di problemi fra loro diversi che si distinguono per le alternative diagnostiche considerate, per la diffusione delle relative patologie, e per l'insieme dei sintomi descritti; 3) fa uso di una formulazione modificata dei problemi, in cui nessun valore di probabilità condizionata è noto fin dall'inizio; 4) comprende, oltre a una scelta diagnostica dicotomica, la richiesta di una stima di probabilità della correttezza delle due diagnosi alternative.

3. Esperimento II

3.1. *Metodi*

Abbiamo elaborato tre problemi diagnostici simili nel formato a quelli dell'Esperimento I. I tre problemi si distinguevano per le malattie e i sintomi considerati, per i valori delle probabilità condizionate contenuti nelle celle

della matrice e per la diffusione delle due malattie nella popolazione. Il rapporto fra la diffusione generale dell'una e dell'altra malattia nei tre problemi era, rispettivamente, 1:1 (identica probabilità a priori), 2:1 (una malattia è a priori due volte più probabile dell'altra), 5:1 (una malattia è a priori cinque volte più probabile dell'altra). Tutte le ipotesi diagnostiche erano presentate in termini astratti (per esempio: la malattia *H*).

In un paragrafo introduttivo si chiedeva ai partecipanti di immaginare di praticare la professione in una clinica dotata di un data-base computerizzato che può essere consultato come strumento di sostegno alla diagnosi. Si informavano i partecipanti che il data-base poteva fornire loro informazioni su precedenti pazienti delle clinica, sulle condizioni patologiche di cui soffrivano e sui sintomi che le accompagnavano. I soggetti dovevano selezionare le due celle della matrice 2 × 2 con le informazioni che essi ritenevano più utili in vista della diagnosi. Dovevano quindi indicare quanto erano convinti di aver effettivamente selezionato la coppia di informazioni maggiormente utili su una scala che variava fra 1 (= «sto soltanto tirando a indovinare») e 10 (= «sono assolutamente certo»). In seguito, dovevano indicare la diagnosi a loro parere più probabile e riportare i valori di probabilità che essi assegnavano a ciascuna delle due diagnosi alternative. Le istruzioni specificavano che queste stime di probabilità dovevano sommare a 100%.

Ogni partecipante riceveva tutti e tre i problemi in un ordine di presentazione stabilito in modo casuale. Nella condizione di informazione completa le probabilità condizionate di ciascun sintomo data la presenza di ciascuna delle due malattie erano tutte visibili. Per contro, nella condizione di informazione parziale i quattro valori di probabilità erano stampati con un inchiostro invisibile che poteva essere reso visibile attraverso l'applicazione di una sostanza chimica. Nella condizione di informazione completa si dovevano indicare le due informazioni selezionate cerchiando le relative cifre. Nella condizione di informazione parziale si dovevano scoprire le due informazioni selezionate ricalcandole con una particolare «penna» che rendeva visibili le immagini latenti. Le due versioni differivano riguardo alla quantità di informazioni disponibili per l'elaborazione di un giudizio diagnostico. Nella condizione di informazione completa tutti i dati erano disponibili fin dall'inizio, mentre nella condizione di informazione parziale erano disponibili solo i due dati che erano stati selezionati. La condizione di informazione parziale determinava quindi specifici vincoli sui giudizi diagnostici dei medici in funzione della qualità delle informazioni che avevano precedentemente selezionato. Nella condizione di informazione completa, per contro, non erano presenti vincoli di questo tipo.

Le due versioni del compito sperimentale sono state assegnate in modo casuale a 108 medici che partecipavano a due corsi di formazione continua presso la Facoltà di Medicina dell'Università del Michigan. Le prestazioni

dei medici nei tre problemi sono state valutate sulla base delle seguenti variabili.

1. La percentuale di problemi nei quali ciascun medico ha selezionato una coppia di informazioni ottimale o non ottimale (dal punto di vista normativo bayesiano). Queste percentuali sono calcolate in base al numero dei problemi completati.
2. La media delle valutazioni della fiducia di ciascun medico nell'utilità delle informazioni che egli aveva scelto. La media è stata calcolata per ogni singolo partecipante rispetto ai problemi che egli aveva completato. Quindi è stata nuovamente calcolata separando i problemi nei quali il medico aveva scelto in modo ottimale da quelli in cui aveva scelto in modo non ottimale. Pertanto, a ogni soggetto possono essere associate fino a tre diverse medie delle valutazioni delle fiducia: la prima relativa a tutti i problemi che ha risolto, la seconda relativa ai soli problemi in cui ha scelto le informazioni in modo ottimale e la terza relativa ai problemi in cui ha scelto le informazioni in modo non ottimale.
3. La qualità delle decisioni diagnostiche di ciascun medico è stata misurata attraverso il confronto con le soluzioni bayesiane normative basate sui dati che, in ciascun problema, quel medico aveva a disposizione. Il livello di qualità è stato espresso per ogni medico attraverso la percentuale dei problemi in cui egli ha elaborato la diagnosi ottimale.
4. Anche l'accuratezza delle stime probabilistiche è stata valutata confrontandole con le corrispondenti soluzioni bayesiane basate sulle informazioni disponibili. L'accuratezza di ciascun medico è stata espressa attraverso la media degli scarti assoluti fra le stime fornite e i valori normativi bayesiani nei problemi che il medico ha completato. Sono stati impiegati gli scarti assoluti per evitare che errori di stima per eccesso e per difetto si cancellassero fra di loro.

3.2. *Risultati*

Non si è trovata alcuna differenza significativa fra le due condizioni nell'appropriatezza delle informazioni che i medici hanno individuato come le più utili. In entrambe le condizioni, i medici hanno selezionato una coppia ottimale di informazioni nel 25% circa dei casi (media = 25,5% nella condizione di informazione completa e 23,6% nella condizione di informazione parziale, $t = 0{,}34$, $p < 0{,}74$).

La fiducia dei medici nell'utilità delle informazioni che avevano selezionato è stata significativamente maggiore nella condizione di informazione totale che nella condizione di informazione parziale: media (deviazione standard) = rispettivamente 5,3 (2,7) *vs.* 3,8 (2,1), F (1, 105) = 9,88, $p < 0{,}002$. È da no-

tare che, in entrambe le condizioni, la fiducia dei medici nelle loro scelte non era influenzata dal fatto che tali scelte fossero o non fossero effettivamente ottimali; in altri termini, erano altrettanto fiduciosi avendo compiuto scelte pseudodiagnostiche di quanto lo fossero avendo compiuto scelte ottimali (t = 1,09, *n.s.*, e 0,27, *n.s.*, rispettivamente per la condizione di informazione completa e quella di informazione parziale).

Indipendentemente dalla qualità delle informazioni che i medici avevano selezionato, nelle diagnosi da loro effettivamente indicate hanno fatto un uso piuttosto buono delle informazioni che avevano a disposizione. In media, nella condizione di informazione totale, i medici hanno elaborato diagnosi ottimali nel 66,7% dei problemi; nella condizione di informazione parziale, hanno elaborato diagnosi ottimali nel 61,8% dei casi (t = 0,80, p < 0,43). Si noti ancora una volta che l'identificazione della diagnosi ottimale è determinata dall'applicazione del teorema di Bayes alle informazioni che ciascun medico aveva a disposizione in ciascun problema. Nella condizione di informazione totale, la diagnosi ottimale era la stessa per tutti i soggetti, indipendentemente dai dati che essi avevano precedentemente selezionato come più utili. Nella condizione di informazione parziale, per contro, la diagnosi più probabile poteva essere differente a seconda che in precedenza fossero stati selezionati dati ottimali o non ottimali. Il motivo di ciò è che coloro che avevano selezionato informazioni non ottimali avrebbero dovuto, in base al teorema di Bayes, fondare la loro diagnosi unicamente sulla diffusione dell'una e dell'altra malattia. Per contro, una selezione ottimale dei dati offriva informazioni che potevano alterare il rapporto fra le probabilità delle due diagnosi, indicando come più probabile a posteriori una diagnosi differente da quella più probabile a priori. Le percentuali qui riportate misurano quindi la capacità dei medici coinvolti di far buon uso delle informazioni che avevano, in quanto distinta dalla capacità di selezionare nel modo migliore le informazioni da ricercare.

Anche l'accuratezza delle stime di probabilità dei medici non ha mostrato differenze significative fra le due condizioni. Lo scarto assoluto medio fra le stime di probabilità dei medici e i valori ottenuti applicando il teorema di Bayes ai dati disponibili è stato del 16,5% (deviazione standard = 7,3) nella condizione di informazione completa a fronte del 18,5% (10,2) nella condizione di informazione parziale (t = 1,09, p < 0,29).

3.3. *Discussione*

Come nell'Esperimento I, nei problemi compresi nell'Esperimento II le due differenti condizioni sperimentali non hanno influito sulla frequenza con cui le informazioni venivano selezionate in modo ottimale. Sia nella condi-

zione di informazione completa sia in quella di informazione parziale i medici hanno selezionato informazioni ottimali in media nel 25% dei casi. Sebbene la qualità delle informazioni raccolte non sia stata differente nelle due condizioni, lo è stata la fiducia dei medici nella correttezza delle loro scelte. I medici si sono mostrati molto più sicuri di aver scelto le informazioni migliori quando avevano a disposizione tutti i valori di probabilità, e indipendentemente dal fatto che avessero effettivamente scelto informazioni ottimali.

Sebbene i medici abbiamo *raccolto* le informazioni conformemente a quanto previsto dal teorema di Bayes solo nel 25% dei problemi, essi hanno *impiegato* le informazioni in accordo con il teorema di Bayes in più del 60% dei casi. In qualche misura, questa differenza potrebbe riflettere la differenza nelle probabilità che i due eventi si verifichino su basi puramente casuali (33% per la selezione ottimale dell'informazione *vs*. 50% per l'individuazione della diagnosi «corretta»). Tuttavia, questo risultato conferma quanto si è osservato nell'Esperimento I, nel quale un'alta percentuale di medici aveva indicato la diagnosi corretta dal punto di vista normativo.

4. Conclusioni

I risultati dei due esperimenti suggeriscono che i medici sono stati più capaci di scegliere l'ipotesi diagnostica normativamente corretta (specie se già in possesso di informazioni ottimali) che non nel selezionare le informazioni più utili da ricercare. Questa conclusione richiama l'osservazione di Beyth-Marom, Fischhoff (1983) secondo la quale le persone mostrano generalmente una corretta comprensione qualitativa del valore diagnostico di informazioni che trovano già selezionate e organizzate, ma sembrano avere molte più difficoltà nel selezionare autonomamente tali informazioni. Questi risultati suggeriscono che decisioni diagnostiche manchevoli possono spesso dipendere da una difficoltà nel valutare quali informazioni siano utili e significative dal punto di vista diagnostico, piuttosto che da una difficoltà nel combinare, integrare e organizzare le informazioni disponibili in vista di una decisione finale – almeno per quanto riguarda compiti relativamente semplici come quelli impiegati in questo studio.

È comunque necessario qualificare in diversi modi le conclusioni tratte dalla presente indagine. Per esempio, il tipo di compito diagnostico studiato era piuttosto semplice, essendo limitato a due sole ipotesi diagnostiche alternative e a due soli sintomi. Inoltre, come criterio di accuratezza è stata impiegata la diagnosi più probabile alla luce del teorema di Bayes, ma nei problemi diagnostici reali altri fattori possono influire su quale sia la scelta diagnostica più corretta. Infine, i medici ricevevano le ipotesi diagnostiche alternative fin dall'inizio, anziché doverle elaborare autonomamente sulla base delle

informazioni cliniche, come tipicamente avviene. Determinare se i nostri risultati potranno essere estesi a situazioni più complesse che comprendono un maggior numero di informazioni e di possibili diagnosi resta un compito per ulteriori indagini.

Un'altra questione significativa rispetto ai problemi che abbiamo indagato riguarda l'influenza dell'esperienza clinica sulle prestazioni dei medici in compiti di questo tipo. Ricerche precedenti hanno mostrato che esperti e non esperti impiegano le ipotesi e le informazioni in modo differente per giungere a una diagnosi (Patel, Groen, 1986; Patel, HoPingKong, Mark, 1984; Patel, Evans, Kaufman, 1989). Queste differenze possono essere attribuite alla diversa ampiezza e al diverso livello di organizzazione della base di conoscenze degli esperti e dei non esperti. Nel presente studio abbiamo deliberatamente cercato di eliminare la possibile influenza di queste differenze descrivendo le ipotesi diagnostiche in termini astratti (per esempio: la malattia *A*), in modo da studiare i processi decisionali a un livello generale e indipendentemente dalle conoscenze specifiche. Sebbene non si siano osservate differenze chiare e sostanziali nei risultati ottenuti nell'Esperimento I (nel quale i soggetti erano medici al primo anno di specializzazione) e nell'Esperimento II (nel quale i soggetti erano medici praticanti a diversi livelli di esperienza), le differenze nei problemi e nella struttura dei due esperimenti non permette di trarre precise conclusioni riguardo al ruolo dell'esperienza clinica. Saranno quindi necessarie ulteriori ricerche volte a determinare in che misura il livello di esperienza influisce sulle prestazioni dei medici in compiti simili a questi nel contesto di specifiche aree di competenza.

La nostra conclusione per cui la raccolta e la selezione delle informazioni sono più problematiche dell'integrazione e dell'impiego delle informazioni disponibili, se corretta, suggerisce che gli interventi educativi dovrebbero mirare al miglioramento delle procedure attraverso le quali i medici raccolgono le informazioni cliniche. Uno dei tentativi finora svolti in questo senso ha offerto alcuni risultati incoraggianti: una sessione di addestramento piuttosto semplice e breve ha prodotto notevoli miglioramenti nella selezione dei dati rilevanti in problemi simili a quelli impiegati nel presente studio (Wolf, Gruppen, Billi, 1988; Lord, Lepper, Preston, 1984). Resta da determinare quanto possano essere duraturi e generalizzabili i risultati di interventi educativi di questo tipo.

Gli autori desiderano ringraziare J. Frank Yates, Ted E. Dielman, Patricia Mullan, Jay Christensen-Szalanski e un *referee* anonimo per i loro utili commenti su precedenti versioni di questo articolo.

4. *I giudizi diagnostici in casi di sospetta batteriemia: euristica della disponibilità e* wishful thinking*

di *Roy M. Poses, Michele Anthony*

1. Introduzione

Spesso i medici devono prendere decisioni relative al trattamento dei pazienti prima di aver elaborato una diagnosi definitiva. Gli esperti di analisi delle decisioni hanno suggerito che questo tipo di scelte dovrebbe fondarsi sui giudizi dei medici riguardo alla probabilità che il paziente soffra di una particolare patologia. Ma queste raccomandazioni potrebbero migliorare le decisioni solo a condizione che i giudizi probabilistici dei medici fossero accurati. Anche i medici che non si riconoscono nelle indicazioni dell'analisi delle decisioni, del resto, possono talvolta trovarsi a decidere sulla base dei loro giudizi di probabilità. In tutti questi casi, giudizi probabilistici manchevoli possono condurre a decisioni non ottimali.

Sfortunatamente, i giudizi intuitivi dei medici riguardo alle probabilità di diversi esiti clinici possono essere poco affidabili (Poses *et al.*, 1985; Christensen-Szalanski, Bushyhead, 1983; Thornbury, Fryback, Edwards, 1975; DeSmet, Fryback, Thornbury, 1979; Eisenberg *et al.*, 1982; Tierney *et al.*, 1986)[1]. Non c'è una chiara spiegazione del fatto che professionisti formati possano incontrare tante difficoltà nello svolgimento di compiti cognitivi che sembrano centrali nella loro disciplina. La ricerca sperimentale in psicologia cognitiva e nella psicologia della decisione ha mostrato seri limiti nell'elaborazione di giudizi probabilistici da parte degli individui. A quanto pare, numerosi fattori possono distorcere i giudizi in modo sistematico, e spesso le persone si affidano a semplici regole, dette «euristiche del giudizio», che talvolta risultano fuorvianti (Hogarth, 1987; Kahneman, Slovic, Tversky, 1982;

* Traduzione di *Availability, wishful thinking, and physicians' diagnostic judgments for patients with suspected bacteremia*, «Medical decision making», 11 (1991), pp. 159-168. Per gentile concessione dell'editore.

1. Si vedano anche, in questo volume, i capp. 1, 2 e 12 [N.d.C.].

Wright, Ayton, 1987). Non è noto, però, in che misura queste osservazioni possano trovare applicazione nel contesto dei giudizi e delle decisioni dei medici riguardo a pazienti reali. In particolare, alcuni psicologi hanno contestato che i risultati sperimentali possano essere generalizzati alle situazioni del mondo reale, dal momento che gran parte degli esperimenti in questione sono stati svolti con studenti universitari ai primi anni di corso e in condizioni molto particolari (Christensen-Szalanski, 1986; Christensen-Szalanski, Fobian, 1988). Comprendere più a fondo le cause delle difficoltà dei medici nell'elaborare giudizi probabilistici può suggerire direzioni di intervento per il miglioramento delle loro decisioni e della qualità delle cure.

L'obiettivo dello studio qui presentato è spiegare le difficoltà dei medici nel giudicare la probabilità della batteriemia in pazienti ospedalizzati cui erano state prescritte delle emocolture. Abbiamo selezionato questo specifico problema clinico per una serie di ragioni. Innanzi tutto, si è mostrato che una tempestiva terapia antibiotica incide positivamente sugli esiti clinici di pazienti affetti da batteriemia (Kreger, Craven, McCabe, 1980; Weinstein *et al.*, 1983a), mentre senza una pronta terapia antibiotica i pazienti batteriemici incorrono in maggiori spese mediche (Dunagan *et al.*, 1989) e in un più elevato rischio di morte (Meyers *et al.*, 1989). I medici devono inoltre fondare la decisione iniziale riguardo alla terapia antibiotica sulla probabilità stimata che il paziente sia affetto da batteriemia, perché la diagnosi definitiva fornita dalle emocolture non è immediata e perché trattare fin dall'inizio tutti i pazienti con svariati agenti antibiotici risulterebbe costoso e potenzialmente pericoloso (Makadon *et al.*, 1987). La batteriemia non è facile da diagnosticare in assenza dei risultati delle emocolture, perché non vi sono riscontri clinici patognomici per questa condizione o tali che permettano chiaramente di escluderla (Makadon *et al.*, 1987; Mellors *et al.*, 1987). Infine, l'evidenza disponibile suggerisce che i medici possono prendere decisioni inappropriate riguardo alla terapia antibiotica in pazienti con sospetta batteriemia (Dunagan *et al.*, 1989) e che talvolta non riescono a somministrare gli antibiotici in maniera sufficientemente tempestiva (Weinstein *et al.*, 1983b).

Le nostre specifiche ipotesi prevedevano che l'utilizzo inappropriato dell'«euristica della disponibilità» (*availability heuristic*) e gli effetti di una «distorsione valutativa» (*value bias*) potessero influenzare i giudizi dei medici. L'euristica della disponibilità consiste nello stimare la probabilità di un evento in base alla facilità con la quale eventi simili possono essere richiamati alla memoria (o in base alla «disponibilità» del ricordo di eventi simili). Si ha invece una distorsione valutativa quando la probabilità di un evento è stimata in funzione della gravità delle sue possibili conseguenze (Dawson, Arkes, 1987).

2. Metodi

2.1. *Schema dello studio e popolazione dei pazienti*

Abbiamo condotto uno studio prospettico su una coorte di 227 pazienti ai quali erano state prescritte una o più serie di emocolture. Abbiamo identificato i primi 82 pazienti sollecitando ripetutamente alcuni specializzandi in medicina a segnalarci i pazienti a cui avevano ordinato questo tipo di esami. Abbiamo quindi identificato i restanti 145 pazienti attraverso la consultazione giornaliera del registro delle emocolture del laboratorio di microbiologia. Tutti i pazienti individuati erano stati ricoverati in un grande ospedale universitario fra novembre 1984 e giugno 1985. L'ospedale era la principale sede di formazione pratica di una Facoltà di Medicina. Tutti i pazienti sono stati inizialmente seguiti da un medico al primo anno di specialità sotto la supervisione di uno specializzando con una maggiore esperienza e da un medico responsabile.

2.2. *I giudizi dei medici*

Per ogni singolo paziente, abbiamo identificato il medico (solitamente uno specializzando al primo anno, in alcuni casi uno specializzando con una maggiore esperienza) che aveva ordinato le emocolture. Entro le prime 24 ore dalla prescrizione della prima serie di emocolture, abbiamo raccolto la stima probabilistica intuitiva del medico riguardo al fatto che «almeno una delle emocolture risulti positiva». Queste stime sono state ottenute attraverso questionari scritti oppure, nei casi in cui i questionari non erano stati volontariamente compilati, attraverso un contatto diretto con il medico da parte di uno degli autori (Michele Anthony) o di un assistente coinvolto nella ricerca. Al momento in cui hanno raccolto i dati, né l'autore né l'assistente erano al corrente delle specifiche ipotesi dell'indagine. Abbiamo ottenuto le stime probabilistiche dei medici prima che fossero disponibili i risultati di ciascuna delle colture che essi avevano ordinato. Gli specializzandi coinvolti erano i medici incaricati di una prima valutazione del sospetto di batteriemia in questi pazienti. Essendo «in prima fila», essi svolgevano normalmente un ruolo centrale nelle decisioni cliniche, anche se formalmente ne condividevano la responsabilità con colleghi più esperti. Come indicato da Kunin (1988), gli specializzandi «devono spesso prendere da sé decisioni difficili e molto specifiche […] con un notevole carico di lavoro e con poco tempo a disposizione per riflettere o per condurre un'approfondita consultazione della letteratura».

2.3. *Variabili associate al ricordo e variabili valutative*

Abbiamo ipotizzato che i medici avrebbero impiegato l'euristica della disponibilità nel giudicare la probabilità delle batteriemia. Abbiamo operazionalizzato questa ipotesi prevedendo che i giudizi di ciascun medico per i singoli pazienti sarebbero stati associati alla disponibilità di ricordi relativi a casi di batteriemia nel momento in cui il medico vedeva ciascuno di quei pazienti. Nei questionari e nelle interviste abbiamo raccolto diverse risposte che ritenevamo potessero riflettere la disponibilità alla memoria di ricordi di questo tipo. Tali «variabili associate al ricordo» (*recalled experience variables*) erano: la proporzione dei pazienti ospedalizzati seguìti dal medico che egli ricordava stessero ricevendo antibiotici in quel momento; il numero di pazienti che il medico ricordava avessero avuto emocolture positive nei mesi precedenti; infine, se il medico ricordasse o meno di avere curato numerosi pazienti affetti da batteriemia. Dal momento che l'euristica delle disponibilità è definita come la tendenza a valutare «la probabilità di un evento sulla base della facilità con cui vengono riportati alla mente esempi o occorrenze dello stesso tipo di evento» (Tversky, Kahneman, 1974), abbiamo raccolto dati relativi ai ricordi di esperienze passate. Per non sollecitare l'impiego dell'euristica della disponibilità da parte dei medici, abbiamo espresso queste domande in termini neutrali, senza far riferimento all'intensità o all'impatto emotivo dei ricordi in questione.

Abbiamo anche ipotizzato che una distorsione valutativa avrebbe influenzato i giudizi dei medici. Abbiamo operazionalizzato questa ipotesi prevedendo che le stime di probabilità dei medici sarebbero state associate con i loro giudizi di valore, cioè con i loro giudizi riguardo alla potenziale importanza della batteriemia per ciascun paziente nel caso che egli ne fosse realmente affetto. Abbiamo identificato questo fattore con il rischio aggiuntivo di morte che, a parere del medico, la batteriemia avrebbe comportato per ciascun paziente, perché la morte prematura è il più catastrofico fra i potenziali esiti della batteriemia. Abbiamo quindi richiesto a ogni medico una stima della probabilità che ciascun paziente sopravvivesse fino al congedo ospedaliero (indichiamo questa stima con «p(sopravvivenza)») e della probabilità che il paziente sopravvivesse assumendo di sapere che era affetto da batteriemia (indichiamo questa stima con «p(sopravvivenza|batteriemia)»). A partire da questi valori di probabilità abbiamo calcolato delle «variabili valutative» (*value variables*). La prima di tali variabili è il rischio relativo (stimato) di morte di ciascun paziente a causa della batteriemia, cioè:

$$\begin{aligned} RR_{morte} &= p(\text{morte}|\text{betteriemia})/p(\text{morte}) = \\ &= [1 - p(\text{sopravvivenza}|\text{batteriemia})]/[1 - p(\text{sopravvivenza})] \end{aligned}$$

La seconda variabile valutativa è il rischio incrementale (stimato) di morte di ciascun paziente a causa della batteriemia:

$$\begin{aligned} RI_{morte} &= p(\text{morte}|\text{betteriemia}) - p(\text{morte}) = \\ &= p(\text{sopravvivenza}) - p(\text{sopravvivenza}|\text{batteriemia}) \end{aligned}$$

Dal momento che la distribuzione di RR_{morte} presentava una curvatura molto pronunciata, abbiamo anche calcolato il suo logaritmo naturale, $ln(RR_{morte})$. Il rischio relativo e il suo logaritmo sono misure epidemiologiche convenzionalmente utilizzate negli studi di coorte (Sackett, Haynes, Tugwell, 1985).

2.4. *Ulteriori dati raccolti*

Abbiamo raccolto dati relativi a variabili cliniche che i medici avrebbero potuto impiegare nei loro giudizi intuitivi e che avrebbero potuto sovrapporsi all'influenza dell'euristica della disponibilità e di una distorsione valutativa. Questi dati comprendono le valutazioni dei medici riguardo alle condizioni generali di ciascun paziente (espresse attraverso valori discreti, variabili fra «non malato» e «morente»), alla suscettibilità del paziente alle infezioni e alla probabilità che il paziente avesse infezioni locali preesistenti. Abbiamo anche registrato dati demografici, dati clinici che potevano essere correlati con la presenza o l'assenza di batteriemia e le informazioni necessarie per calcolare l'APACHE II Acute Physiology Score (un punteggio che si è dimostrato predittivo della severità generale delle patologie di pazienti in condizioni critiche; Knaus *et al.*, 1985). Abbiamo infine considerato se i pazienti avevano recentemente ricevuto trattamenti antibiotici o trattamenti farmacologici di altro genere che avrebbero potuto alterare le loro reazioni alle infezioni. Un assistente coinvolto nella ricerca ha classificato, sulla base di criteri convenzionali (McCabe, Jackson, 1962), ciascuna delle patologie soggiacenti dei pazienti come «non fatale», «fatale» o «rapidamente fatale».

2.5. *La qualità dei giudizi diagnostici dei medici*

La misurazione della qualità dei giudizi probabilistici dei medici richiede che essi vengano messi a confronto con un livello di controllo. A questo scopo, abbiamo raccolto i risultati finali delle emocolture di ciascun paziente. Abbiamo registrato la presenza degli organismi identificati e il corrispondente schema di suscettibilità a trattamenti antibiotici.

Dal momento che non esiste una singola tecnica statistica generalmente accettata per valutare le stime probabilistiche (Poses, Cebul, Centor, 1988),

abbiamo impiegato tre diversi metodi. La «curva di calibrazione» confronta graficamente i giudizi quantitativi con le effettive frequenze dei corrispondenti esiti (Lichtenstein, Fischhoff, Phillips, 1980) e fornisce una rappresentazione del «grado in cui le probabilità stimate da un soggetto corrispondono, dal punto di vista delle frequenze relative, a ciò che di fatto si verifica» (Winkler, 1986). Per costruire la curva di calibrazione, abbiamo innanzi tutto diviso i pazienti in gruppi ai quali erano stati assegnati giudizi di probabilità identici o simili. Per ciascuno dei gruppi ottenuti, abbiamo associato alla media delle stime di probabilità la frequenza effettivamente osservata della malattia. Abbiamo quindi calcolato il livello di significatività statistica della differenza fra i valori stimati e quelli effettivamente osservati impiegando il teorema binomiale o la sua approssimazione normale. Infine, abbiamo diviso il valore critico di p per il numero dei gruppi analizzati, utilizzando la correzione di Bonferroni per i confronti multipli (Ingelfinger *et al.*, 1983). Per la nostra analisi principale abbiamo identificato la batteriemia con la crescita entro 24 ore di un organismo «patogeno» (esclusi, cioè, *corynebacteria*, difteroidi, o *staphylococcus epidermidis*) in una o più emocolture o con la crescita, nello stesso lasso di tempo, di un qualsiasi singolo organismo in almeno due emocolture separate (Haley *et al.*, 1980).

Applicando un secondo metodo, suggerito da Spiegelhalter (1985), abbiamo diviso le stime di probabilità in decili e calcolato i valori z per ciascun decile a partire dai tassi osservati e da quelli attesi in base alle stime dei medici. La somma dei quadrati di questi valori z è una statistica χ^2 (con tanti gradi di libertà quanti sono i gruppi) che controlla l'affidabilità generale delle stime.

Abbiamo infine analizzato la capacità di discriminazione utilizzando l'analisi della curva ROC (*receiver operating characteristic*) (Levi, 1985; Berwick, Thibodeau, 1983). La capacità di discriminazione si distingue dall'accuratezza (calibrazione) e dall'affidabilità e consiste nella capacità di suddividere correttamente i pazienti in gruppi distinti caratterizzati da diversi tassi di emocolture positive. Per determinare la capacità di discriminazione abbiamo considerato una serie di diversi valori-soglia (*cut-off*) per ciascun decile e abbiamo identificato come «positive» le stime maggiori o uguali a ognuno dei valori-soglia e come «negative» quelle minori di ognuno dei valori-soglia. Abbiamo quindi costruito una curva ROC associando al tasso di veri positivi (sensibilità) delle stime il tasso di falsi positivi (1 – specificità). L'area al di sotto della curva, anche detta «indice-*c*» o «probabilità di concordanza» (Lee *et al.*, 1986), misura la capacità di discriminazione ed equivale alla probabilità che sia stata assegnata una stima di probabilità più alta al paziente batteriemico in una coppia selezionata casualmente e formata da un paziente batteriemico e da uno non batteriemico. Un'area uguale a 1 significa che le stime di probabilità assegnate ai pazienti con emocolture positive sono sempre

superiori a quelle assegnate a quelli con emocolture negative. Un'area uguale a 0,5 significa che le stime di probabilità sono distribuite in modo casuale fra i pazienti batteriemici e quelli non batteriemici. Abbiamo calcolato l'area utilizzando il metodo di Hanley, McNeil (1982) adattato da Centor, Schwartz (1983).

2.6. *Altre analisi statistiche*

Abbiamo innanzi tutto analizzato il rapporto fra le variabili individuali e le stime probabilistiche dei medici utilizzando statistiche univariate. Per le variabili categoriche abbiamo impiegato il test ANOVA. Per le variabili continue abbiamo impiegato la regressione lineare semplice. Abbiamo controllato i rapporti delle variabili associate ai ricordi e delle variabili valutative con le stime di probabilità usando il livello convenzionale di significatività statistica di 0,05. Per le altre analisi abbiamo modificato i valori *p* ottenuti attraverso la correzione di Bonferroni per i confronti multipli (Ingelfinger *et al.*, 1983).

Abbiamo inoltre condotto una serie di analisi basate sulla regressione lineare multipla. Il loro principale obiettivo era controllare se le variabili cliniche potessero aver interferito nei rapporti delle variabili associate al ricordo e delle variabili valutative con i giudizi diagnostici dei medici (per esempio, era plausibile aspettarsi che le variabili relative alla severità delle condizioni del paziente fossero correlate sia con le probabilità stimate di batteriemia sia con le probabilità stimate di morte a causa della batteriemia). Un altro obiettivo era controllare il peso relativo delle variabili associate al ricordo e delle variabili valutative in confronto alle variabili cliniche. Da un punto di vista concettuale, questo tipo di analisi si rifà al «modello a lente» (*lens model*) di Brunswik. Tale modello prevede l'impiego della regressione lineare per analizzare gli effetti di una molteplicità di elementi di informazione (*cues*) sulla probabilità stimata (trattata come una variabile dipendente continua) (Wigton, 1988). Abbiamo costruito una serie di modelli impiegando come variabili indipendenti diverse variabili associate al ricordo, o combinazioni di coppie di tali variabili, e diverse variabili valutative. Ciascun modello includeva come variabili indipendenti anche svariati fattori clinici che le precedenti analisi univariate o il nostro giudizio clinico indicavano come potenzialmente correlate con le stime di probabilità dei medici.

3. Risultati

3.1. *Dati descrittivi*

Abbiamo raccolto dati relativi a 227 pazienti (51% uomini, 49% donne) e a 34 specializzandi. I medici al primo anno di specialità hanno fornito la maggior parte (206) delle stime di probabilità. Il numero di pazienti considerati da ciascun medico variava fra 1 e 20. Nessun singolo medico ha quindi fornito più del 9% delle stime di probabilità raccolte in questo studio. La maggior parte delle stime proveniva da 10 medici, ciascuno dei quali ha fornito fra le 10 e le 20 stime. L'età dei pazienti variava fra 17 e 91 anni, con una media di 61. Il loro tasso di sopravvivenza fino al congedo ospedaliero è stato del 71,4%.

Per la maggior parte dei pazienti (85%) sono state svolte diverse emocolture (almeno due) entro un periodo di 24 ore. Per 27 pazienti si sono ottenuti risultati positivi in relazione a uno o più organismi. In 10 delle 27 emocolture positive sono cresciuti due differenti organismi. La distribuzione dei diversi microrganismi rilevati è riportata in tab. 1. In base alla nostra definizione, 8 pazienti erano affetti da batteriemia e 9 avevano emocolture contaminate.

Tab. 1 - Risultati delle emocolture

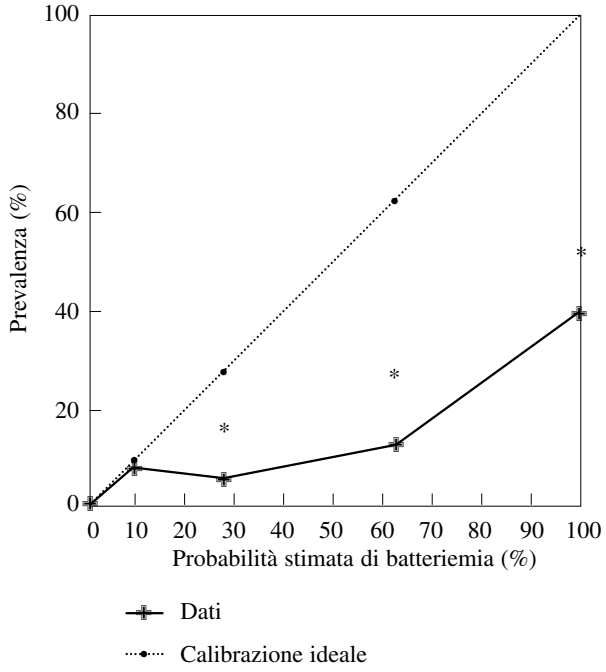
Organismo	*Pazienti con colture positive*
Acinetobacter	1
Corynebacteria	1
Difteroidi	1
Enterococco	3
Erysipelothrix	1
Escherichia coli	7
Klebsiella oxytoca	1
Morganella	1
Proteus mirabilis	1
Staphylococcus aureus	4
Staphylococcus epidermidis	9
Streptococcus pneumoniae	4

3.2. *La qualità dei giudizi diagnostici dei medici*

Le valutazioni intuitive della probabilità di batteriemia da parte dei medici non sono state molto accurate. In particolare, per la maggior parte dei pazien-

ti, i medici coinvolti hanno sovrastimato la probabilità di batteriemia in modo significativo e notevole. La fig. 1 illustra la curva di calibrazione ottenuta dalle loro stime aggregate. La curva riflette il rapporto fra le medie delle stime dei medici (sull'asse *x*) e l'effettiva prevalenza della batteriemia (sull'asse *y*) per gruppi di pazienti cui erano state assegnate stime probabilistiche simili. La curva associata a stime perfettamente calibrate compare come una linea retta a 45°. Per il gruppo di 71 pazienti cui erano state assegnate stime di probabilità comprese fra 16% e 40% (media: 26,8%), il tasso effettivo di batteriemia era del 4,2% ($z = 4{,}4$; $p < 0{,}00005$). Per il gruppo di 74 pazienti cui erano state assegnate stime di probabilità comprese fra 41% e il 99% (media: 62%) il tasso effettivo era del 12,2% ($z = 8{,}9$, $p < 0{,}00005$). Dei cinque pazienti nei quali i medici avevano diagnosticato con certezza la batteriemia (assegnando stime di probabilità del 100%), soltanto due effettivamente l'avevano ($p = 0$). Una curva di calibrazione costruita impiegando una diversa definizione della malattia (cioè assumendo la presenza della batteriemia nel caso in cui una qualsiasi coltura fosse positiva per un qualsiasi organismo) è risultata molto simile a quella presentata.

Fig. 1 - Curva di calibrazione ricavata dalle stime di probabilità di batteriemia dei medici

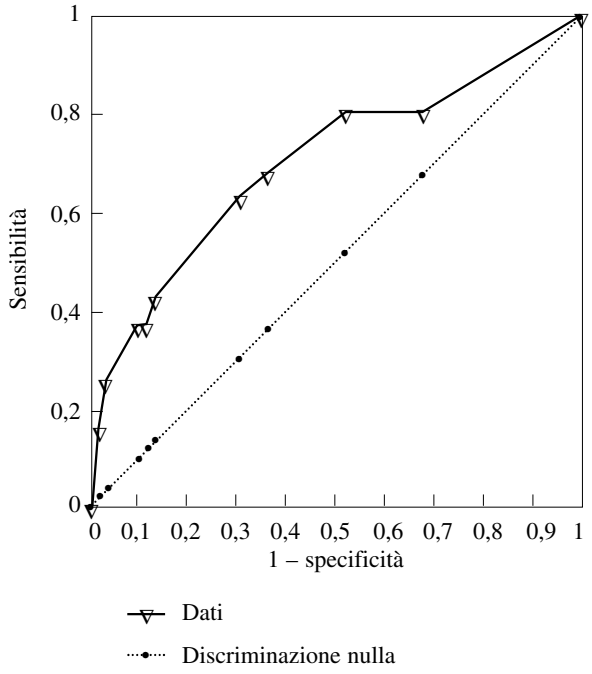


* L'effettiva prevalenza è significativamente differente dalla probabilità stimata ($p < 0{,}00005$).

Il test di affidabilità di Spiegelhalter ha prodotto una statistica χ^2 di 121,0 (10 gradi di libertà), suggerendo che le stime dei medici si sono allontanate significativamente ($p < 0{,}001$) da un livello ideale di affidabilità.

La capacità di discriminazione dei medici fra pazienti con e senza batteriemia non è stata elevata. La curva ROC ricavata dalle stime dei medici è illustrata in fig. 2. La capacità di discriminazione dei medici è indicata dalla misura in cui la curva si estende verso l'alto e verso sinistra. La «curva» ROC corrispondente a una capacità di discriminazione nulla compare come una linea retta a 45°. La curva associata ai giudizi dei medici delimita un'area di 0,687 (SE = 0,073) e indica una capacità di discriminazione non più che moderata.

Fig. 2 - Curva ROC ricavata dalle stime di probabilità di batteriemia dei medici



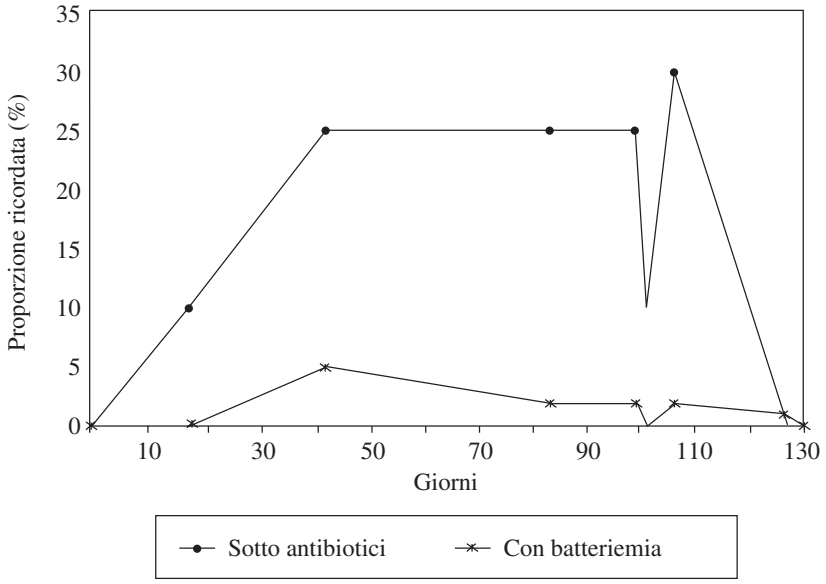
3.3. *Euristica della disponibilità e distorsione valutativa*

I risultati del nostro studio suggeriscono che i medici hanno fatto uso dell'euristica della disponibilità nel giudicare la probabilità della batteriemia. In particolare, si è osservato un legame forte e significativo fra due delle tre

variabili associate al ricordo e i giudizi intuitivi dei medici. Un paziente visto da un medico in un momento in cui quest'ultimo richiamava alla mente di aver seguito frequentemente casi di batteriemia riceveva in media una stima di probabilità più elevata (36,3%, $n = 183$) di quella assegnata in media a un paziente visto da un medico in un momento in cui quest'ultimo non richiamava alla mente di aver seguito frequentemente casi di batteriemia (22,1%, $n = 41$; $p < 0{,}0025$) (per tre pazienti non erano disponibili dati relativi a questa variabile). Si è anche rilevata una correlazione positiva fra la proporzione degli individui sotto antibiotici che i medici richiamavano alla mente fra i loro pazienti e le stime di probabilità di batteriemia dei medici riguardo ai singoli pazienti coinvolti nello studio ($r = 0{,}15$, $p < 0{,}026$). Il numero di recenti emocolture positive che i medici richiamavano alla memoria fra i loro pazienti mostra invece una correlazione debole e statisticamente non significativa con le stime di probabilità dei medici ($r = 0{,}07$).

Per controllare l'assunzione che le variabili associate al ricordo riflettessero effettivamente esperienze recenti e non credenze prefissate, abbiamo aggregato i valori di tali variabili per ciascun medico e li abbiamo disposti in funzione del tempo. La fig. 3 riporta un esempio rappresentativo dei grafici ottenuti. Come ci si sarebbe dovuti aspettare, per ciascun medico i tre fattori

Fig. 3 - Un grafico rappresentativo delle esperienze richiamate alla memoria da un singolo medico in momenti diversi. Le variabili considerate sono: la proporzione richiamata alla memoria dei propri pazienti recentemente sottoposti a terapia antibiotica e quella dei propri pazienti recentemente colpiti da batteriemia



si sono rivelati coerenti fra loro nel breve periodo e alquanto variabili nel lungo periodo.

Tab. 2 - Variabili correlate ai giudizi diagnostici

Variabile	*Condizione*				*Significatività (p)*
Variabili dicotomiche	*Media delle stime se la condizione è*				
	presente		*assente*		
Tumore ematologico maligno	53,0		32,2		< 0,05
Insufficienza renale cronica	43,2		32,7		< 0,05
Chemioterapia	44,3		32,3		< 0,05
Neutropenia	66,3		33,0		< 0,05
Brividi	44,6		31,7		< 0,01
Lesioni cutanee	45,6		31,6		< 0,005
Alterazioni dello stato mentale	40,1		31,3		< 0,05
Shock	44,6		32,5		< 0,05
Variabili categoriche	*Morente*	*Acuto*	*Cronico*	*Non malato*	
Stato clinico generale	48,6	38,1	26,8	21,1	< 0,001
	Sì	*Forse*	*No*		
Infezioni locali*	38,6	29,8	30,0		< 0,05
Infezioni genitourinarie*	44,1	25,8	32,9		< 0,05
Infezioni cutanee locali*	50,7	27,8	31,9		< 0,01
Variabili continue	*r*				
Temperatura	+0,2213				< 0,001
Frequenza respiratoria	+0,2468				< 0,001
Conteggio delle piastrine	–0,1381				< 0,05
Frequenza cardiaca**	+0,2269				< 0,001
Bicarbonato nel siero**	+0,1370				< 0,05
Sodio nel siero**	+0,1645				< 0,05
Emoglobina**	+0,1784				< 0,05
Punteggio APACHE II	+0,3043				< 0,000001

* Valutazioni dei medici.
** Valore assoluto della differenza fra i valori effettivi e quelli normali (per esempio: |frequenza cardiaca effettiva – 80|).

I risultati del nostro studio suggeriscono anche che i giudizi dei medici sono stati influenzati da una distorsione valutativa. Si è infatti rilevata una significativa correlazione inversa fra il logaritmo delle stime dei medici del ri-

schio relativo di morte da batteriemia, $ln(RR_{morte})$, e le loro stime di probabilità di batteriemia ($r = -0{,}244$, $p < 0{,}001$); il confronto con la seconda variabile valutativa, RI_{morte}, ha dato risultati analoghi ($r = -0{,}231$, $p < 0{,}001$). I medici hanno quindi assegnato stime di probabilità di batteriemia *più basse* a quei pazienti che a loro parere, nel caso fossero stati effettivamente colpiti da batteriemia, avrebbero avuto un più alto rischio di morirne durante la loro degenza in ospedale.

Nelle nostre analisi molte altre variabili sono risultate rilevanti per le stime di probabilità dei medici. Nella tab. 2 abbiamo elencato soltanto quelle che mostravano una relazione univariata con tali stime ($p < 0{,}05$, senza correzioni). Alcuni dei fattori considerati possono essere plausibili indicatori della presenza di infezioni severe, come ad esempio brividi, lesioni cutanee, shock, alterazioni dello stato mentale, temperatura, frequenza del battito cardiaco, frequenza del respiro o dinimuzione del livello di piastrine. Altre variabili, come il punteggio APACHE II, possono essere correlate con una maggiore suscettibilità alle infezioni, a una condizione clinica generalmente negativa, o a severe alterazioni fisiologiche dovute a sepsi. Dal momento che abbiamo svolto analisi relative a numerose variabili potenzialmente rilevanti (approssimativamente 100), alcuni di questi test statistici potrebbero essere falsamente positivi. La correzione di Bonferroni richiede un valore p inferiore a $0{,}05/100 = 0{,}0005$ come criterio di significatività statistica. Solo la relazione fra il punteggio APACHE II e le stime di probabilità soddisfa questo criterio restrittivo.

Eravano particolarmente preoccupati della possibilità che la relazione fra il rischio di morte da batteriemia e le stime di probabilità di batteriemia potesse essere falsata dalla severità delle condizioni cliniche soggiacenti. I pazienti con severe patologie croniche possono risultare più suscettibili a infezioni rispetto a pazienti con un minor grado di comorbidità. Pazienti di questo tipo potevano dunque avere bassi valori di RR_{morte} per il fatto che il loro rischio di morte era già alto anche in assenza di batteriemia. Le nostre analisi multivariate suggeriscono però che non è possibile ricondurre a questo fattore gli effetti attribuiti all'euristica della disponibilità e alla distorsione valutativa. In una serie di modelli multivariati che impiegavano diverse combinazioni di variabili associate al ricordo, variabili valutative e variabili cliniche, le variabili associate al ricordo e le variabili valutative hanno mantenuto un legame indipendente e significativo con i giudizi dei medici.

*Tab. 3 - Variabili correlate con le stime probabilistiche dei medici in base alla regressione lineare multipla**

Variabile	*Coefficiente*	*Errore standard*	*Coefficiente standardizzato*	*Valore p*
Logaritmo di RR_{morte}	–3,1	1,4	–0,15	0,02
Neutropenia	29,2	11,5	0,15	0,01
Antibiotici	2,2	4,8	0,03	0,7
APACHE II	0,81	0,23	0,22	0,001
Infezioni locali	1,0	3,7	0,02	0,8
Tumore ematologico maligno o chemioterapia	12,9	4,5	0,17	0,001
Infezioni genitounrinarie o cutanee	5,4	4,8	0,08	0,1
Disponibilità	12,9	5,0	0,16	0,01
Brividi o lesioni cutanee	12,1	3,7	0,21	0,001

* $r^2 = 0,29$.

La tab. 3 riporta un esempio rappresentativo di questo tipo di analisi. Nel modello presentato le variabili associate ai ricordi e le variabili valutative erano significativamente correlate con i giudizi probabilistici, indipendentemente da diversi altri fattori che riflettevano la severità delle condizioni cliniche e/o la suscettibilità a infezioni. I coefficienti standardizzati delle variabili rappresentano il contributo relativo delle diverse variabili alle predizioni derivate da questo modello di regressione riguardo alle stime probabilistiche dei medici. I valori dei coefficienti standardizzati della variabile valutativa (0,15) e della variabile associata al ricordo (0,16) indicano che tali variabili contribuiscono ampiamente alle predizioni ricavate dal modello. Complessivamente, questo modello era in grado di spiegare il 29% della varianza dei giudizi dei medici.

4. Discussione

I medici si trovano in difficoltà nel diagnosticare la batteriemia. Gli specializzandi coinvolti nel nostro studio hanno ampiamente sovrastimato la probabilità della batteriemia e hanno mostrato una capacità di discriminazione soltanto moderata fra i pazienti che ne erano affetti e quelli che non lo erano. Questi risultati sono simili a quelli di studi precedenti (Makadon *et al.*, 1987; Mellors *et al.*, 1987) e suggeriscono l'esistenza di un problema generale nel modo in cui i medici elaborano questi giudizi diagnostici.

Molti autori hanno sostenuto che i medici dovrebbero prendere le decisioni di trattamento per pazienti con diagnosi incerte sulla base dei loro giudizi in-

tuitivi di probabilità. Per pazienti con un sospetto di batteriemia, un medico potrebbe considerare la probabilità di questa patologia per decidere se ordinare delle emocolture, se iniziare un trattamento antibiotico e quali antibiotici impiegare. I nostri risultati suggeriscono però che, se i medici utilizzassero a questo scopo i loro giudizi di probabilità, potrebbero essere indotti a prendere decisioni sbagliate. La loro modesta capacità di discriminazione potrebbe condurre a un uso di antibiotici insufficiente per pazienti che sono realmente infetti e a un uso eccessivo per pazienti che non lo sono. In particolare, la loro frequente sovrastima della probabilità di batteriemia potrebbe condurre a un impiego sproporzionato di regimi farmacologici tossici e dispendiosi, composti da molteplici agenti antibiotici, per pazienti che in realtà non soffrono di infezioni complesse o rare.

I nostri dati suggeriscono che i medici fanno affidamento sull'euristica della disponibilità nel tentativo di diagnosticare la batteriemia.

L'euristica della disponibilità può in molte occasioni condurre a giudizi di buona qualità. Per eventi fra loro legati da nessi causali, come i diversi casi di una malattia contagiosa, l'euristica della disponibilità può permettere di riconoscere tempestivamente reali cambiamenti nella prevalenza della malattia. Ed è possibile che un individuo richiami più facilmente alla memoria eventi comuni che non eventi rari. Tuttavia, con poche eccezioni, la batteriemia non è contagiosa. L'euristica della disponibilità può quindi indurre a ritenere erroneamente che uno o più casi recenti di batteriemia, che in realtà riflettono fluttuazioni casuali, preannuncino un effettivo mutamento nella prevalenza della malattia. A ciò si deve aggiungere che eventi importanti o drammatici ma inusuali (come appunto un caso di batteriemia) possono risultare più facili da richiamare alla mente che non eventi meno salienti ma più comuni (come un caso di polmonite). L'euristica della disponibilità può pertanto spingere a considerare come più comuni condizioni cliniche che sono più serie o più salienti di altre. Taluni lamentano in effetti che gli specializzandi tendono a «cercare le zebre» (anziché i cavalli), cioè malattie rare ed esotiche. Indagare sulla possibilità di un problema clinico raro e curabile è certamente giustificabile. L'euristica della disponibilità può però anche suggerire indagini cliniche di scarsa utilità su condizioni improbabili.

I nostri dati suggeriscono inoltre che i medici possono mostrarsi riluttanti a diagnosticare la batteriemia in pazienti nei quali tale condizione produrrebbe un maggior rischio di esiti avversi (come la morte prematura). Si tratta di una forma di distorsione valutativa del giudizio chiamata «*wishful thinking*» (Hogarth, 1987).

Una distorsione valutativa raramente produce buoni giudizi, ma può talvolta dar luogo a buone decisioni. Si prenda il caso in cui si presenta una distorsione valutativa nella direzione *opposta* a quella che abbiamo qui osservato, vale a dire una tendenza ad assegnare stime di probabilità *più alte* a quei pa-

zienti che rischiano maggiormente un esito avverso a causa della malattia. Questa tendenza può suggerire decisioni adeguate, anche se per le ragioni sbagliate. Per esempio, può essere del tutto ragionevole iniziare un trattamento antibiotico nei pazienti più a rischio di morte da batteriemia prima che siano disponibili i risultati delle emocolture. L'esperto di analisi delle decisioni giustificherebbe una scelta di questo tipo osservando che, per pazienti ad alto rischio di morte da batteriemia, si dovrebbe impiegare una soglia di probabilità più bassa per individuare i casi in cui è opportuno prescrivere un trattamento antibiotico. La distorsione valutativa cui ci stiamo riferendo potrebbe determinare questa stessa decisione inducendo il medico, non ad abbassare (opportunamente) la soglia di probabilità minima per avviare il trattamento, ma piuttosto a sovrastimare (erroneamente) la probabilità della malattia per il paziente in questione.

Inizialmente ci aspettavamo di osservare *quest'ultima* forma di distorsione valutativa nel nostro studio. Disponevamo in precedenza di alcuni risultati equivoci riguardo alla possibilità che una tendenza di questo genere influenzasse i medici nel giudicare la probabilità della faringite da streptococchi in pazienti con mal di gola (Poses *et al.*, 1985). Wallsten aveva mostrato che questa stessa tendenza influenzava i radiologi nel giudicare la probabilità che gli esiti di una pielografia endovenosa indicassero la presenza di un tumore (Wallsten, 1981). Siamo stati quindi sorpresi nel rilevare la presenza del *wishful thinking*, cioè della tendenza ad assegnare stime di probabilità *più basse* di una certa malattia ai pazienti ritenuti maggiormente a rischio di un esito clinico avverso a causa di quella malattia. La nostra sorpresa ci ha spinti a considerare l'interferenza di altri fattori come una possibile spiegazione alternativa dei nostri dati. Estese analisi statistiche con numerose variabili non hanno però rivelato tale interferenza.

Può essere arduo per i medici evitare il *wishful thinking*. Far fronte a una nuova complicanza in un paziente già malato può essere difficile. Per uno specializzando già affaticato dai ritmi lavorativi e con un notevole investimento emotivo sul benessere del paziente le difficoltà possono essere ancora maggiori. La letteratura sulla psicologia della decisione suggerisce che il *wishful thinking* può emergere dalla percezione dell'incapacità di evitare un evento negativo o di migliorarne le conseguenze, e da una forte identificazione con le persone che potrebbero soffrirne (Zakay, 1983). Gli specializzandi coinvolti in questo studio potrebbero aver dubitato dell'utilità degli antibiotici per i loro pazienti più severamente malati nel caso fossero stati affetti da batteriemia; potrebbero inoltre essersi fortemente identificati con tali pazienti dopo avere diligentemente lavorato al loro servizio. Il *wishful thinking* può quindi ritardare il riconoscimento della malattia proprio nei pazienti che rischiano maggiormente per le sue possibili conseguenze.

Non sono molti, finora, gli studi che mostrano l'influenza degli errori cognitivi o dell'impiego di euristiche inappropriate sui giudizi dei medici (Dawson, Arkes, 1987). Tra gli stessi psicologi si è discusso se gli errori di giudizio riscontrati negli studi sperimentali influenzino in misura rilevante i giudizi probabilistici degli individui in situazioni familiari e nelle importanti decisioni del mondo reale (Christensen-Szalanski, 1986; Christensen-Szalanski, Fobian, 1988). Molti dei dati che indicano specifiche debolezze nella cognizione umana provengono infatti da studi condotti in forma scritta in cui studenti che non hanno forti incentivi a rispondere in modo corretto esprimono giudizi probabilistici riguardo a eventi poco familiari. Sebbene Kassirer e Kopelman (Kassirer, Kopelman, 1989) abbiamo riportato un aneddoto che suggerisce l'impiego inappropriato dell'euristica della disponibilità da parte di un medico nel caso di un singolo paziente, e sebbene alcuni esempi di *wishful thinking* siano comparsi nella letteratura psicologica ed economica (Hogarth, 1987; Morlock, 1967), nessuno aveva finora mostrato che l'euristica della disponibilità o il *wishful thinking* potessero avere effetti frequenti o importanti sulla cura dei pazienti da parte dei medici.

È opportuno, d'altra parte, considerare alcuni limiti del presente studio. Abbiamo studiato soltanto pazienti per i quali i medici avevano prescritto emocolture. È quindi possibile che alcuni pazienti batteriemici siano rimasti esclusi, anche se dubitiamo che possa trattarsi di un numero significativo. La prevalenza della batteriemia nella popolazione di pazienti coinvolti era ridotta e questo suggerisce che i medici hanno adottato una soglia piuttosto bassa per stabilire se ordinare le emocolture. Del resto, agli specializzandi viene insegnato di non esitare a ordinare emocolture (Lipsky, 1988). In uno studio recentemente pubblicato, per esempio, gli specializzandi di un ospedale universitario hanno prescritto emocolture per il 90% dei loro pazienti con febbre (Filice *et al*., 1989).

Abbiamo studiato stime relative alla probabilità che almeno una coltura sarebbe risultata positiva entro un periodo di 24 ore. Abbiamo identificato i risultati delle colture, e non la «batteriemia», come l'evento da giudicare per il fatto che i partecipanti potevano non condividere una comune definizione di «batteriemia». Abbiamo valutato la qualità dei giudizi impiegando due diverse definizioni di «emocoltura positiva». Per minimizzare l'impatto delle emocolture falsamente positive, una delle due era una definizione convenzionale e restrittiva con la quale si intende escludere le possibili contaminazioni (Haley *et al*., 1980). È quindi possibile che alcuni risultati «negativi» fossero in effetti falsi negativi, specialmente quando abbiamo impiegato questa definizione restrittiva di posivitità del test. Dubitiamo, però, che si possa trattare di un numero significativo. Alla maggioranza (85%) dei pazienti coinvolti è stato infatti prelevato sangue da impiegare in una o più serie di emocolture entro 24 ore, e la sensibilità di una singola serie di emocolture è dell'80%-91,5%, quella di due serie dell'89%-99,3% (Aronson, Bor, 1987).

La nostre analisi basate sulla regressione lineare multipla hanno spiegato soltanto il 29% della varianza nei giudizi dei medici. Il tipo di modelli ricavati non rende conto delle differenze individuali fra i medici, di giudizi incoerenti espressi da singoli partecipanti e degli effetti di eventuali variabili cliniche diverse da quelle che sono state misurate. Inoltre, la costruzione di modelli lineari implica alcune assunzioni riguardo alla distribuzione dei dati che il nostro insieme di dati potrebbe non soddisfare. Tuttavia, i «modelli a lente» di Bruswik sono lo strumento convenzionale per la rappresentazione del modo in cui le persone soppesano molteplici elementi di informazione per elaborare un giudizio, anche in assenza della normalità multivariata (Wigton, 1988).

Nelle nostre analisi abbiamo aggregato i giudizi di diversi medici, determinando il peso dei giudizi di ciascuno in funzione del numero dei pazienti di quel particolare medico che sono stati inclusi nello studio. Questo tipo di analisi non avrebbe potuto rilevare differenze fra i singoli medici nella loro suscettibilità alla distorsione valutativa e all'uso dell'euristica delle disponibilità. Esistono probabilmente differenze di questo tipo; ma, dal momento che nessun singolo medico ha contribuito in misura sproporzionata (cioè in più del 9% dei pazienti), i nostri risultati suggeriscono che i due fenomeni hanno riguardato almeno una minoranza significativa e, più probabilmente, la maggioranza dei medici.

Abbiamo osservato che le variabili associate al ricordo e le variabili valutative erano predittori forti e indipendenti delle stime probabilistiche dei medici. Sulla base delle nostre conoscenze fisiopatologiche, tali variabili non dovrebbero essere considerate rilevanti per l'effettiva presenza della batteriemia. I nostri dati suggeriscono quindi che fare affidamento su questi elementi estranei ha peggiorato la qualità dei giudizi probabilistici dei medici, sebbene tali dati non permettano di dimostrare in maniera definitiva questa asserzione di tipo causale.

Abbiamo studiato un gruppo di specializzandi nella seconda metà del loro primo anno di corso. I nostri risultati potrebbero pertanto non essere generalizzabili a medici più esperti. Gli specializzandi coinvolti erano però incaricati della valutazione e del trattamento iniziale dei loro pazienti, quindi la qualità dei loro giudizi dovrebbe risultare rilevante per l'appropriatezza delle cure offerte a quei pazienti. Inoltre, la sola esperienza clinica può non essere sufficiente a superare limitazioni di tipo cognitivo: si è infatti osservato che anche medici esperti possono elaborare giudizi inaffidabili (Poses *et al.*, 1985, 1989). Infine, è da notare che la formazione medica a livello universitario, post-universitario e professionale include raramente, o non include affatto, tecniche specifiche per il miglioramento dei giudizi probabilistici.

Le nostre variabili valutative e quelle associate ai ricordi potrebbero non riflettere perfettamente i fattori che i medici hanno considerato nei loro processi mentali. È materia di dibattito in che misura l'euristica delle disponibilità sia mediata dalla vivacità di specifici ricordi (Shedler, Manis, 1986). Per evitare

di provocare l'impiego dell'euristica delle disponibilità abbiamo ridotto al minimo l'impatto delle domande poste ai medici. In questo modo, è possibile che le variabili impiegate siano in effetti indicatori imperfetti della presenza di ricordi vivaci che influenzano i giudizi dei medici. Si osservi però che, nel presente studio, l'impiego di variabili di minimo impatto poteva soltanto rendere più difficile rilevare significativi effetti dell'euristica delle disponibilità. Si deve inoltre notare che una delle variabili, cioè il numero di pazienti batteriemici che i medici potevano richiamare alla memoria, potrebbe aver risentito del carico di lavoro affidato ai diversi specializzandi; il che spiegherebbe forse la mancanza di un legame statisticamente significativo con i giudizi probabilistici. Infine, abbiamo considerato una sola dimensione valutativa, cioè il rischio di morte durante il ricovero. Una misurazione più complessa di questo fattore, che comprendesse altri possibili esiti della batteriemia, avrebbe forse messo in rilievo ulteriori aspetti della distorsione valutativa.

In conclusione, gli errori cognitivi e l'uso inappropriato di euristiche possono influenzare i giudizi diagnostici intuitivi dei medici. È possibile esplorare diverse strade per sviluppare migliori strategie di giudizio. La soluzione più semplice per lo specifico problema clinico che abbiamo trattato potrebbe essere un test per la batteriemia rapido, discriminante, non invasivo ed economico. Al momento attuale non è però disponibile e, se anche lo fosse, i medici dovrebbero comunque affidarsi al loro giudizio per decidere quando prescriverne l'esecuzione. In alternativa, i medici potrebbero elaborare i propri giudizi con l'ausilio di un valido modello multivariato per la diagnosi di batteriemia. Si è infatti mostrato che mettere al corrente i medici delle variabili impiegate in un modello e delle predizioni da esso ricavate può migliorare i loro giudizi diagnostici in casi reali (Poses *et al.*, 1992). Se i modelli finora proposti per la diagnosi della batteriemia (Mellors *et al.*, 1987; Bates *et al.*, 1990) dovessero rivelarsi validi e generalizzabili, questa soluzione diverrebbe praticabile. Oltre a ciò, si potrebbe forse stabilire di combinare matematicamente i giudizi indipendenti di diversi medici che seguono lo stesso paziente. Questo processo, potenzialmente in grado di neutralizzare almeno in parte errori di giudizio individuali, si è mostrato efficace nel migliorare giudizi prognostici riguardo a pazienti in condizioni critiche (Poses *et al.*, 1990). Infine, è possibile che una sufficiente conoscenza della psicologia del giudizio e della decisione induca nei medici la necessaria consapevolezza dei limiti cognitivi che influenzano il giudizio umano e li aiuti così a evitare strategie cognitive inappropriate. Ulteriori ricerche in tutti questi ambiti potrebbero quindi migliorare i giudizi dei medici, le loro decisioni e la salute dei pazienti.

Gli autori sono in debito con Fiore J. Copare per l'aiuto nella raccolta dei dati, e con John Holmes per l'assistenza nelle analisi statistiche. Teri Thorowgood ha contribuito nel lavoro di segreteria.

5. *L'influenza di informazioni pseudodiagnostiche nella valutazione dell'ischemia miocardica**

di *Lee A. Green, J. Frank Yates*

1. Introduzione

Negli ultimi trent'anni gli studi epidemiologici hanno individuato alcuni utili indicatori per la previsione del rischio a lungo termine di patologie coronariche (Kannel, 1976; Stamler, 1973; Harris *et al*., 1988). Questi classici fattori di rischio comprendono: storia familiare di patologie coronariche precoci, età, sesso maschile, fumo di sigarette, diabete mellito, crescita dei livelli di colesterolo del siero e ipertensione (American Heart Association, 1976). I medici sono stati sensibilizzati dalla letteratura e dalla loro formazione professionale a sospettare la presenza di ischemia miocardica nei pazienti in cui ricorrono questi fattori di rischio.

Più recentemente, studi volti a discriminare fra pazienti con e senza ischemia miocardica acuta in reparti di medicina di emergenza hanno mostrato che la natura e la localizzazione dei sintomi, la storia personale di patologie ischemiche e alcuni specifici riscontri elettrocardiografici sono di gran lunga gli indicatori più efficaci per predire l'ischemia acuta, l'infarto e la mortalità (Goldman *et al*., 1982; Goldman *et al*., 1988; Pozen *et al*., 1984; Selker, Griffith, D'Agostino, 1991a e 1991b; Lee *et al*., 1992; Solomon *et al*., 1989). Jayes *et al*. (1992) hanno sottolineato la differenza fra i fattori rilevanti per la valutazione del rischio a lungo termine e quelli utili per le decisioni cliniche in condizioni di emergenza. Indagando i fattori predittivi dell'ischemia coronarica acuta nei reparti di emergenza, essi hanno riscontrato che la valutazione dei classici fattori di rischio di tipo epidemiologico incrementava di poco (fra gli uomini) o per nulla (fra le donne) il potere predittivo delle sole informazioni relative ai sintomi e ai risultati elettrocardiografici.

* Traduzione di *Influence of pseudiognostic information on the evaluation of ischemic heart disease*, «Annals of emergency medicine», 25 (1995), pp. 451-457, copyright © 1995 American College of Emergency Physicians.

A oggi le ricerche indicano che la considerazione dei fattori di rischio classici limita la validità delle previsioni di ischemia acuta nei reparti di emergenza. Tali ricerche non indagano però in modo esplicito se e in che misura questi fattori permettano di prevedere le decisioni dei medici. In base all'enfasi che hanno ricevuto nella letteratura e nella formazione medica, i fattori di rischio classici possono influenzare le decisioni dei medici nei reparti di emergenza a dispetto della loro scarsa validità nel prevedere l'ischemia miocardica acuta.

Il presente studio mira a determinare se i classici fattori di rischio di tipo epidemiologico agiscono sulle decisioni dei medici dei reparti di emergenza in casi di sospetta ischemia miocardica. In particolare, abbiamo ipotizzato che una storia di ipertensione o di diabete (rilevata dalle annotazioni nelle cartelle cliniche dei pazienti) avesse un valore indipendente nel prevedere le decisioni di ricovero dei medici in un modello multivariato che comprende anche quei fattori (ricavati dalla storia clinica e dall'elettrocardiogramma) che predicono in modo accurato l'ischemia acuta. I dati a disposizione nel presente studio non hanno permesso di valutare l'influenza di fumo di sigarette, livelli di colesterolo e storia familiare.

2. Materiali e metodi

Lo studio qui presentato consiste in un'analisi retrospettiva delle cartelle cliniche di pazienti presentatisi nel 1988 nei reparti di emergenza di due ospedali non universitari nel nord della Penisola Inferiore del Michigan. Il primo ospedale ha un bacino di utenza di 45.700 persone, il secondo di 44.920 persone (calcolato in base al «metodo di raggruppamento per quote di mercato»; si veda Griffith *et al.*, 1985). Nessuna delle due comunità ospita minoranze etniche di dimensioni significative. Entrambe le comunità sono costituite in prevalenza da operai e sono caratterizzate da un'economia incentrata sulle manifatture e le risorse naturali. Il primo ospedale ha 110 posti letto, il secondo 113. I posti disponibili in unità per patologie coronariche sono rispettivamente 9 e 8.

Le decisioni di ricovero considerate nello studio sono state prese da 19 medici di reparti di emergenza. Tutti i medici in questo gruppo erano abilitati a livello nazionale in medicina d'urgenza, in medicina interna o in medicina di famiglia, e il periodo intercorso dalla laurea variava fra i 5 e i 19 anni. Abbiamo considerato soltanto le decisioni iniziali relative al ricovero, nelle quali erano stati coinvolti esclusivamente medici dei reparti di emergenza. I nostri dati non riflettono quindi il possibile coinvolgimento di altri specialisti (come cardiologi o chirurghi) in questo tipo di decisioni. In servizi ospedalieri di piccole comunità, come quelli qui studiati, il ricorso a consulti specialistici è

piuttosto limitato. Nei reparti di emergenza non sono presenti medici di diverse specialità e non è pratica comune effettuare un consulto prima di una decisione di ricovero.

Nessuno dei due ospedali è attrezzato per la cateterizzazione cardiaca o la chirurgia toracica. Entrambi gli ospedali sono gli unici nelle rispettive comunità e, per ragioni geografiche, è praticamente certo che un paziente con una sintomatologia acuta in una di queste due aree si presenterà nel relativo ospedale. Le stesse ragioni di ordine geografico assicurano che tutti i pazienti che si presentano in uno di questi due reparti di emergenza verranno valutati all'interno di quella struttura. Non si fa riferimento ad altri centri prima di una valutazione iniziale e della stabilizzazione delle condizioni del paziente.

Un laureando in medicina coinvolto nella ricerca ha esaminato i registri dei reparti di emergenza dei due ospedali e ha compilato una lista di tutti i pazienti presentatisi nel 1988 con sintomi che potevano indicare la presenza di ischemia miocardica acuta (si veda la tab. 1). Abbiamo quindi ottenuto le relative cartelle cliniche e raccolto informazioni sui pazienti per i quali i medici avevano effettivamente valutato la possibilità di un'ischemia miocardica acuta. I dati raccolti da ciascuna cartella clinica comprendevano: età, genere, storia di precedenti patologie, disturbi inizialmente riportati, risultati del primo ECG, decisioni di trattamento nel reparto di emergenza (e in ospedale, per i pazienti successivamente ricoverati), condizioni cliniche e diagnosi al momento del congedo.

La raccolta e l'inserimento dei dati si sono svolti in un unico passo (evitando errori di trascrizione dovuti all'uso di documenti di codifica) e utilizzando moduli, criteri e sistemi di controllo degli errori definiti attraverso la programmazione di un data-base relazionale su un computer portatile. Le procedure e i criteri di raccolta dei dati sono stati sviluppati nel dipartimento del primo autore come parte di un progetto di ricerca in corso riguardo alle decisioni dei medici in casi di sospetta ischemia miocardiaca acuta e derivano dal lavoro di Selker, Griffith, D'Agostino (1991a) (si veda Green, Becker, 1994, per una descrizione dettagliata). La storia familiare, i livelli di colesterolo e il fumo di sigarette non sono compresi nell'insieme dei dati raccolti. Non è stato possibile includere fra le variabili la storia familiare e i livelli di colesterolo perché solo raramente comparivano nelle annotazioni dei medici dei reparti di emergenza. In uno studio pilota si è inoltre osservato che il fumo di sigarette veniva annotato in cartella clinica in meno della metà dei casi.

La raccolta dei dati dalle cartelle cliniche è stata limitata ai pazienti per i quali i medici avevano considerato il ricovero a causa di sospetta ischemia miocardica acuta. L'espressione «ischemia miocardica acuta» non si identifica con i casi in cui si deve innanzi tutto cercare di escludere l'infarto miocardico («*rule out myocardial infarction*»), ma comprende tutte le condizioni caratterizzate da ischemia acuta del miocardio, come angina instabile o angina

in crescendo, così come la presenza di un infarto miocardico in corso. Si è ritenuto che l'ischemia miocardica acuta fosse stata presa in considerazione per un particolare paziente se le note del medico del reparto di emergenza lo indicavano esplicitamente, se erano stati ordinati isoenzimi CK o se il paziente era stato sottoposto a elettrocardiogramma (ECG) per sospetta ischemia miocardica acuta. Abbiamo escluso i pazienti per i quali l'ischemia miocardica acuta era stata considerata, ma che in base alle annotazioni raccolte nel reparto di emergenza dovevano essere ricoverati per qualche motivo indipendente (come crisi ipertensiva, aggravamento di un'ostruzione polmonare cronica o diabete fuori controllo). In questi casi, infatti, la decisione del ricovero non dipendeva dalla stima del medico della probabilità di ischemia miocardica acuta. Abbiamo infine codificato come decisioni di ricovero i casi dei pochi pazienti che dopo la stabilizzazione delle loro condizioni erano stati trasferiti in altri centri. Non sono stati utilizzati altri criteri di esclusione.

Tab. 1 - Termini associati a sospetta ischemia miocardica acuta nelle annotazioni dei reparti di emergenza

Angina
Angina pectoris
Angina instabile
Dolore al petto
Compressione del petto
Oppressione del petto
Dolore o compressione che si irradia al braccio sinistro o alla mascella
Shock cardiogenico
MI («*myocardial infarction*»)*
Infarto miocardico
MI subendocardiaco
ROMI O R/O MI («*rule out myocardial infartction*»)*
*Rule out myocardial infarction**
Rule out MI*
Attacco cardiaco
AIHD («*acute ischemic heart disease*»)*
CVD («*cardiovascular disease*»)*
Patologia coronarica
Patologia coronarica acuta
Ischemia coronarica acuta
Patologia vascolare (occlusiva) coronarica
Patologia cardiovascolare
Angina di Prinzmetal
(Vaso)spasmo coronarico
Spasmo arterocoronarico
Occlusione arterocoronarica

* Sono stati conservati gli acronimi e le originarie espressioni inglesi [N.d.T.].

Abbiamo innanzi tutto effettuato una serie di semplici analisi univariate con test χ^2 dei rapporti di verosimiglianza (*likelihood ratio*) in confronti fra variabili categoriche 2 × 2. Abbiamo inoltre impiegato la statistica di Kolmogorov-Smirnov per la variabile «età».

Per controllare la nostra ipotesi di studio abbiamo utilizzato la regressione logistica (Afifi, 1990). Il modello impiegato era una formula logistica standard con soli termini lineari (effetti principali). Sarebbe stato interessante includere termini relativi a possibili interazioni (per esempio, al fatto che le alterazioni dell'ECG possano aver esercitato una maggiore influenza sia per i pazienti più anziani che per quelli più giovani), ma l'elaborazione di stime stabili per questi termini avrebbe richiesto un numero di pazienti molto superiore a quello disponibile nella presente indagine. Le variabili indipendenti erano: età, genere, dolore al petto come principale disturbo iniziale, storia di infarti cardiaci, onde Q, anomalie del tratto ST e delle onde T nel primo ECG, storia di diabete mellito e storia di ipertensione. L'età è stata inclusa come quantità scalare. Le altre variabili sono state incluse in forma dicotomica, come presenti o assenti. Le informazioni relative agli ECG sono state ricavate dai tracciati (non dall'interpretazione che i medici ne avevano dato), utilizzando i metodi e i criteri definiti da Selker, Griffith, D'Agostino (1991a). Le variabili dicotomiche sono state codificate come presenti («1») o assenti («0»). Dati equivoci, come un sovraslivellamento del tratto ST di meno di 1 mm, sono stati codificati con il valore «0,5». La variabile dipendente era costituita dalle decisioni dei medici del reparto di emergenza in favore del ricovero («1») o del congedo («0»).

Un possibile limite teorico per la validità del nostro modello era rappresentato da eventuali multicollinearità fra fattori di rischio classici (diabete e ipertensione) e alterazioni elettrocardiografiche; per questo motivo, abbiamo calcolato e analizzato una matrice di covarianza delle variabili indipendenti.

Infine, in un modello basato sullo stesso insieme di fattori, abbiamo applicato la regressione logistica per la previsione di due ulteriori variabili dipendenti: la diagnosi di infarto cardiaco al momento del congedo e la morte in ospedale.

3. Risultati

Nei registri dei reparti di emergenza sono stati identificati in tutto 1.062 pazienti per i quali erano disponibili cartelle cliniche complete. Ottantatre ulteriori cartelle erano del tutto irreperibili o incomplete in relazione a documenti essenziali, come le annotazioni iniziali nel reparto di emergenza. I dati mancanti riflettevano errori casuali di registrazione e archiviazione; non si è identificata alcuna causa sistematica di rimozione o di occultamento in nessu-

na delle due strutture coinvolte nello studio. Le uniche informazioni universalmente disponibili erano l'età e il genere dei pazienti. L'età media dei pazienti le cui cartelle cliniche erano mancanti era leggermente ma non significativamente inferiore a quella degli altri pazienti (62,4 *vs*. 59,9, $t = 1{,}28$, $p > 0{,}20$). Le cartelle mancanti mostravano inoltre una distribuzione uniforme in relazione al genere (8,9% fra le donne *vs*. 9,8% fra gli uomini, $\chi^2_{(1)} = 0{,}22$, $p > 0{,}5$). Fra le 1.062 cartelle cliniche complete, 787 appartenevano a pazienti per i quali era stata valutata principalmente o esclusivamente una diagnosi di ischemia miocardica acuta, e che pertanto sono stati inclusi nello studio. La distribuzione di questi pazienti in termini di età, genere e variabili cliniche è riportata in tab. 2.

Tab. 2 - Statistiche descrittive

Totale	787
Donne	328 (41,7%)
Ricoverati in ospedale	512 (65,1%)
Dolore al petto come principale disturbo iniziale	646 (82,1%)
Storia di infarti miocardici	217 (27,6%)
Storia di ipertensione	220 (28,0%)
Diabete mellito	121 (15,4%)
Diagnosi di infarto miocardico al momento del congedo	110 (14,0%)
Morti in ospedale	40 (5,1%)
Onde Q	
Risultati anomali	179 (27,7%)
Risultati non specifici	91 (11,6%)
Alterazioni del tratto ST	
Risultati anomali	223 (28,3%)
Risultati non specifici	200 (25,4%)
Onde T	
Risultati anomali	239 (30,4%)
Risultati non specifici	40 (5,1%)

Le associazioni univariate fra variabili indipendenti categoriche e decisioni di ricovero sono presentate in tab. 3. I pazienti ricoverati erano di età più avanzata di quelli non ricoverati (età media, rispettivamente: 64,4 e 58,5; *D* Kolmogorov-Smirnov = 0,236; $p < 0{,}001$).

La regressione logistica mostra che, in relazione al ricovero, il rapporto incrociato aggiustato (*adjusted odds ratio*) era più alto per pazienti con una storia di ipertensione (8,00; intervallo di fiducia al 95%: 4,64-13,80) che per pazienti con sovra- o sottoslivellamento del tratto ST (3,78; intervallo di fiducia al 95%: 2,41-5,93) o con una storia di infarti cardiaci (2,13; intervallo di fiducia al 95%: 1,34-3,38). Anche una storia di diabete ha influenzato in modo

significativo le decisioni di ricovero (rapporto incrociato = 1,87; intervallo di fiducia al 95%: 1,02-3,44). Sebbene le anomalie relative alle onde Q e alle onde T siano risultate significativamente rilevanti nell'analisi univariata, non hanno raggiunto la significatività statistica nel modello logistico; nessuna delle due variabili aveva quindi un potere predittivo indipendente in relazione alle decisioni di ricovero (onde Q: rapporto incrociato = 1,13; intervallo di fiducia al 95%: 0,70-1,84; onde T: rapporto incrociato = 1,16; intervallo di fiducia al 95%: 0,77-1,72). Anche il dolore al petto come principale disturbo iniziale non ha raggiunto la significatività (rapporto incrociato = 1,35; intervallo di fiducia al 95%: 0,86-2,13). L'età era positivamente correlata con le decisioni di ricovero (peso fattoriale: 0,017 ± 0,0065, $p < 0{,}001$). La matrice di covarianza delle variabili indipendenti indica che i risultati delle presenti analisi non risentono dei rapporti fra fattori di rischio classici e alterazioni dell'ECG; gli indici di covarianza sono infatti compresi fra 0,0052 e 0,00054.

Tab. 3 - Associazioni univariate delle variabili categoriche indipendenti con le decisioni di ricovero

Variabile	*Ricoveri (%)*
Donne *vs*. uomini	63,4 *vs*. 66,2
Dolore al petto come disturbo principale *vs*. altri disturbi principali	65,6 *vs*. 62,4
Precedenti di infarto miocardico *vs*. nessun precedente noto	82,5 *vs*. 57,5*
Presenza di diabete *vs*. assenza di diabete	86,0 *vs*. 61,3*
Presenza di ipertensione *vs*. assenza di ipertensione	92,3 *vs*. 54,5*
Onde Q *vs*. scarsa progressione dell'onda R *vs*. assenza di onde Q	80,0 *vs*. 65,9 *vs*. 59,8*
Anomalie del tratto ST *vs*. risultati non specifici *vs*. risultati normali	86,6 *vs*. 67,5 *vs*. 50,6*
Anomalie delle onde T *vs*. risultati non specifici *vs*. risultati normali	74,1 *vs*. 60,0 *vs*. 61,2*

* $p < 0{,}01$ nel test χ^2 (massima verosimiglianza).

Abbiamo rielaborato il modello includendo termini di interazione fra i fattori di rischio classici e i fattori relativi alla condizione acuta per controllare la possibilità che i primi potessero aver influenzato le decisioni dei medici solo quando i secondi erano assenti o equivoci. Nessuno dei termini di interazione è però risultato significativo e il modello che li includeva non ha raggiunto una maggiore accuratezza nel prevedere le decisioni dei medici.

La regressione logistica applicata allo stesso insieme di variabili indipendenti per la previsione della diagnosi di infarto miocardico acuto al momento del congedo ha rivelato come significative soltanto le anomalie del tratto ST (rapporto incrociato = 12,06; intervallo di fiducia al 95%: 6,16-23,62) e il dolore al petto come principale disturbo iniziale (rapporto incrociato = 3,63; in-

tervallo di fiducia al 95%: 1,63-8,10). Si è osservata una associazione negativa non statisticamente significativa fra ipertensione e diagnosi di infarto cardiaco al momento del congedo (rapporto incrociato = 0,66; intervallo di fiducia al 95%: 0,40-1,08).

La regressione per la previsione dell'ultima variabile dipendente considerata, la morte in ospedale, non è riportata. Si sono verificate in tutto soltanto 40 morti, una proporzione insufficiente per rendere significativa l'analisi del modello.

4. Discussione

Le decisioni di ricovero dei medici coinvolti nello studio sembrano essere state fortemente influenzate da informazioni prive di un genuino valore diagnostico e prognostico: ipertensione e diabete – due dei classici fattori di rischio di tipo epidemiologico per le patologie coronariche. In particolare, l'ipertensione ha esercitato un'influenza molto maggiore in confronto ai migliori indicatori per la previsione di ischemia e di morte in condizioni acute e il diabete (in casi in cui non era fuori controllo e non richiedeva quindi il ricovero per ragioni indipendenti) ha esercitato un'influenza paragonabile a quella di fattori con un genuino potere predittivo.

I classici fattori di rischio epidemiologici qui considerati non sono semplicemente non diagnostici o scarsamente diagnostici. Sono piuttosto *pseudo*-diagnostici (Kern, Doherty, 1982): rappresentano, cioè, informazioni aggiuntive che sono correlate a una diagnosi che si intende valutare ma che non permettono di distinguere tale diagnosi da altre possibili. In contesti sperimentali si è osservato che in condizioni di informazione incompleta le persone, compresi medici e studenti di medicina, tendono a ricercare dati pseudodiagnostici e a utilizzarli per le loro decisioni (Kern, Doherty, 1982; cap. 3 in questo volume). Questo risultato ha suggerito che i medici valutano le ipotesi diagnostiche in modo seriale anziché metterle a confronto le une con le altre. Nel presente studio è stato possibile determinare se i medici incaricati di decidere del ricovero di pazienti con sospetta ischemia miocardica acuta ricercano informazioni utili per il confronto fra ipotesi diagnostiche in competizione. I medici coinvolti, diversamente dai soggetti degli studi sperimentali, non erano vincolati nel numero di elementi aggiuntivi di informazione che potevano ricercare. I nostri risultati documentano quindi l'impiego di informazioni pseudodiagnostiche al di là dei contesti sperimentali e all'interno dei reparti di medicina di emergenza. Quali che siano le altre informazioni che essi hanno ricercato, i medici coinvolti in questo studio hanno raccolto informazioni pseudodiagnostiche e ne sono stati influenzati.

La portata della presente indagine riguarda le decisioni di un numero limitato di medici in due ospedali. Tuttavia si tratta di medici esperti, non di specializzandi o studenti, e di strutture ospedaliere ordinarie. È quindi improbabile che i nostri risultati siano fortemente atipici o minimamente rappresentativi delle pratiche cliniche diffuse.

Il nostro studio riguarda una singola questione clinica: la valutazione di una sospetta ischemia miocardica acuta. Si tratta di un caso che offre ai medici l'opportunità di giudicare in maniera ottimale, perché il problema è relativamente semplice ed è stato studiato in modo estremamente approfondito. È quindi improbabile che l'impatto di informazioni pseudodiagnostiche sulle decisioni dei medici sia confinato a questo specifico problema clinico. È plausibile infatti che tale impatto risulti amplificato in situazioni più complesse in cui il valore diagnostico delle diverse informazioni non è altrettanto noto, i fattori coinvolti sono in numero maggiore e l'insieme delle possibilità diagnostiche da considerare è più esteso.

Idealmente, questo studio avrebbe dovuto affiancare alle informazioni relative a ipertensione e diabete quelle relative a fumo di sigarette, livelli di colesterolo e storia familiare. Occorre però notare che difficilmente l'inclusione di questi dati avrebbe potuto modificare i risultati ottenuti. Se la considerazione di questi ulteriori fattori di rischio non avesse avuto alcun effetto sulle decisioni dei medici, ipertensione e diabete si sarebbero comunque rivelate informazioni pseudodiagnostiche dotate di una loro specifica influenza. Più probabilmente, l'inclusione di questi fattori avrebbe rivelato che essi esercitano a loro volta un qualche effetto, rafforzando i risultati qui riportati.

Il presente studio risente dei possibili limiti dell'impiego delle cartelle cliniche come fonti per la raccolta dei dati. La completezza delle informazioni registrate varia fra i diversi medici e certamente non tutte le informazioni disponibili vengono annotate. Inoltre, sebbene i criteri qui impiegati avessero lo scopo di catturare quante più informazioni fosse possibile, non si può escludere che sinonimi o abbreviazioni idiosincrasiche per qualche condizione o sintomo siano sfuggite nel corso della raccolta dei dati. Naturalmente, l'impiego di cartelle cliniche implica anche che alcuni elementi di informazione non verbale ai quali i medici potrebbero essere sensibili non sono disponibili per l'indagine.

Nella raccolta dei dati abbiamo utilizzato, per ciascun paziente, il tracciato dell'ECG e non la lettura che i medici ne avevano dato. È noto che molte anomalie dell'ECG passano inosservate nella lettura dei medici dei reparti di emergenza, sebbene le anomalie mancate siano spesso prive di significato clinico (Westdrop, Gratton, Watson, 1992). È possibile che la considerazione delle anomalie degli ECG che i medici non avevano notato avrebbe modificato alcune delle loro decisioni di ricovero. Se così fosse, impiegare le interpretazioni dei medici anziché i tracciati dell'esame avrebbe aumentato il peso del-

le anomalie degli ECG nella regressione logistica. L'impatto delle informazioni pseudodiagnostiche sarebbe rimasto inalterato, ma l'influenza dei dati realmente diagnostici sarebbe potuta risultare, in termini relativi, meno modesta.

Si potrebbe ipotizzare che le decisioni di ricovero riflettano il giudizio clinico dei medici e che tale giudizio clinico non possa essere colto dall'analisi statistica di alcuni fattori predittivi; oppure che i medici coinvolti fossero in qualche modo a conoscenza di caratteristiche specifiche del loro contesto territoriale e abbiano agito in base a esse. Si è per esempio osservato che la variabilità geografica nell'uso delle radiografie in casi di sospetta polmonite riflette le caratteristiche delle popolazioni cui si rivolgono le diverse strutture ospedaliere (Tape *et al.*, 1991). Se i medici coinvolti nel nostro studio avessero deciso in modo ottimale in relazione al loro contesto territoriale, i fattori che hanno influenzato le loro decisioni avrebbero dovuto mostrarsi utili anche nel prevedere la diagnosi di infarto al momento del congedo. Ma i risultati della regressione per questa variabile dipendente non sostengono l'ipotesi di una superiore conoscenza dell'ambito locale da parte dei medici. Tali risultati sono invece in accordo con quelli degli studi più ampi già citati nell'Introduzione.

Le implicazioni di queste osservazioni per la pratica clinica sono problematiche. Idealmente, le decisioni dei medici dovrebbero basarsi il più possibile su informazioni in grado di predire gli sviluppi delle condizioni cliniche dei pazienti. I classici fattori di rischio per patologie coronariche non sono completamente irrilevanti. Jayes *et al.* (1992) hanno evidenziato che, per alcuni pazienti (quelli per cui il significato dei fattori predittivi in condizioni acute è equivoco), i fattori di rischio classici possono fornire una qualche utilità predittiva. Tuttavia, tali fattori di rischio non dovrebbero prevalere sulle alterazioni del tratto ST e il loro peso non dovrebbe nemmeno avvicinarsi a quello del dolore al petto come principale disturbo iniziale o di una storia di precedenti infarti miocardici. Inoltre, la mancanza di significative interazioni fra fattori di rischio classici e fattori predittivi in condizioni acute mostra che i primi hanno influenzato le decisioni dei medici anche quando il significato dei secondi non era equivoco.

Non dovrebbe stupire che i medici assegnino tanta importanza ai fattori pseudodiagnostici qui individuati; essi sono ampiamente trattati e discussi, in letteratura e nell'insegnamento, come dati cruciali in relazione alle patologie cardiache. Nella letteratura scientifica anteriore al lavoro di Jayes *et al.* (1992) la distinzione fra i fattori di rischio classici e gli indicatori della condizione acuta non viene enfatizzata, e talvolta non è neppure segnalata. I presenti risultati sono in accordo con l'osservazione ben documentata che i medici non elaborano decisioni di ricovero ottimali in casi di sospetta ischemia miocardica acuta (Selker, Griffith, D'Agostino, 1991b) e ne suggeriscono una possibile spiegazione.

La considerazione dei medici per informazioni pseudodiagnostiche può inoltre essere alimentata da inferenze scorrette tratte dalla pratica clinica. Per esempio, l'ipertensione cronica ha una notevole prevalenza fra i pazienti affetti da ischemia miocardica acuta. È però molto diffusa anche fra i pazienti che non sono colpiti da ischemia miocardiaca acuta, quindi non contribuisce in modo sostanziale a distinguere pazienti con e senza ischemia miocardiaca acuta.

In termini formali, p(IT) rappresenta la probabilità di ipertensione in assenza di informazioni aggiuntive e specifiche riguardo al paziente; si tratta quindi della prevalenza dell'ipertensione nella popolazione dei pazienti dei reparti di emergenza. p(IMA) è la probabilità di ischemia miocardica acuta in assenza di specifiche conoscenze riguardo al paziente (vale a dire, la «probabilità a priori» dell'ischemia miocardica acuta). p(IMA|IT) (cioè la probabilità dell'ischemia miocardica acuta data la presenza di ipertensione) è la prevalenza dell'ischemia miocardica acuta fra i pazienti ipertesi ed è maggiore di p(IMA|non-IT) (la probabilità dell'ischemia miocardica acuta data l'assenza di ipertensione) precisamente se l'ipertensione è un'utile indicatore per la previsione dell'ischemia miocardica acuta. Se però p(IMA|IT) è uguale a p(IMA|non-IT), allora l'ipertensione è irrilevante: non fornisce alcuna indicazione utile riguardo alla probabilità dell'ischemia miocardica acuta. Quanto più p(IMA|IT) e p(IMA|non-IT) differiscono, tanto più la presenza di ipertensione è utile nel valutare la probabilità che un paziente sia colpito da ischemia miocardica acuta.

Nei reparti di emergenza molti pazienti soffrono sia di ipertensione sia di ischemia miocardica acuta. Un medico che abbia osservato una probabilità p(IMA|IT) relativamente alta (cioè numerosi casi di ischemia miocardica acuta fra i pazienti ipertesi) potrebbe assumere che la probabilità p(IMA|non-IT) è bassa, e che quindi esiste un legame fra ipertensione e ischemia miocardiaca acuta che può essere di aiuto nella diagnosi (Yates, 1990). È possibile calcolare queste probabilità utilizzando i dati raccolti in questa indagine. Non disponiamo di un dato relativo all'effettiva prevalenza p(IMA), ma a scopo illustrativo possiamo utilizzare la proporzione di diagnosi di infarto miocardico al momento del congedo, che chiameremo p(IMC). Nel nostro caso, p(IT) = 0,280 e p(IMC) = 0,140. Inoltre, p(IMC|IT) = 0,186 e p(IMC|non-IT) = 0,122. Come si vede, la probabilità di infarto miocardico fra i pazienti non ipertesi è due terzi della probabilità di infarto miocardico fra quelli ipertesi. Questa osservazione riflette la modesta utilità diagnostica dell'ipertensione, ed è in constrasto con l'impatto sproporzionato di questo fattore sulle decisioni di ricovero.

L'influenza di informazioni pseudodiagnostiche sulle decisioni di ricovero suggerisce possibili direzioni di indagine per migliorare le decisioni mediche. In generale, si è osservato che gli effetti della formazione nel modificare i

comportamenti dei medici nella pratica clinica sono modesti (Eisenberg, 1986). Tuttavia, programmi specifici mirati ad insegnare ai medici a discriminare fra informazioni diagnostiche e pseudodiagnostiche potrebbero rivelarsi più efficaci. Si è inoltre mostrato che, in casi simulati, la capacità di giudicare il valore diagnostico delle informazioni può essere migliorata inducendo i medici ad adottare una strategia detta «euristica delle ipotesi in competizione» (Wolf, Gruppen, Billi, 1985, 1988). Potrebbe essere opportuno studiare le potenzialità di questo metodo nella pratica clinica attraverso studi controllati.

5. Conclusioni

Ipertensione e diabete (rilevati dalle annotazioni nelle cartelle cliniche) non sono indicatori efficaci nel prevedere l'infarto miocardico fra i pazienti dei reparti di emergenza. Ciononostante, hanno influenzato le decisioni dei medici, offrendo un esempio di impiego di informazioni pseudodiagnostiche. L'influenza di queste informazioni pseudodiagnostiche potrebbe essere associata all'attenzione che viene loro dedicata nella formazione medica. Le decisioni cliniche potrebbero quindi essere migliorate da programmi di formazione che insegnino a distinguere fra fattori di rischio di tipo epidemiologico e fattori dotati di reale valore diagnostico in condizioni acute.

Gli autori esprimono la loro gratitudine per il sostegno fornito dalla Agency for Health Care Policy and Research (RO1 HS06409) a questo progetto di ricerca.

6. *La distorsione retrospettiva nella valutazione della probabilità delle diagnosi**

di *Hal R. Arkes, Paul D. Saville, Robert L. Wortmann, Allan R. Harkness*

Chi sa che un certo evento si è verificato potrebbe sostenere che, se gli fosse stato chiesto di predire quello stesso evento, molto probabilmente ci sarebbe riuscito. In effetti, le persone in possesso della conoscenza a posteriori di un evento tendono ad assegnare a quell'evento stime di probabilità più alte di chi deve formulare una previsione in mancanza di tale conoscenza a posteriori (Fischhoff, 1975). Questo effetto è noto come «distorsione retrospettiva» (*hindsight bias*).

L'impatto della distorsione retrospettiva è stato illustrato sperimentalmente in studi relativi a eventi storici e casi psichiatrici (Fischhoff, 1975), fatti di cultura generale (Wood, 1978) e risultati di esperimenti scientifici (Slovic, Fischhoff, 1977). Nell'indagine qui presentata abbiamo cercato di estendere lo studio di questo fenomeno alla diagnosi medica. Questo approccio riflette la tendenza a considerare il processo diagnostico come un compito cognitivo, nel quale dovrebbe essere possibile osservare sperimentalmente fenomeni cognitivi stabiliti da ricerche svolte in altri ambiti (Arkes, Harkness, 1980).

1. Metodi

1.1. *Partecipanti*

Settantacinque medici hanno partecipato allo studio. Ognuno di loro era uno specializzando fino a cinque anni dalla laurea in medicina oppure un docente della Facoltà di Medicina dell'Università del Wisconsin. L'insieme dei partecipanti è stato suddiviso in cinque gruppi di quindici.

* Traduzione di *Hindsight bias among physicians weighing the likelihood of diagnoses*, «Journal of applied psychology», 66 (1981), pp. 252-254, copyright © 1981 American Psychological Association.

1.2. *Stimoli sperimentali*

Ai medici nel gruppo di controllo o «gruppo di previsione» (*foresight group*) è stato presentato un caso reale che esemplifica un problema clinico comune. Le informazioni disponibili suggerivano la considerazione di quattro possibili diagnosi, dal momento che nessuno dei sintomi presenti permetteva di eliminare qualcuna delle quattro diagnosi né rappresentava una manifestazione esclusiva di qualcuna di esse. Il materiale presentato a questo gruppo di medici era il seguente.

> Questa è la descrizione di un caso clinico. Vorremmo che la leggessi e decidessi quale probabilità avresti assegnato a ciascuna delle quattro possibili patologie, se ti fossi trovato a fare una diagnosi. Assegna una probabilità a ogni diagnosi, accertandoti che le probabilità sommino al 100%.
>
> Un barista di trentasette anni è stato in buona salute finché recentemente ha sviluppato un dolore di crescente intensità al ginocchio sinistro, che è diventato caldo e gonfio. Qualche giorno dopo si sono manifestati dolore, gonfiore e calore al polso sinistro e al ginocchio destro. L'esame fisico ha in effetti rivelato gonfiore, riscaldamento e versamento a entrambe le ginocchia e al polso sinistro. Non ci sono deformità. Il fegato risulta ingrandito di due centimetri al di sotto del livello del costato. I risultati dell'esame emocromo-citometrico sono normali. L'esito del test della velocità di eritrosedimentazione (ves) è 30 mm. Un Latex test è risultato negativo. Non sono ancora disponibili i risultati dei test chem-7 sul siero sanguigno e degli antigeni Hb_sAg. Il liquido sinoviale risulta contenere 20.000 globuli bianchi per mm cubo, con 80% di neutrofili e bassa viscosità. Si è trovato pus in eccesso nelle urine. Febbre a 38°.
>
> Ora ti chiediamo di assegnare a ognuna delle quattro possibili diagnosi la probabilità che ritieni avresti assegnato. Accertati che le probabilità sommino al 100%.
>
> - sindrome di Reiter (incompleta);
> - artrite batterica da streptococchi (febbre reumatica);
> - gotta;
> - epatite da siero in fase anitterica.

Lo studio comprendeva anche quattro gruppi sperimentali o «gruppi retrospettivi» (*hindsight group*) ai quali veniva presentato lo stesso materiale, con la sola eccezione della frase iniziale, che era differente. Le quattro frasi iniziali erano rispettivamente: «questa è la descrizione di un caso di sindrome di Reiter (incompleta)», «questa è la descrizione di un caso di artrite batterica da streptococchi (febbre reumatica)», «questa è la descrizione di un caso di gotta» e «questa è la descrizione di un caso di epatite da siero in fase anitterica». Si osservi che la seconda frase del questionario richiedeva a tutti i soggetti nei gruppi retrospettivi una loro valutazione individuale e indipendente del caso del paziente descritto.

1.3. *Procedura*

Sessantanove medici hanno ricevuto il questionario in incontri di gruppo. Altri sei hanno svolto il compito individualmente. Non sono stati richiesti né il nome né altri segni identificativi, garantendo quindi ai partecipanti l'anonimato.

2. Risultati

Le probabilità medie assegnate a ciascuna diagnosi in ognuno dei gruppi sono presentate in tab. 1, dove sono anche indicati i confronti fra ogni gruppo retrospettivo e le stime corrispondenti nel gruppo di previsione. In due delle righe della tabella le cifre non sommano esattamente a 100 a causa di piccoli errori aritmetici da parte di due partecipanti, che hanno fornito stime la cui somma era diversa dal 100%. I dati relativi a questi soggetti sono stati comunque inclusi nell'analisi.

Poiché la statistica parametrica riguarda la stima di un singolo parametro e poiché le stime di probabilità medie di una particolare diagnosi in ciascuno dei gruppi retrospettivi è influenzata dalle altre stime di probabilità in quello stesso gruppo, per l'analisi dei dati sono state impiegate tecniche statistiche non parametriche.

Tab. 1 - Tutte le probabilità sono moltiplicate per 100. Le cifre tra parentesi indicano il numero di medici le cui stime per quella particolare diagnosi superano la stima corrispondente nel gruppo di previsione

			Ipotesi diagnostica			
Gruppo	*n*	*Diagnosi nota*	*Sindrome di Reiter*	*Artrite batterica*	*Gotta*	*Epatite*
Previsione	15	Nessuna	43,9	11,1	29,0	15,7*
Retrospettivo	15	Reiter	39,2 (7)	15,2	24,7	20,9
Retrospettivo	15	Artrite	36,7	31,0 (11)	10,0	22,3
Retrospettivo	15	Gotta	38,6	10,3	34,9 (7)	16,1**
Retrospettivo	15	Epatite	34,8	16,0	12,2	37,7* (13)

* Le cifre su questa riga non sommano a 100,00 a causa di un errore aritmetico commesso da un partecipante.
** Le cifre su questa riga non sommano a 100,00 a causa dell'arrotondamento dei decimali.

In tutto, trentotto dei sessanta medici nei gruppi retrospettivi hanno assegnato alla diagnosi nota una stima di probabilità più alta di quella corrispon-

dente nel gruppo di previsione ($p < 0{,}02$; sign test). Per ogni gruppo retrospettivo, abbiamo esaminato in modo indipendente il numero di medici che hanno fornito alla diagnosi nota stime di probabilità più alte di quella corrispondente nel gruppo di previsione. Nel gruppo retrospettivo associato alla diagnosi di artrite batterica da streptococchi, i medici che hanno fornito stime di probabilità più alte di quella corrispondente nel gruppo di previsione sono stati 11 su 15 ($p < 0{,}0625$), mentre nel gruppo retrospettivo associato alla diagnosi di epatite da siero sono stati 13 su 15 ($p < 0{,}005$). In ciascuno dei due gruppi retrospettivi associati alle diagnosi che erano state considerate più probabili nel gruppo di previsione (sindrome di Reiter e gotta), solo 7 medici su 15 hanno assegnato stime di probabilità più alte delle rispettive stime del gruppo di previsione.

3. Discussione

Il nostro risultato principale è che i giudizi dei medici mostrano gli effetti della distorsione retrospettiva. Va però notato che il fenomeno è stato osservato soltanto in relazione alle due diagnosi giudicate meno probabili nel gruppo di previsione. I medici nei relativi gruppi retrospettivi hanno asserito che avrebbero considerato queste due diagnosi (sindrome di Reiter e artrite batterica) da due a tre volte più probabili di quanto è stato stimato nel gruppo di previsione. Questo risultato replica le osservazioni di Fischhoff (1975) e Wood (1978), che hanno osservato una distorsione retrospettiva più forte per eventi che vengono inizialmente giudicati meno plausibili.

Nel nostro studio, come nella maggior parte di quelli precedenti nello stesso ambito, è stato impiegato come strumento di indagine materiale presentato in forma scritta. Resta da determinare la generalizzabilità di questi risultati in situazioni cliniche reali. Elstein, Shulman, Sprafka (1978) hanno mostrato che indagini in cui venivano impiegate simulazioni altamente realistiche (con attori professionisti che descrivevano i «loro» sintomi) davano risultati molto simili a quelli di indagini in cui venivano impiegate simulazioni meno realistiche (descrizioni scritte dei sintomi). Questo risultato, unito al fatto che è stata utilizzata la descrizione di un caso reale, ci rende fiduciosi riguardo alla validità dei risultati del presente studio.

L'entità dell'effetto di distorsione retrospettiva può forse essere meglio apprezzata considerando un'importante differenza fra la presente indagine e il precedente lavoro di Fischhoff (1975). Nello studio di Fischhoff gli stimoli sperimentali riguardavano eventi storici e casi psichiatrici. Per esempio, ai partecipanti nel gruppo di previsione venivano presentate alcune informazioni sulle fazioni in lotta nella guerra fra l'Impero Britannico e la popolazione himalayana dei Gurka. Si chiedeva quindi di assegnare stime di probabilità a

diversi esiti, come una vittoria britannica o dei Gurka. I partecipanti nei gruppi retrospettivi, a cui venivano forniti gli stessi dati con l'aggiunta dell'informazione che la guerra si era poi conclusa in un determinato modo, non avevano nessuna conoscenza precedente in base alla quale valutare la probabilità di un certo esito della guerra o la rilevanza di informazioni storiche alquanto esoteriche riguardo a una possibile vittoria britannica o dei Gurka. In altre parole, i partecipanti si trovavano a esprimere un giudizio a partire da una condizione di ignoranza virtualmente completa, che poteva renderli particolarmente soggetti a errori di tipo cognitivo. Per contro, i medici dispongono di una mole notevole di conoscenze riguardo alla probabilità di diverse patologie e alla relazione fra i sintomi presenti e le possibili diagnosi. In confronto ai soggetti studiati da Fischhoff, i medici sono quindi attrezzati molto meglio per elaborare decisioni basate su informazioni adeguate. E tuttavia, anche in questa condizione relativamente favorevole, i medici risultano essere vittime di errori sistematici nei loro giudizi. Questo suggerisce che la distorsione retrospettiva è un fenomeno molto robusto, anche se limitato alle diagnosi meno probabili.

L'osservazione della distorsione retrospettiva in medici praticanti ha diverse importanti implicazioni. La presenza di questa fonte di errore suggerisce che gli individui coinvolti in questo studio hanno cercato di dare un senso all'insieme delle informazioni di cui disponevano e non di analizzare i dati iniziali in modo indipendente. Questa è precisamente la situazione in cui può trovarsi un medico che sia chiamato a fornire un secondo parere o a prendere in cura un paziente seguìto in precedenza da un altro medico. La distorsione retrospettiva tenderà in questi casi a produrre un secondo parere in linea con le valutazioni già elaborate da altri. In breve, l'accuratezza delle diagnosi può essere compromessa in modo significativo dalla conoscenza delle diagnosi che sono già state proposte.

L'osservazione della distorsione retrospettiva ha anche importanti implicazioni per la formazione dei medici. Gran parte della formazione dei medici ha luogo «sul campo»: gli studenti lavorano a contatto con gruppi di medici direttamente responsabili delle diagnosi e della cura dei pazienti. Avendo sistematicamente accesso alle diagnosi, è probabile che gli studenti sopravvalutino l'accuratezza del proprio giudizio diagnostico e non si rendano pienamente conto della difficoltà iniziale a elaborare quelle stesse diagnosi. Tenuto conto dell'impatto della distorsione retrospettiva sui giudizi clinici, questo tipo di apprendimento «a posteriori» non rappresenta probabilmente una strategia ottimale.

II. Preferenze e decisioni

7. *L'elicitazione delle preferenze per terapie alternative**

di *Barbara J. McNeil, Stephen G. Pauker, Harold C. Sox, Amos Tversky*

Nell'opinione pubblica e nella professione medica vi è una crescente considerazione per la necessità di incorporare le preferenze dei pazienti nelle decisioni cliniche. Perché ciò avvenga, il medico deve fornire al paziente informazioni sui possibili esiti delle opzioni di trattamento disponibili e il paziente deve essere in grado di comprendere e utilizzare tali informazioni. In questo studio indaghiamo in che modo le persone impiegano informazioni di tipo statistico riguardo ai possibili esiti di terapie alternative. Ci siamo concentrati su un particolare problema medico (il cancro ai polmoni operabile) e abbiamo chiesto ai partecipanti di scegliere fra l'intervento chirurgico e la terapia radiologica sulla base di semplici descrizioni delle possibili conseguenze dei due trattamenti. Abbiamo considerato quattro variabili: il tipo di dati presentati ai partecipanti (probabilità cumulative o aspettativa di vita), la caratterizzazione o «incorniciamento» (*framing*) degli esiti clinici (in termini di mortalità o in termini di sopravvivenza), l'identificazione del tipo di trattamento (intervento chirurgico e terapia radiologica *vs*. trattamenti non identificati e indicati come «trattamento *A*» e «trattamento *B*») e infine la popolazione di soggetti coinvolti (medici, pazienti e studenti universitari post-laurea).

1. Metodi

1.2. *Il problema clinico*

Per il nostro studio abbiamo selezionato il cancro al polmone perché presenta la possibilità di una scelta netta fra due terapie alternative (radiologica

* Traduzione di *On the elicitation of preferences for alternative therapies*, «New England journal of medicine», 306 (1982), pp. 1259-1262. Per gentile concessione della Massachusetts Medical Society.

o chirurgica) che incidono in maniera differente sulle probabilità di sopravvivenza. Uno studio precedente su questo stesso problema, basato sull'approccio dell'analisi formale delle decisioni (Raiffa, 1968; Keeney, Raiffa, 1976), ha mostrato che un numero considerevole di pazienti preferiva la terapia radiologica a quella chirurgica nonostante la terapia radiologica fosse associata a una più bassa probabilità di sopravvivenza a lungo termine (McNeil, Weichselbaum, Pauker, 1978), presumibilmente perché essa non implica il rischio di una possibile morte perioperatoria.

Come nel precedente studio sul cancro al polmone, abbiamo fatto uso di dati riportati da Mountain (1976) e Mountain, Carr, Anderson (1974) riguardo agli esiti dell'intervento chirurgico e di dati riportati da Hilton (1960) riguardo agli esiti della terapia radiologica per il cancro al polmone operabile. Questi studi, in accordo con altri, indicano che per un paziente di 60 anni trattato chirurgicamente il tasso di mortalità perioperatoria è in media del 10% e il tasso di sopravvivenza a cinque anni è in media del 34% circa. I tassi di sopravvivenza a uno, due, tre e quattro anni sono, rispettivamente, del 68%, 51%, 40% e 35%. Per la terapia radiologica non vi è essenzialmente alcun rischio di morte legato al trattamento e il tasso di sopravvivenza a cinque anni è in media del 22%; i tassi di sopravvivenza a uno, due, tre e quattro anni sono, rispettivamente, 77%, 44%, 28% e 23%. Per popolazioni di pazienti di età differente, sono state ricavate stime di sopravvivenza modificate attraverso dati raccolti dal National Cancer Institute sulla crescita del rischio di morte per cancro al polmone e sui tassi di mortalità annua per specifiche fasce di età (Axtell, Cutler, Myers, 1972). Il confronto fra i due trattamenti mostra che il trattamento chirurgico offre migliori prospettive a lungo termine a fronte di un più elevato rischio immediato.

1.2. *Tipo di dati forniti*

Abbiamo impiegato due diversi tipi di dati. I dati del primo tipo, detti «probabilità cumulative», indicavano la probabilità di sopravvivenza (o di morte) subito dopo il trattamento, un anno dopo il trattamento e a cinque anni dal trattamento. Il termine di un anno è stato scelto perché rappresenta la prospettiva di breve periodo nella quale la probabilità di sopravvivenza è più alta con il trattamento radiologico che con quello chirurgico; il termine dei cinque anni è stato scelto perché è di uso comune in medicina nella valutazione di diversi trattamenti fra loro alternativi. I dati del secondo tipo, detti «aspettative di vita», indicavano la probabilità di sopravvivenza (o di morte) subito dopo il trattamento e l'aspettativa di vita associata a ciascun trattamento, cioè il periodo medio di sopravvivenza a seguito di ciascun trattamento.

La curva di sopravvivenza che descrive gli esiti dell'intervento chirurgico ha una coda più lunga (cioè si estende in misura maggiore verso destra) di quella associata al trattamento radiologico. Pertanto, la proporzione di pazienti che sopravviverà, per esempio, oltre i dieci anni è più alta con la terapia chirurgica che con quella radiologica. Ci si dovrebbe quindi aspettare che l'impiego delle aspettative di vita (che risentono in modo significativo della lunghezza della coda della curva di sopravvivenza) faccia apparire l'opzione dell'intervento chirurgico più attraente di quanto essa non appaia in base ai tassi di sopravvivenza a uno e a cinque anni (che non sono influenzati dalla lunghezza della coda della curva di sopravvivenza).

1.3. *Identificazione dei trattamenti*

Per la metà circa dei partecipanti i dati presentati venivano identificati come esiti del trattamento chirurgico e di quello radiologico; per i restanti partecipanti, i trattamenti in questione non venivano identificati e le terapie alternative erano semplicemente indicate come «trattamento *A*» o «trattamento *B*». I dati relativi agli esiti del trattamento *A* erano identici a quelli associati all'intervento chirurgico, i dati relativi agli esiti del trattamento *B* erano identici a quelli associati alla radioterapia. Abbiamo incluso nello studio la manipolazione di questa variabile per indagare in che misura le scelte possono essere influenzate da preconcetti (siano essi corretti o meno) relativi all'intervento chirurgico e alla terapia radiologica.

1.4. *Incorniciamento degli esiti*

Le probabilità presentate a metà circa dei soggetti si riferivano alla sopravvivenza dopo un particolare periodo di tempo (per esempio, 68% di probabilità di sopravvivere più di un anno). Le probabilità presentate ai restanti soggetti si riferivano alla mortalità (per esempio, 32% di probabilità di morire prima di un anno). Recenti studi di psicologia cognitiva sull'«incorniciamento» (*framing*) dei problemi decisionali indicano che caratterizzare gli esiti di possibili corsi di azione in termini di probabilità di sopravvivenza piuttosto che in termini di probabilità di morte può avere un effetto notevole sulle preferenze degli individui (Kahneman, Tversky, 1979; Tversky, Kahneman, 1981). Più precisamente, questi studi suggeriscono che l'impatto della mortalità perioperatoria sul confronto fra i due trattamenti dovrebbe essere maggiore quando il dato relativo è presentato come una differenza fra 0% e 10% nel tasso di mortalità che non quando è presentato come una differenza 100% e 90% nel tasso di sopravvivenza. Dal momento che il rischio di morte perio-

peratoria è il principale svantaggio del trattamento chirurgico in confronto alla radioterapia, abbiamo ipotizzato che il trattamento chirurgico sarebbe stato scelto con più frequenza a fronte della descrizione del problema in termini di probabilità di sopravvivenza che non a quella in termini di probabilità di morte.

1.5. *Popolazioni di partecipanti*

Abbiamo studiato tre gruppi di partecipanti: medici, pazienti e studenti. A nessuno dei partecipanti era stato diagnosticato un cancro al polmone. I pazienti erano 238 uomini con problemi medici cronici seguìti come pazienti ambulatoriali da internisti del Palo Alto Veterans Administration Medical Center. La loro età variava fra 40 e 80 anni, con una media di 58 e una distribuzione simile a quella dei pazienti colpiti da cancro al polmone. I medici erano 424 radiologi di età compresa fra 28 e 67 anni, con una media di 43, e stavano frequentando corsi post-laurea di specializzazione o aggiornamento presso la Facoltà di Medicina di Harvard e il Brigham and Women's Hospital. Dal momento che i medici hanno normalmente un ruolo centrale nella scelta delle terapie, lo studio delle preferenze di questo gruppo di partecipanti è di considerevole interesse. Il terzo gruppo di partecipanti comprendeva 491 studenti post-laurea della Stanford Business School che avevano completato diversi corsi di statistica e teoria della decisione. La loro età media era di 29 anni. Sono stati inclusi nello studio per esaminare gli eventuali effetti dell'età e della conoscenza di strumenti formali di ragionamento.

Ci aspettavamo che gli studenti, che erano in media più giovani sia dei pazienti sia dei medici, scegliessero l'intervento chirurgico con più frequenza rispetto agli altri due gruppi. Ci aspettavamo inoltre che i medici e gli studenti post-laurea, distinti dai pazienti in virtù del loro addestramento formale, fossero meno sensibili a variazioni nell'incorniciamento degli esiti.

Tab. 1 - Numero dei partecipanti che ricevevano i dati nei diversi modi

Popolazione	*Esiti presentati in termini di probabilità di morte*		*Esiti presentati in termini di probabilità di sopravvivenza*	
	Trattamenti identificati	*Trattamenti non identificati*	*Trattamenti identificati*	*Trattamenti non identificati*
Pazienti	60	60	59	59
Medici	80	135	87	122
Studenti	196	64	101	130

1.6. *Procedura*

Ogni partecipante è stato assegnato a una delle quattro condizioni sperimentali ottenute dalle possibili combinazioni fra le diverse presentazioni dei trattamenti (identificati o non identificati) e le diverse presentazioni degli esiti (sopravvivenza o morte). Il numero di soggetti compresi in ciascun gruppo è riportato in tab. 1.

Tutti i partecipanti ricevevano sia le probabilità cumulative sia i dati relativi all'aspettativa di vita, in questo ordine. Ai partecipanti che ricevevano i dati con trattamenti identificati e con esiti descritti in termini di probabilità di morte veniva presentato il testo seguente.

> L'intervento chirurgico per il cancro al polmone implica un'operazione ai polmoni. La maggior parte dei pazienti resta in ospedale per due o tre settimane e accusa un qualche dolore nella zona dell'incisione. Una volta dimessi, la convalescenza a casa richiede ancora un mese circa; dopodiché i pazienti generalmente si sentono bene.
> La terapia radiologica per il cancro al polmone implica l'impiego di radiazioni per distruggere il tumore e richiede che il paziente si rechi in ospedale circa quattro volte alla settimana per sei settimane. Ogni seduta di trattamento dura alcuni minuti, durante i quali il paziente è steso su un tavolo come se fosse sottoposto a esami radiografici. Durante il periodo del trattamento alcuni pazienti soffrono di nausea e vomito, ma intorno alla fine della sesta settimana generalmente si sentono bene.
> Quindi, dopo un periodo che si aggira intorno alle sei settimane dall'inizio del trattamento, i pazienti sottoposti a intervento chirurgico e quelli sottoposti a terapia radiologica si sentono all'incirca nello stesso modo.

A questo punto, ai partecipanti venivano fornite le seguenti probabilità cumulative, che erano anche riassunte in una tabella.

> Su 100 persone sottoposte a intervento chirurgico, 10 moriranno durante il trattamento, 32 saranno morte dopo un anno e 66 saranno morte dopo cinque anni. Su 100 persone sottoposte a terapia radiologica, nessuna morirà durante il trattamento, 23 saranno morte dopo un anno e 78 saranno morte dopo cinque anni.
> Quale trattamento preferiresti?

Dopo che i partecipanti avevano scelto, si diceva loro che questi dati sintetizzavano l'esperienza relativa a svariati ospedali e che ora avrebbero dovuto considerare nuove informazioni che riguardavano uno specifico ospedale ed effettuare una nuova scelta sulla base di tali informazioni.

> In questo particolare ospedale, il 10% dei pazienti sottoposti a intervento chirurgico muore durante il periodo perioperatorio. I pazienti che sopravvivono al

trattamento hanno una aspettativa di vita (il numero medio di anni di vita restanti) di 6,8 anni. L'aspettativa di vita dell'insieme di tutti i pazienti sottoposti a intervento chirurgico (compresi quelli che muoiono nel periodo postoperatorio) è di 6,1 anni. Con la terapia radiologica, nessun paziente muore durante il trattamento, e l'aspettativa di vita dei pazienti sottoposti a terapia radiologica è di 4,7 anni.
Quale trattamento preferiresti?

Ai soggetti che ricevevano i dati nel formato in cui i trattamenti alternativi non venivano identificati veniva presentato un diverso insieme di informazioni preliminari.

Sia il trattamento *A* sia il trattamento *B* sono farmaci che vengono somministrati in ospedale a pazienti affetti da cancro. La somministrazione è in entrambi i casi per via endovenosa e nessuno dei due farmaci ha significativi effetti collaterali. I trattamenti *A* e *B* sono considerati uguali tranne che per quanto riguarda i tassi di sopravvivenza.

I dati presentati (probabilità cumulative e aspettative di vita) erano identici a quelli nel formato con trattamenti identificati, tranne che per il fatto che «intervento chirurgico» e «terapia radiologica» erano sostituiti rispettivamente da «trattamento *A*» e «trattamento *B*». Per i partecipanti a cui venivano presentati i dati in termini di sopravvivenza, tutte le probabilità di morte erano sostituite con le corrispondenti probabilità di sopravvivenza.

I pazienti sono stati intervistati individualmente. I medici e gli studenti hanno risposto a questionari scritti.

2. Risultati

La percentuale dei partecipanti che in ciascuna condizione sperimentale ha scelto la radioterapia piuttosto che l'intervento chirurgico è riportata in *Tab. 2*. I risultati relativi alle condizioni con probabilità cumulative e con aspettative di vita sono stati sottoposti a due separate analisi della varianza $3 \times 2 \times 2$ dopo una trasformazione in arcseno (Snedecor, Cochran, 1967). Gli effetti di ognuna delle quattro variabili considerate sono risultati statisticamente significativi ($p < 0{,}001$). La tab. 2 mostra uno schema di risultati estremamente regolare: con una sola eccezione di piccola entità, non si riscontrano «effetti incrociati» fra le principali variabili (tipo di dati, identificazione dei trattamenti e incorniciamento degli esiti). Per esempio, tutte le cifre relative alla condizione «probabilità cumulative» superano le corrispondenti cifre relative alla condizione «aspettative di vita».

Considereremo ora brevemente i principali effetti osservati, uno alla volta.

Tab. 2 - Percentuali dei partecipanti che hanno preferito la radioterapia all'intervento chirurgico

Tipo di dati	*Variabili relative a trattamenti ed esiti*				*Totale*
	Probabilità di morte		*Probabilità di sopravvivenza*		
	Trattamenti identificati	*Trattamenti non identificati*	*Trattamenti identificati*	*Trattamenti non identificati*	
(N. di partecipanti)	(336)	(259)	(247)	(311)	(1153)
Probabilità cumulative					
Pazienti	40	68	22	31	40
Medici	50	62	16	51	47
Studenti	43	53	17	27	35
Totale	44	61	18	37	40
Aspettative di vita					
Pazienti	35	50	19	27	28
Medici	28	39	9	41	31
Studenti	21	41	9	24	22
Totale	25	42	11	31	27

2.1. *Tipo di dati forniti*

Come previsto, i partecipanti che avevano ricevuto dati relativi alle aspettative di vita hanno nel complesso scelto la radioterapia con meno frequenza (27%) di quelli che avevano ricevuto probabilità cumulative (40%). Un esame delle scelte individuali mostra che il 59% dei soggetti ha scelto l'intervento chirurgico a fronte di entrambi i tipi di dati e che il 26% dei soggetti ha scelto la radioterapia a fronte di entrambi i tipi di dati. Quindi, l'85% dei partecipanti ha scelto nello stesso modo nelle due condizioni. Il 14% dei partecipanti ha scelto la radioterapia nella condizione con probabilità cumulative e l'intervento chirurgico nella condizione con aspettative di vita; soltanto l'1% ha scelto in modo opposto.

2.2. *Identificazione dei trattamenti*

Complessivamente, la radioterapia è stata scelta il 42% delle volte quando non veniva identificata e soltanto il 26% delle volte quando veniva identificata. Evidentemente, l'identificazione dei due trattamenti favorisce l'intervento chirurgico nei confronti della radioterapia.

2.3. *Incorniciamento degli esiti*

Come previsto, l'intervento chirurgico era relativamente meno attraente nella presentazione in termini di mortalità (probabilità di morte) che nella presentazione in termini di sopravvivenza (probabilità di sopravvivenza). In media, la radioterapia è stata preferita all'intervento chirurgico il 42% delle volte nella presentazione in termini di mortalità e il 25% delle volte nella presentazione in termini di sopravvivenza.

2.4. *Popolazioni di soggetti*

La radioterapia ha avuto il livello minimo di popolarità fra gli studenti (28% del totale delle risposte), una popolarità leggermente superiore fra i pazienti (34%) e il livello massimo di popolarità fra i medici (39%). In tutti e tre i gruppi, però, a dispetto delle notevoli differenze di età, reddito e stile di vita, le preferenze variavano secondo uno schema simile in funzione di ciascuna delle altre variabili considerate.

3. Discussione

Abbiamo presentato a un gran numero di pazienti ambulatoriali, medici e studenti post-laurea informazioni che descrivevano i possibili esiti di due terapie alternative per il cancro al polmone. I partecipanti hanno mostrato di comprendere questi dati e di farne uso nelle loro risposte. In particolare, interviste condotte con i pazienti in seguito allo svolgimento dell'esperimento hanno indicato che essi avevano compreso i dati che erano stati presentati loro e che erano in grado di richiamare alla mente importanti elementi di informazione. Tuttavia sia le scelte dei soggetti meno esperti (i pazienti) sia quelle dei soggetti più esperti (i medici) sono state significativamente influenzate da una serie di differenze nella natura e nella forma dei dati presentati.

L'osservazione che i dati relativi alle aspettative di vita favoriscono l'intervento chirurgico mentre le probabilità cumulative favoriscono la radioterapia non è sorprendente alla luce del fatto che la distribuzione della sopravvivenza associata all'intervento chirurgico si estende nel tempo in misura maggiore rispetto a quella associata alla radioterapia. Ma questo risultato illustra le difficoltà che si incontrano nel selezionare il tipo di dati più appropriato per sintetizzare le caratteristiche dei diversi trattamenti. È probabile che in molti casi diverse misure statistiche che possono sembrare ragionevolmente adatte (come, per esempio, la media o la mediana di una distribuzione) possano influenzare chi decide indirizzandolo verso una o verso l'altra delle alternative disponibili.

La minore attrattiva esercitata dalla radioterapia quando i trattamenti venivano identificati indica che le persone si affidano di più a credenze precedenti riguardo ai trattamenti che non ai dati statistici che vengono loro presentati. Non sappiamo però se tali credenze si fondino su informazioni valide o riflettano un pregiudizio ampiamente diffuso contro la radioterapia. Nel primo caso, l'insieme dei dati forniti dovrebbe essere esteso in modo da includere le informazioni aggiuntive presumibilmente impiegate dai soggetti nella condizione in cui i trattamenti venivano identificati. Nel secondo caso, si dovrebbero informare i pazienti in modo appropriato prima del processo di elicitazione delle preferenze, nel tentativo di correggere i loro pregiudizi.

Il risultato forse più notevole del nostro studio riguarda l'effetto delle diverse presentazioni, in termini di sopravvivenza o di morte, sulle scelte degli individui. L'intervento chirurgico è apparso molto più attraente quando gli esiti dei trattamenti erano presentati in termini di probabilità di sopravvivenza che non quando erano presentati in termini di probabilità di morte. Attribuiamo questo risultato al fatto che il rischio di morte perioperatoria ha un impatto maggiore se presentato in termini di mortalità piuttosto che in termini di sopravvivenza. A differenza dei risultati precedenti, che possono essere giustificati o quanto meno razionalizzati, il diverso effetto dell'impiego di una diversa terminologia per descrivere gli stessi esiti rappresenta una illusione cognitiva. L'effetto generale osservato nel presente studio è ampio (25% *vs.* 42%) e robusto: si presenta sia con le probabilità cumulative sia con le aspettative di vita, sia con trattamenti identificati sia con trattamenti non identificati e in tutte e tre le popolazioni di soggetti. Con nostra grande sorpresa, l'effetto non è stato in generale di minore entità né fra i medici (che avevano una considerevole esperienza nel valutare dati clinicamente rilevanti) né fra gli studenti post-laurea (che avevano una formazione statistica) in confronto al gruppo dei pazienti (che non avevano né l'una né l'altra).

A partire dai nostri risultati, si potrebbe essere tentati di concludere che non ha senso mettere a punto metodi specifici ed espliciti per l'elicitazione delle preferenze dei pazienti, dal momento che tali preferenze sono tanto suscettibili al modo in cui i dati sono presentati, alle considerazioni che essi implicitamente sollecitano e ad altre potenziali cause di distorsione. Occorre però notare che le preferenze espresse dai medici, che attraverso i loro consigli ai pazienti svolgono certamente un ruolo centrale nelle decisioni, si sono rivelate sensibili alle medesime fonti di distorsione. Inoltre, sembra improbabile che procedure più informali, nelle quali le alternative di trattamento vengono descritte in termini generali e senza dati statistici quantitativi, possano risultare meno suscettibili agli effetti di diversi modi di presentazione.

Le variazioni fra diverse presentazioni dei dati possono essere impiegate per valutare la robustezza delle preferenze riguardo alle alternative disponibili. Se, per esempio, un paziente mostra di preferire l'intervento chirurgico al-

la radioterapia sia con probabilità cumulative sia con dati relativi alle aspettative di vita e sia con probabilità di morte sia con probabilità di sopravvivenza, allora questa preferenza si può ragionevolmente considerare accertata. Se, per contro, una modifica nel modo di presentazione produce una inversione delle preferenze, allora sono probabilmente necessari dati aggiuntivi, ulteriori discussioni o analisi più approfondite. In conclusione, vorremmo suggerire che una maggiore consapevolezza fra i medici e fra i pazienti degli effetti delle diverse presentazioni dei dati può contribuire a ridurre gli errori e a migliorare la qualità delle decisioni mediche.

Siamo in debito con i volontari che hanno partecipato a questo studio per aver generosamente messo a nostra disposizione il loro tempo.

Questa ricerca è stata in parte finanziata dalla Henry J. Kaiser Family Foundation.

Stephen G. Pauker è titolare di un Research Career Development Award (K04-GM-00349).

8. *È sempre preferibile avere più opzioni? L'effetto di attrazione nelle scelte di trattamento**

di *Janet A. Schwartz, Gretchen B. Chapman*

1. Introduzione

Nelle decisioni relative ai trattamenti è comune che i medici affrontino scelte fra diversi farmaci, ognuno dei quali è sicuro ed efficace. In questi casi, la situazione dei medici è simile a quella dei consumatori, che spesso si trovano a decidere fra diverse alternative comparabili. Un aspetto del comportamento dei consumatori che risulta rilevante anche per le decisioni mediche è che il numero delle opzioni disponibili è in continua crescita. Diverse ricerche dedicate alle scelte dei consumatori hanno identificato alcuni errori sistematici che conseguono precisamente dall'aumento del numero delle alternative comprese nell'insieme delle opzioni disponibili (Tversky, Shafir, 1992b). Tenuto conto della presenza di questi errori, sembra opportuno studiarli in ambiti della decisione medica che presentano problemi affini a quelli caratteristici delle scelte dei consumatori, come la selezione fra diversi farmaci o diverse opzioni di trattamento.

Da un punto di vista normativo, le preferenze per un'alternativa all'interno di un insieme di opzioni di scelta non dovrebbero aumentare con l'aggiunta di una nuova opzione; un fenomeno di questo tipo rappresenta infatti una violazione del cosiddetto *principio di regolarità* (Shafir, Tversky, 1995). Tuttavia, numerosi studi sulle scelte dei consumatori documentano casi in cui l'aggiunta di una terza alternativa a un insieme di due opzioni di scelta può produrre violazioni di principi normativi come quello di regolarità (Tversky, Shafir, 1992a e 1992b; Slovic, 1975; Shafir, Simonson, Tversky, 1993). Per esempio, in uno studio si chiedeva a un gruppo di partecipanti se avrebbero preferito acquistare subito un noto lettore CD che era in saldo per un giorno

* Traduzione di *Are more options always better? The attraction effect in physicians' decisions about medications*, «Medical decision making», 19 (1999), pp. 315-323. Per gentile concessione dell'editore.

soltanto oppure rinviare la decisione per raccogliere ulteriori informazioni sul prodotto. A un secondo gruppo di partecipanti venivano presentate le stesse due opzioni più una terza: un altro interessante modello di lettore CD (con una qualità di riproduzione leggermente superiore e un prezzo leggermente più alto), anch'esso in saldo per un giorno soltanto. Si è osservato che era più probabile che i partecipanti scegliessero di rinviare la decisione (e conservare lo status quo) di fronte alla scelta fra tre opzioni che di fronte a quella fra due (Tversky, Shafir, 1992b).

Gli autori dello studio hanno suggerito che l'aggiunta di un nuovo modello di lettore CD, i cui vantaggi e svantaggi erano simili a quelli del modello già presente, aveva aumentato il livello di conflitto percepito dai partecipanti nel prendere la decisione. In altri termini, quando si trattava di valutare un solo modello di lettore CD, il suo valore appariva apprezzabile e l'alternativa di rinviare la decisione sembrava poco giustificata; ma quando si trattava di considerare due modelli di lettore CD fra loro simili non emergeva alcuna chiara giustificazione per scegliere uno piuttosto che l'altro. Dal momento che uno dei due lettori CD era più conveniente ma l'altro era di un produttore associato a una migliore qualità, i partecipanti si trovavano di fronte alla difficoltà di dover soppesare la fama dell'azienda produttrice contro il prezzo. L'opzione di rinviare la decisione appariva così più attraente, perché permetteva ai partecipanti di evitare questo difficile confronto.

In uno studio collegato a quello appena descritto si è osservata una tendenza a conservare lo status quo nelle decisioni dei medici (Redelmeier, Shafir, 1995). I medici coinvolti nello studio leggevano la descrizione di un caso clinico e dovevano indicare quale trattamento avrebbero consigliato al paziente unitamente alla raccomandazione di una visita specialistica. Un gruppo di medici doveva scegliere fra la prescrizione di ibuprofene e nessun trattamento. Per un secondo gruppo di medici le opzioni di scelta comprendevano un ulteriore farmaco (piroxicam). La proporzione dei medici che hanno scelto di non raccomandare alcun trattamento è stata significativamente maggiore quando il problema comprendeva due diverse opzioni di trattamento farmacologico che non quando il problema ne comprendeva una soltanto. Questi risultati sono coerenti con quelli ottenuti nel precedente studio relativo alle scelte dei consumatori e indicano che gli errori cognitivi prodotti dall'aumento del numero delle alternative di scelta possono influenzare la cura dei pazienti.

Dal momento che il comportamento dei medici può risentire del numero di alternative di trattamento disponibili, è interessante scoprire se nelle decisioni dei medici si manifestino anche altri tipi di errori. In particolare, l'obiettivo del presente studio è quello di determinare se un altro errore comunemente osservato negli studi sul comportamento dei consumatori, noto come «effetto di attrazione» (Huber, Payne, Puto, 1982), si presenta nelle decisioni dei medici. In modo simile a quanto avviene nel fenomeno studiato da Redelmeier,

Shafir (1995), nell'effetto di attrazione l'aggiunta di una terza opzione a un insieme di alternative influenza le preferenze relative alle due opzioni originarie. In questo caso, però, l'opzione aggiuntiva non rinforza le preferenze per l'inazione o per la conservazione dello status quo (per esempio, la scelta di non prescrivere alcun trattamento) e non rappresenta un'alternativa in grado di competere con le altre. Supponiamo, per esempio, che siano disponibili due trattamenti per una certa condizione: il trattamento *X* è molto efficace, ma stando agli studi ha numerosi effetti collaterali; il trattamento *Y* è moderatamente efficace, ma produce pochi effetti collaterali. Ora supponiamo che un terzo trattamento *Z* (che funge da «esca») sia moderatamente efficace e allo stesso tempo produca un numero significativo di effetti collaterali. Chiaramente, nessuno preferirebbe *Z* a *Y*, perché *Z* causa più effetti collaterali di *Y* ma non è più efficace di quest'ultimo. L'aggiunta di *Z* all'insieme delle opzioni di scelta, però, fa aumentare la probabilità che venga scelto il trattamento *Y*.

L'effetto di attrazione è un altro esempio di violazione del principio di regolarità, perché prevede che l'aggiunta di una opzione di valore inferiore a un insieme di alternative di scelta aumenti la probabilità che venga scelta una delle opzioni già disponibili. Questo fenomeno è anche noto come «effetto di dominanza asimmetrica» perché l'esca (*decoy*) (per esempio, il trattamento *Z*) è «dominata» dall'opzione «bersaglio» (*target*) (vale a dire, è chiaramente inferiore a una delle opzioni originarie, che con l'aggiunta dell'esca risulterà più attraente; per esempio, il trattamento *Y*), mentre non è dominata dal «competitore» (*competitor*), cioè dall'altra opzione originaria (per esempio, il trattamento *X*). Una delle possibili spiegazioni di questo effetto è simile a quella suggerita da Tversky, Shafir (1992b). Il fatto che un'alternativa presenti vantaggi evidenti rispetto a un'altra può fornire a molte persone una giustificazione convincente per compiere una scelta (Simonson, 1989). In altri termini, considerando soltanto il bersaglio e il competitore, la decisione risulta difficile perché le due opzioni sono entrambe attraenti e la scelta fra le due richiede di soppesare le loro specifiche caratteristiche. Quando però entra in scena l'opzione-esca, chi è chiamato a decidere è in grado di identificare un aspetto (il confronto, appunto, con l'esca) riguardo al quale una delle due opzioni originarie appare chiaramente superiore all'altra. La scelta a favore dell'opzione-bersaglio sembra così offrire un'agevole via d'uscita dal problema decisionale.

Sebbene svariati studi abbiano mostrato l'effetto di attrazione nelle decisioni dei consumatori, esso non è mai stato indagato nelle decisioni dei medici. L'effetto di attrazione potrebbe risultare particolarmente rilevante per le scelte di trattamento, precisamente per via del modo in cui i diversi trattamenti sono presentati in molti testi di consultazione. Nel *Physician's Desk Reference* (1997), per esempio, l'efficacia dei trattamenti e gli effetti associati

al loro impiego vengono spesso riportati in termini percentuali, in un formato simile a quello utilizzato nel presente studio. In un confronto fra svariati trattamenti efficaci descritti in modo simile, è possibile che i medici siano inclini a evitare il difficile compito di soppesare fra loro i diversi attributi rilevanti, e a scegliere quindi un certo trattamento per il semplice fatto che ne domina un altro. Questo punto risulta particolarmente saliente se si considera che spesso vi sono numerosi trattamenti disponibili per una stessa condizione clinica e che, come abbiamo accennato, il numero delle alternative è in continua crescita. L'effetto di attrazione può quindi riflettere una euristica impiegata per semplificare le decisioni a fronte di ampi insiemi di alternative di scelta.

Per determinare se l'effetto di attrazione si manifesta nelle decisioni dei medici abbiamo presentato a specializzandi di medicina interna tre casi clinici. Gli scenari impiegati descrivevano i sintomi e la storia clinica di tre pazienti con tre diagnosi piuttosto comuni: depressione, sinusite e vaginite. Ai partecipanti si chiedeva di valutare diversi trattamenti e di scegliere quello che giudicavano preferibile sulla base delle seguenti caratteristiche: efficacia e possibili effetti collaterali negli scenari relativi alla depressione e alla sinusite; durata del trattamento e possibili effetti collaterali nello scenario relativo alla vaginite[1].

2. Partecipanti e procedura

I partecipanti in questo studio erano specializzandi di medicina interna fino a tre anni dalla laurea. Abbiamo collocato i questionari nelle loro caselle postali presso l'ospedale. Dei 72 soggetti inizialmente contattati, 40 (28 uomini, 11 donne, e 1 che non ha indicato il genere; età media: 28,2 anni) hanno completato e restituito il questionario (per un tasso di risposte del 56%). Il sistema utilizzato era il seguente. Due settimane dopo la distribuzione iniziale è stata collocata una cartolina nelle caselle postali di coloro che non avevano ancora restituito il questionario per richiamare la loro attenzione. Due settimane dopo che erano state distribuite le cartoline, una nuova versione completa del questionario è stata collocata nelle caselle postali di coloro che non avevano ancora risposto. Abbiamo ripetuto fino a quattro volte questa procedura. La partecipazione allo studio era volontaria e la scelta di partecipare o di non partecipare da parte degli specializzandi contattati non è stata resa nota al personale responsabile della loro formazione medica. Prima dello svolgimento dell'analisi

1. L'efficacia e gli effetti collaterali sono attributi comuni rispetto ai quali si distinguono farmaci diversi e per questo sono stati impiegati nei primi due scenari. Nello scenario relativo alla vaginite uno degli attributi utilizzati è stato la durata del trattamento perché rappresentava una differenza saliente fra le opzioni disponibili in questo ambito.

dei dati, le risposte dei partecipanti e le loro informazioni identificative sono state separate. Tutti i partecipanti sono stati retribuiti per il loro tempo.

Ogni questionario comprendeva la descrizione di tre casi clinici relativi a depressione, sinusite e vaginite. Gli scenari presentati erano stati sottoposti a un professore di medicina interna e a un tutor degli specializzandi per assicurarsi che fossero realistici da un punto di vista clinico. L'ordine dei tre casi presentati e la posizione di ciascuna opzione di trattamento (bersaglio, competitore o esca) sono stati combinati con uno schema sperimentale a quadrato latino bilanciato. Ogni scenario clinico poteva comparire in una delle tre forme seguenti: una condizione di controllo con due sole opzioni *A* e *B* (senza esca), una condizione sperimentale in cui il bersaglio era l'opzione *A*, e una condizione sperimentale in cui il bersaglio era l'opzione *B*. In altri termini, ciascuno dei farmaci che comparivano nella condizione di controllo con due sole opzioni compariva in entrambe le condizioni sperimentali che comprendevano un'opzione-esca, essendo una volta il bersaglio e una volta il competitore. Questo schema ci ha permesso di controllare alcuni potenziali fattori interferenti, come la familiarità con particolari farmaci o la predilezione per certi specifici marchi commerciali. Infatti, se la familiarità con il prodotto fosse alla base della scelta per un certo farmaco, ci si dovrebbe aspettare che esso sia preferito agli altri in tutte e tre le condizioni, compresa quella in cui è presente un'opzione-esca e in cui il farmaco in questione è il competitore. Ogni questionario conteneva i tre scenari, ciascuno in una condizione diversa (controllo, opzione *A* come bersaglio, opzione *B* come bersaglio). Fra i diversi partecipanti, ogni scenario compariva nelle tre diverse condizioni previste.

Ai partecipanti si chiedeva di leggere ciascun caso clinico e di considerare la relativa lista di possibili trattamenti. La descrizione dei trattamenti è stata ricavata dal *Physician's Desk Reference* (1997), con minimi aggiustamenti necessari per rispettare le caratteristiche dello schema sperimentale. In alcuni casi, per conservare l'accuratezza della descrizione dei farmaci, abbiamo manipolato le relazioni di dominanza fra le alternative servendoci della descrizione dei sintomi e della storia clinica del paziente. Gli stessi medici che hanno valutato la plausibilità dei casi clinici presentati hanno anche valutato la plausibilità e l'accuratezza delle descrizioni dei farmaci.

Ai partecipanti si diceva di immaginare che i trattamenti indicati fossero gli unici disponibili. Un insieme di opzioni di scelta di due o tre alternative pone limiti più restrittivi di quelli comuni nelle reali scelte di trattamento, ma non sembra comunque irragionevole tenuto conto del modo in cui tali scelte risultano spesso vincolate dalle diverse polizze assicurative, dalla storia del paziente o dall'interazione fra farmaci. Dopo aver preso in considerazione la lista dei trattamenti presentati, ciascun partecipante valutava l'appropriatezza di ciascun trattamento su una scala da 1 a 5, nella quale «1» indicava il valore minimo e «5» il valore massimo di appropriatezza; ciascun partecipante indicava quindi quale avrebbe scelto di prescrivere fra i trattamenti proposti.

Infine, si chiedeva ai partecipanti di immaginare che il medico responsabile avesse messo in discussione la loro scelta e di fornire argomenti in difesa della scelta effettuata.

2.1. *Primo scenario: depressione*

Nella descrizione del caso di depressione gli specializzandi leggevano il testo seguente.

> Il paziente è una donna di 56 anni. È rimasta vedova sei settimane fa, dopo aver assistito per diversi mesi alle sofferenze del marito a causa della malattia da cui era affetto. Durante una visita di controllo la paziente riferisce di aver avuto nelle ultime tre settimane difficoltà a dormire, scarso appetito, ansia e mancanza di interesse per attività precedentemente praticate con piacere. Riferisce inoltre che gli incontri settimanali con il suo terapista, che le erano stati di grande aiuto durante la malattia del marito, negli ultimi tempi non l'hanno più aiutata. A parte questo, la paziente non presenta altre particolari condizioni cliniche e non ha precedenti di tipo psichiatrico. La tua diagnosi indica l'emergere di una condizione di depressione. La paziente è una candidata appropriata per un trattamento farmacologico antidepressivo. Ti chiediamo di leggere la lista dei trattamenti indicati di seguito e di assumere che siano le uniche opzioni disponibili. L'assicurazione sosterrà le spese di qualunque prescrizione a condizione che la paziente copra la quota usuale di 3 dollari.

La tab. 1 mostra le opzioni di trattamento presentate in questo scenario e la loro descrizione. Nella condizione di controllo con due sole opzioni, gli unici trattamenti proposti erano la paroxetina HCl e il nefazodone HCl. Nella condizione *A* con tre opzioni, la presenza dell'opzione aggiuntiva rappresentata dalla fluoxetina (l'esca) faceva della paroxetina (superiore alla fluoxetina per una minore incidenza di effetti collaterali) il bersaglio della manipolazione sperimentale. Nella condizione *B* con tre opzioni, la presenza dell'opzione aggiuntiva rappresentata dalla imipramina (l'esca) faceva del nefazodone (superiore all'imipramina per una minore incidenza di effetti collaterali) il bersaglio della manipolazione sperimentale. A ogni partecipante veniva presentato uno soltanto dei tre insiemi di opzioni di scelta compresi in tab. 1. In tutte e tre le condizioni relative al caso di depressione, la paroxetina HCl e il nefazodone HCl erano descritti nello stesso modo e la descrizione del paziente era identica.

Tab. 1 - I farmaci dello scenario relativo alla depressione

Farmaco	*Efficacia*	*Possibili effetti collaterali*
Paroxetina HCl (es. Paxil)	Negli studi clinici controllati si è osservato che la paroxetina HCl era superiore a un placebo rispetto a **2 su 4** validi indicatori della condizione di depressione	Il **5%** dei pazienti riporta bocca secca, nausea e sonnolenza
Nefazodone HCl (es. Serzone)	Negli studi clinici controllati si è osservato che il nefazodone HCl era superiore a un placebo rispetto a **3 su 4** validi indicatori della condizione di depressione	Il **25%** dei pazienti riporta bocca secca, nausea e sonnolenza
Fluoxetina HCl (es. Prozac) [esca per la paroxetina HCl]	Negli studi clinici controllati si è osservato che la fluoxetina HCl era superiore a un placebo rispetto a **2 su 4** validi indicatori della condizione di depressione	Il **15-20%** dei pazienti riporta bocca secca, nausea e sonnolenza; il 10% dei pazienti riporta disturbi del sonno
Imipramina HCl (es. Tofranil) [esca per il nafazodone HCl]	Negli studi clinici controllati si è osservato che la imipramina HCl era superiore a un placebo rispetto a **3 su 4** validi indicatori della condizione di depressione	Il **35%** dei pazienti riporta bocca secca, nausea e sonnolenza; riporta inoltre anomalie nell'elettrocardiogramma, allucinazioni e fissazioni

La tab. 2 mostra le risposte dei partecipanti rispetto alla scelta del trattamento e alle valutazioni di appropriatezza in ognuna delle tre condizioni.

Tab. 2 - Scelte e valutazioni di appropriatezza nello scenario relativo alla depressione

		Numero e % delle scelte		*Valutazioni di appropriatezza*	
Condizione	*Farmaco**	*n*	*%*	*Media*	*Dev. std*
Tre opzioni A (bersaglio superiore per effetti collaterali)	Paroxetina HC1 (*b*)	12	86	4,57	0,65
	Nefazodone (*c*)	2	14	3,71	0,73
	Fluoxetina (*e*)	0	0	2,64	1,08
Due opzioni	Paroxetina HC1	12	75	4,31	0,79
	Nefazodone	4	25	3,38	1,02
Tre opzioni B (bersaglio superiore per effetti collaterali)	Paroxetina HC1 (*c*)	2	20	3,40	0,97
	Nefazodone (*b*)	7	70	4,10	1,10
	Imipramina (*e*)	1	10	2,80	1,13

* *b* = bersaglio; *c* = competitore; *e* = esca.

Nella condizione *A* con tre opzioni, in cui la paroxetina HCl era il bersaglio, la frequenza con cui i partecipanti l'hanno scelta non è stata significativamente superiore rispetto alla condizione di controllo con due sole opzioni: 86% *vs.* 75%, χ^2 (1, n = 30) = 0,536, *n.s.*, potenza = 0,11. In altri termini, non si è manifestato un significativo effetto di attrazione in questa condizione, il che potrebbe essere dovuto a un «effetto soffitto» (un numero sproporzionato di soggetti ha scelto la paroxetina HCl nella condizione di controllo con due sole opzioni) e alla modesta potenza dovuta alle limitate dimensioni del campione. L'effetto di attrazione è stato invece rilevato nella condizione sperimentale *B* con tre opzioni; una proporzione significativamente maggiore di partecipanti ha scelto il nefazodone HCl nella condizione in cui esso era il bersaglio che non nella condizione di controllo con due sole opzioni: 70% *vs.* 25%, χ^2 (1, n = 23) = 6,512, $p \leq 0{,}05$, potenza = 0,72.

Abbiamo condotto un'analisi fattoriale della varianza (ANOVA) 2 × 3 (farmaco per condizione) sulle valutazioni di appropriatezza della paroxetina e del nefazodone e abbiamo trovato un'interazione significativa fra farmaco e condizione, che indica l'effetto di attrazione: $F(1, 37) = 5{,}5$, p 0,05, potenza = 0,51. Ciò significa che sia le valutazioni di appropriatezza della paroxetina HCl sia quelle del nefazodone HCl sono in media cresciute nelle condizioni nelle quali l'una e l'altro erano il bersaglio rispetto alle corrispondenti valutazioni nella condizione di controllo. Solo le valutazioni di appropriatezza della paroxetina HCl sono diminuite nella condizione in cui essa compariva come competitore; quelle del nefazodone sono invece leggermente cresciute. Comunque, in media, ognuno di questi farmaci è stato valutato come il più appropriato fra i due quando era il bersaglio e il meno appropriato fra i due quando era il competitore. Questo risultato indica quindi la presenza dell'effetto di attrazione.

2.2. *Secondo scenario: sinusite*

Nella descrizione del caso di sinusite gli specializzandi leggevano il testo seguente.

> Il paziente è un uomo di 25 anni che lamenta i seguenti sintomi: dolore e una sensazione di pressione intorno e dietro agli occhi, sensibilità al contatto sopra e sotto gli occhi, forte mal di testa, congestione, gocciolamento dal naso, tosse produttiva e una temperatura orale di 37,8 gradi. Dopo aver visitato il paziente, formuli la diagnosi di sinusite. Il paziente è un ex-tossicodipendente in fase di recupero [il che esclude l'opzione di trattamento con un narcotico antitussivo]. Per il resto il paziente è in salute e attivo. Per contrastare l'infezione, prescrivi Biaxin, 500 mg ogni 12 ore per 10 giorni. Ti chiediamo di leggere la lista dei trattamenti indicati di seguito e di assumere che siano gli unici disponibili per la cura dei sintomi secondari della sinusite. L'assicurazione sosterrà le spese di qualunque prescrizione a condizione che il paziente copra la quota usuale di 3 dollari.

La frase relativa all'eliminazione di un narcotico antitussivo non era inclusa nello scenario in cui il Nucofed era l'opzione-esca. L'Appendice A riporta la descrizione dei trattamenti nello scenario relativo alla sinusite. La tab. 3 mostra le risposte dei partecipanti rispetto alla scelta del trattamento e alle valutazioni di appropriatezza in ognuna delle tre condizioni. L'analisi delle scelte non rivela un effetto di attrazione significativo nelle due condizioni sperimentali con tre opzioni. Nella condizione *A* con tre opzioni, l'incremento nelle scelte per il bersaglio (Cardec) non è stato statisticamente significativo rispetto alla condizione di controllo con due sole opzioni: 57% *vs.* 30%, χ^2 (1, n = 24) = 1,73, *n.s.*, potenza = 0,26. Nella condizione *B* con tre opzioni, l'incremento nelle scelte per il bersaglio (Tylenol) non è stato statisticamente significativo rispetto alla condizione di controllo con due sole opzioni: 87% *vs.* 70%, χ^2 (1, n = 26) = 1,21, *n.s.*, potenza = 0,20. Tuttavia in entrambi i casi i risultati mostrano una differenza fra le proporzioni di scelta nella direzione prevista e la non significatività dei risultati dell'analisi statistica potrebbe riflettere la dimensione ridotta dei campioni. Per entrambi i trattamenti, rispetto alla condizione di controllo con due sole opzioni, la percentuale dei partecipanti che hanno scelto un determinato trattamento è infatti cresciuta quando quel trattamento era il bersaglio ed è diminuita quando era il competitore.

Abbiamo svolto sulle valutazioni di appropriatezza dei partecipanti un'analisi fattoriale ANOVA 2 × 3 (farmaco per condizione) simile a quella impiegata per lo scenario relativo alla depressione. Si è riscontrata una significativa interazione fra farmaco e condizione, che indica l'effetto di attrazione: $F(2, 36) = 5{,}92$, $p \leq 0{,}01$, potenza = 0,35. Per ciascuno dei due farmaci compresi nella condizione di controllo le valutazioni medie sono cresciute significativamente quando il farmaco in questione era il bersaglio e sono diminuite significativamente quando quel farmaco era il competitore (si veda tab. 3).

Tab. 3 - Scelte e valutazioni di appropriatezza nello scenario relativo alla sinusite

		Numero e % delle scelte		*Valutazioni di appropriatezza*	
Condizione	*Farmaco**	*n*	*%*	*Media*	*Dev. std*
Tre opzioni A	Cardec (*b*)	8	57	4,21	0,90
(bersaglio superiore	Tylenol (*c*)	6	43	3,71	1,13
per efficacia)	Rhinosyn (*e*)	0	0	2,79	0,80
Due opzioni	Cardec	3	30	3,40	0,79
	Tylenol	7	70	4,20	1,17
Tre opzioni B	Cardec (*c*)	2	13	3,37	0,72
(bersaglio superiore	Tylenol (*b*)	14	87	4,43	0,63
per effetti collaterali)	Nucofed (*e*)	0	0	2,19	1,33

* *b* = bersaglio; *c* = competitore; *e* = esca.

2.3. *Terzo scenario: vaginite*

Nella descrizione del caso di vaginite gli specializzandi leggevano il testo seguente.

> Il paziente è una donna di 27 anni che per la prima volta lamenta i seguenti sintomi: prurito, bruciore e perdite vaginali. Dopo aver visitato la paziente formuli la diagnosi di candidiasi vaginale. Oltre a soffrire per la prima volta dell'infezione, la paziente sta allattando un bambino di due mesi [il che esclude l'opzione di trattamento con il farmaco di uso orale Diflucan (fluconazolo)]. Ti chiediamo di leggere la lista dei trattamenti indicati di seguito e di assumere che siano gli unici disponibili. L'assicurazione sosterrà le spese di qualunque prescrizione a condizione che la paziente copra la quota usuale di 3 dollari.

La frase relativa al fluconazolo non compariva nello scenario in cui tale farmaco era l'opzione-esca. L'Appendice B riporta le descrizioni dei farmaci impiegati nello scenario relativo alla vaginite. La tab. 4 mostra le risposte dei partecipanti rispetto alla scelta del trattamento e alle valutazioni di appropriatezza in ognuna delle tre condizioni.

Tab. 4 - Scelte e valutazioni di appropriatezza nello scenario relativo alla vaginite

		Numero e % delle scelte		*Valutazioni di appropriatezza*	
Condizione	*Farmaco**	*n*	*%*	*Media*	*Dev. std*
Tre opzioni A (bersaglio superiore per la durata del trattamento)	Butoconazolo (*b*)	9	90	4,60	0,70
	Tioconazolo (*c*)	1	10	2,50	1,18
	Clotrimazolo (*e*)	0	0	3,50	1,18
Due opzioni	Butoconazolo	8	57	3,86	0,95
	Tioconazolo	6	43	3,29	1,33
Tre opzioni B (bersaglio superiore per effetti collaterali)	Butoconazolo (*c*)	14	88	4,81	0,40
	Tioconazolo (*b*)	2	12	3,38	0,89
	Fluconazolo (*e*)	0	0	1,06	0,25

* *b* = bersaglio; *c* = competitore; *e* = esca.

Nella condizione sperimentale *A* con tre opzioni la proporzione di partecipanti che ha scelto il bersaglio (butoconazolo) non si è distinta in modo statisticamente significativo da quella che ha scelto lo stesso trattamento nella condizione di controllo con due sole opzioni: 90% *vs*. 57%, χ^2 (1, n = 24) =

3,05, *n.s.*, potenza = 0,42. I risultati mostrano comunque una differenza nelle proporzioni di scelta nella direzione prevista. Per contro, nella condizione sperimentale *B* con tre opzioni nella quale il tioconazolo era il bersaglio, i risultati hanno evidenziato una differenza nella direzione opposta a quella prevista (12% *vs*. 43%). Ciò significa che le scelte per il competitore sono aumentate e quelle per il bersaglio sono diminuite in confronto con la condizione con due sole opzioni, sebbene questa differenza non sia risultata statisticamente significativa: χ^2 (1, n = 30) = 3,52, *n.s.*, potenza = 0,47.

Abbiamo svolto sulle valutazioni di appropriatezza dei partecipanti un'analisi fattoriale ANOVA 2 × 3 (farmaco per condizione) simile a quella impiegata negli altri due scenari. In questo scenario non è stata rilevata alcuna interazione significativa: F(2, 37) = 2,66, *n.s.*, potenza = 0,26. In altri termini, i risultati nello scenario relativo alla vaginite non indicano la presenza dell'effetto di attrazione. Questo risultato molto probabilmente riflette il fatto che in tutte e tre le condizioni il butoconazolo è stato preferito a tutte le altre opzioni di trattamento disponibili. Una successiva analisi fattoriale ANOVA 2 × 2 è stata quindi condotta senza includere la condizione in cui il fluconazolo compariva come opzione-esca. Sebbene nella direzione prevista, gli esiti di questa analisi non hanno dato risultati statisticamente significativi.

2.4. *Analisi entro i soggetti*

Oltre a indagare l'effetto di attrazione in ogni specifico scenario, lo abbiamo analizzato nel comportamento dei singoli partecipanti nei diversi scenari (*within subjects*). A questo scopo, abbiamo considerato se ciascun partecipante, attraverso le sue scelte e le sue valutazioni di appropriatezza nei diversi problemi, ha giudicato nel modo migliore il farmaco bersaglio e nel modo peggiore il competitore. La proporzione di partecipanti che ha scelto l'opzione-bersaglio in entrambe le condizioni che includevano un'opzione-esca (10 su 39, 26%) è stata in effetti significativamente più alta della proporzione di partecipanti che ha scelto un competitore in entrambe le condizioni che includevano un'opzione-esca (2 su 39, 5%): test McNemar, χ^2 (1, n = 39) = 7,143, p < 0,01, potenza = 0,76. Il restante 69% dei partecipanti ha scelto un'opzione-bersaglio in un caso e un competitore nell'altro, con l'eccezione di un partecipante che ha scelto un'opzione-esca ed è stato escluso da questa analisi.

Per determinare se emergeva un effetto entro i soggetti, abbiamo anche analizzato le valutazioni di appropriatezza. In ogni scenario e per ogni partecipante abbiamo sottratto la valutazione del trattamento *B* dalla valutazione del trattamento *A* (il trattamento *A* era il bersaglio nella condizione sperimentale *A* e il trattamento *B* era il bersaglio nella condizione sperimentale *B*). Questo significa che, per quanto riguarda la condizione sperimentale *A*, ab-

biamo sottratto la valutazione assegnata al competitore da quella assegnata al bersaglio, mentre per quanto riguarda la condizione sperimentale *B* abbiamo sottratto la valutazione assegnata al bersaglio da quella assegnata al competitore. Ci aspettavamo una forte differenza positiva fra le valutazioni assegnate ai due trattamenti nella condizione sperimentale *A*, una differenza prossima allo zero nella condizione di controllo con due sole opzioni e una differenza negativa per quanto riguarda la condizione sperimentale *B*. Le medie delle differenze (calcolate su tutti e tre gli scenari) sono state rispettivamente di 1,250 per quanto riguarda la condizione sperimentale *A*, di 0,775 per quanto riguarda la condizione di controllo e di 0,225 per quanto riguarda la condizione sperimentale *B*. Un'analisi fattoriale ANOVA 3 × 3 (condizione per differenze fra le valutazioni di appropriatezza) ha mostrato un effetto significativo riconducibile alla condizione: $F(2, 37) = 6{,}97$, $p \leq 0{,}02$, potenza = 0,42. Ciò significa che le differenze fra le valutazioni di appropriatezza erano ordinate secondo le previsioni (sebbene la differenza media per quanto riguarda la condizione sperimentale *B* non sia stata negativa, il che indica una tendenza generale a preferire quei trattamenti che figuravano come bersaglio nella condizione sperimentale *A*)[2]. Nell'insieme dei tre scenari che ciascuno di loro si è trovato di fronte, i partecipanti hanno valutato meglio le opzioni-bersaglio e peggio i competitori. Di conseguenza, anche se le scelte e le valutazioni non differivano in modo significativo in ciascun singolo scenario, l'analisi dei dati entro i soggetti indica complessivamente la presenza dell'effetto di attrazione.

3. Discussione

Data la diffusione di errori sistematici della decisione negli studi sul comportamento dei consumatori, e in particolare di violazioni del principio di regolarità, sembra opportuno esaminare la presenza di errori simili in ambiti

2. Sospettavamo che la ragione per cui, nell'analisi entro i soggetti, la differenza media fra le valutazioni di appropriatezza nella condizione sperimentale *B* fosse positiva anziché negativa e per cui la differenza nella condizione di controllo con due sole opzioni non fosse prossima a zero risiedesse nel fatto che nella condizione sperimentale *B* relativa al caso di vaginite il competitore era stato giudicato in media più appropriato del bersaglio. Ulteriori analisi hanno mostrato che, una volta escluso lo scenario relativo alla vaginite, la differenza media nella condizione sperimentale *B* era in effetti negativa (–0,583). Tuttavia, tali analisi hanno anche rivelato che la presenza dello scenario della vaginite non spiegava una differenza media diversa da zero nella condizione di controllo con due sole opzioni. Senza lo scenario della vaginite, le differenze medie nella condizione di controllo con due sole opzioni e nella condizione sperimentale *A* erano infatti, rispettivamente, 0,884 e 0,967, il che suggerisce, ancora una volta, che i soggetti hanno mostrato una leggera preferenza generale per quei farmaci che figuravano come bersaglio nella condizione sperimentale *A*.

della decisione medica che presentano problemi affini a quelli propri delle scelte dei consumatori. In questo studio, abbiamo osservato che una violazione della regolarità nota come effetto di attrazione (Huber, Payne, Puto, 1982) si presenta nelle decisioni dei medici di fronte alla scelta fra diverse opzioni di trattamento farmacologico. Ciò significa che l'aggiunta di un'alternativa di valore inferiore (un'esca) a un insieme dato di opzioni di trattamento aumenta la probabilità che un particolare trattamento (il bersaglio) venga scelto.

Sia nello scenario relativo alla depressione sia in quello relativo alla sinusite i risultati sono stati piuttosto chiari: quando un certo farmaco era il bersaglio, i partecipanti l'hanno scelto con più frequenza e l'hanno considerato più appropriato rispetto a quanto accadeva nella condizione di controllo con due sole opzioni. I risultati dello scenario relativo alla vaginite non sono stati altrettanto chiari. Da una parte, quando il butoconazolo era il bersaglio, la proporzione di partecipanti che lo sceglievano cresceva rispetto a quando lo stesso trattamento compariva nella condizione di controllo con due sole opzioni. D'altra parte, la percentuale di individui che sceglievano il butoconazolo cresceva anche quando esso era il competitore. È possibile suggerire diverse spiegazioni per questa osservazione. Innanzi tutto, gli attributi impiegati in questo scenario erano effetti collaterali e durata del trattamento, mentre negli altri due scenari venivano impiegati efficacia clinica ed effetti collaterali. È possibile che i potenziali effetti collaterali siano stati considerati molto più importanti della durata del trattamento e che questo abbia indotto i partecipanti a rispondere in modo differente di fronte al caso di vaginite. Tuttavia, se le cose stessero così, un numero sproporzionato di partecipanti dovrebbe aver selezionato il butoconazolo in tutte e tre le condizioni, compresa la condizione di controllo con due sole opzioni.

Una spiegazione alternativa dei risultati nello scenario della vaginite riguarda il riferimento a una paziente in periodo di allattamento. I medici si trovavano qui a decidere quale fosse il trattamento più appropriato non soltanto per la paziente, ma anche per il suo bambino. Una delle ragioni per cui l'opzione che richiedeva tre giorni di trattamento è sembrata così attraente quando era il competitore potrebbe risiedere nel fatto che i partecipanti hanno ritenuto rischioso per il bambino un trattamento concentrato in una sola dose giornaliera. Sebbene nessuno dei trattamenti intravaginali descritti in questo studio sia controindicato per madri in fase di allattamento, il fluconazolo (che si assume oralmente) lo è. Inoltre, la condizione in cui il fluconazolo compariva come opzione-esca era l'unica in cui la scheda descrittiva dei farmaci presentava un'esplicita avvertenza contro la prescrizione del farmaco durante l'allattamento (si veda Appendice B). Questa avvertenza potrebbe avere indotto molti partecipanti a considerare potenzialmente dannoso anche l'altro trattamento da assumere in una singola dose (il tioconazolo). Questo potrebbe anche spiegare perché la distribuzione delle scelte fra i due trattamenti in-

travaginali era relativamente equilibrata nella condizione con due sole opzioni; infatti, in questo caso, i medici non si sono trovati di fronte a un'esplicita controindicazione relativa a un trattamento da assumere in una singola dose. È da notare comunque che il fatto che il competitore, e non il bersaglio, sia stato scelto più frequentemente quando è stata aggiunta un'alternativa rappresenta pur sempre una violazione della regolarità – solo, non quella che avevamo previsto.

A parte la condizione con il fluconazolo come esca nello scenario della vaginite, il presente studio fornisce sostegno alla presenza dell'effetto di attrazione nelle decisioni dei medici riguardo alla prescrizione di farmaci in casi ipotetici. Lo studio ha, d'altra parte, alcuni limiti relativi al campione utilizzato. Innanzi tutto la generalizzabilità dei risultati potrebbe essere limitata dal fatto che i partecipanti coinvolti erano medici fino a tre anni dalla laurea. Future ricerche dovrebbero indagare la presenza di questo tipo di errore nelle decisioni di medici più esperti. In secondo luogo, il campione impiegato nello studio è di dimensioni ridotte e si è quindi avuta solo una modesta capacità di individuare differenze significative nelle percentuali di scelta in ciascun singolo scenario. L'analisi dei dati entro i soggetti, tuttavia, ha fornito risultati più robusti e ha indicato chiaramente la presenza dell'effetto di attrazione. In generale, era infatti più probabile che i medici scegliessero un trattamento bersaglio piuttosto che un competitore e che giudicassero il primo, e non il secondo, come il più appropriato fra i due.

Nel presente studio sono stati utilizzati casi clinici ipotetici descritti in forma scritta e non si è quindi mostrata un'influenza dell'effetto di attrazione sulle decisioni mediche in contesti reali o un impatto dell'effetto stesso sugli esiti clinici dei pazienti. Inoltre, gli scenari impiegati in questo studio non presentavano la totalità dei trattamenti disponibili per le rispettive patologie. Ciononostante, i risultati suggeriscono che l'effetto di attrazione può avere un impatto sulle decisioni mediche reali perché le descrizioni dei trattamenti usate in questo studio rispecchiano piuttosto fedelmente il modo in cui i diversi farmaci sono elencati e descritti nei testi di consultazione e nei formulari. Lo schema utilizzato in questi testi agevola l'individuazione di rapporti di dominanza e rende così possibile il verificarsi di errori decisionali. Un'azienda farmaceutica potrebbe, per esempio, cercare sapientemente di far comparire un suo prodotto vicino a un trattamento che esso domina, per incrementarne la quota di mercato. E insiemi di opzioni di scelta di questo tipo possono presentarsi anche senza che vi sia un'intenzione deliberata. Un flusso costante di nuovi trattamenti farmaceutici inonda annualmente il mercato, e molti di questi farmaci rendono obsoleti prodotti precedenti. I farmaci più datati che rimangono sul mercato possono quindi agire come esche nelle scelte mediche. Sebbene sia improbabile che i medici prescrivano un vecchio farmaco, la sua presenza in un insieme di opzioni di scelta può incrementare la disponibilità a prescrivere un prodotto molto simile ma più avanzato.

È importante notare che una relazione di dominanza dipende non solo dalle caratteristiche di due prodotti farmaceutici, ma anche dalle caratteristiche del paziente. La presenza di una particolare condizione clinica concomitante, una specifica storia clinica o un determinato sintomo possono far sì che un certo trattamento sia dominato da un trattamento alternativo in relazione a un paziente ma non a un altro. Alcune delle opzioni-esca impiegate in questo studio costituirebbero opzioni di scelta opportune per pazienti con sintomi o storie cliniche differenti da quelli descritti. Il Nucofed, per esempio, risultava dominato per il fatto che il paziente era stato descritto come un ex-tossicodipendente in fase di recupero ma non perché esso sia, in generale, un farmaco di minor valore. Non è difficile immaginare una situazione differente nella quale questo stesso trattamento avrebbe potuto essere l'opzione privilegiata. Pertanto non sempre è semplice identificare un'opzione-esca. I medici dovrebbero senz'altro essere consapevoli delle tecniche di promozione e di marketing che possono far apparire un prodotto più attraente di un altro, ma dovrebbero anche riconoscere le caratteristiche individuali dei pazienti che possono far sì che un'opzione di trattamento ne domini un'altra.

Una questione che si potrebbe sollevare in relazione ai risultati del presente studio riguarda l'influenza dei marchi commerciali dei prodotti farmaceutici. Abbiamo riportato sia i nomi commerciali sia quelli generici, per assicurarci che tutti i farmaci presentati risultassero riconoscibili. Abbiamo cioè voluto evitare che i medici scegliessero un particolare trattamento per il fatto che riconoscevano un certo nome generico ma non un altro. I possibili legami con determinati marchi commerciali o la familiarità con certi prodotti non dovrebbero aver influenzato la probabilità che emergesse l'effetto di attrazione. Se un particolare trattamento fosse risultato preferibile per la sua fama o riconoscibilità commerciale, questo fattore avrebbe dovuto avere la stessa influenza nella condizione di controllo con due sole opzioni, nella condizione sperimentale in cui il prodotto in questione compariva come competitore, così come nella condizione sperimentale in cui esso compariva come bersaglio. In altri termini, se i soggetti avessero scelto il Tylenol quando era l'opzione-bersaglio per il semplice fatto che era il trattamento a loro più familiare, allora ci si sarebbe dovuti aspettare che il Tylenol ricevesse la maggioranza delle preferenze in *tutte* le condizioni, compresa quella in cui era il competitore; ma questo, nel nostro studio, non si è verificato. L'inclusione dei nomi commerciali non ha quindi interferito con la validità dei nostri risultati.

Come i risultati riportati da Redelmeier, Shafir (1995), quelli del presente studio sottolineano il fatto che aumentare il numero delle alternative disponibili in un insieme di opzioni di scelta può modificare gli esiti di una decisione. L'effetto di attrazione e la conservazione dello status quo sono solo due esempi di errori sistematici nelle decisioni dei consumatori che possono risultare rilevanti per le decisioni mediche. Poiché molti aspetti delle cure sanita-

rie presentano sempre più spesso a medici e pazienti problemi affini a quelli caratteristici delle scelte dei consumatori, è utile identificare tali errori e sviluppare strategie che possano aiutare a evitarli.

In che modo è possibile ridurre o correggere l'effetto di attrazione? Una strategia correttiva efficace richiede che la relazione di dominanza venga notata, in modo che l'alternativa dominata non venga scelta, ma anche che la relazione di dominanza non influenzi la scelta fra le opzioni rimanenti. Le alternative dominate dovrebbero pertanto essere identificate e quindi eliminate dall'insieme delle opzioni di scelta in modo da non influenzare le decisioni relative ai trattamenti rimanenti.

La correzione di questo errore non è semplice. L'effetto di attrazione può infatti manifestarsi, almeno in parte, perché chi decide vuole evitare le difficoltà coinvolte nel soppesare fra loro diversi aspetti o perché è orientato a una scelta che possa essere giustificata in modo efficace. Ricerche precedenti hanno mostrato che le persone sono più facilmente soggette all'effetto di attrazione se viene richiesto loro di difendere le proprie decisioni (Simonson, 1989). In questo studio, abbiamo chiesto ai partecipanti di difendere le loro scelte, ma non abbiamo controllato la specifica influenza della responsabilità e della giustificabilità delle scelte (*accountability*). Tali fattori sono comunque rilevanti per le decisioni mediche a causa della crescente enfasi sul controllo della qualità. Molte organizzazioni sanitarie conducono indagini sulla qualità o pubblicano resoconti sulle prestazioni nelle cure cliniche. Sebbene tali pratiche comportino notevoli benefici per la qualità delle cure, richiedere esplicitamente ai medici e ad altri membri del personale ospedaliero di difendere le proprie decisioni può favorire le scelte più semplici da giustificare, alimentando così errori sistematici come l'effetto di attrazione. Future ricerche potranno esaminare le conseguenze della responsabilità e giustificabilità delle scelte su errori come l'effetto di attrazione nel contesto delle decisioni mediche. La consapevolezza della tendenza a evitare decisioni difficili può comunque aiutare i medici a compiere scelte più appropriate.

Ringraziamo Elaine Leventhal per l'aiuto nell'elaborare i casi clinici; Jack Cappitelli e Laura Niedermayer per l'aiuto nello svolgimento dello studio; Alan Schwartz, Patrice Tremoulet e un *referee* anonimo per utili commenti su versioni precedenti di questo articolo.

Appendice A - Descrizione dei farmaci nello scenario relativo alla sinusite

Farmaco	*Efficacia*	*Possibili effetti collaterali*
Pseudoefedrina HCl (60 mg), destrometorfano HBr (15 mg), carbinoxamina maleato (4 mg) (es. Cardec)	Si è osservato che questo farmaco (5 ml *qid*) ha **un'efficacia del 70%** nel contrastare i sintomi secondari della sinusite	Meno del **10%** dei pazienti riporta effetti collaterali a causa della pseudoefedrina HCl; nessuno ne riporta a causa del destrometorfano HBr
Destrometorfano HBr (30 mg), acetaminofeno (650 mg), pseudoefedrina HCl (60 mg) (es. Tylenol tosse decongestionante)	Si è osservato che questo farmaco (5 ml *qid*) ha **un'efficacia del 90%** nel contrastare i sintomi secondari della sinusite	Meno del **10%** dei pazienti riporta effetti collaterali a causa della pseudoefedrina HCl; il **15%** riporta effetti collaterali a causa del destrometorfano HBr; il **5%** riporta disturbi allo stomaco a causa dell'acetaminofeno
Pseudoefedrina HCl (30 mg), clorfeniramina maleato (2 mg), destrometorfano (15 mg) (es. Rhinosyn) [esca per il Cardec]	Si è osservato che questo farmaco (5 ml *qid*) ha **un'efficacia del 60%** nel contrastare i sintomi secondari della sinusite	Meno del **10%** dei pazienti riporta effetti collaterali a causa della pseudoefedrina HCl; il **5%** dei pazienti riporta sonnolenza a causa della clorfeniramina
Pseudoefedrina HCl (60 mg), fosfato di codeina (20 mg) (es. Nucofed) [esca per il Tylenol]	Si è osservato che questo farmaco (5 ml *qid*) ha **un'efficacia del 90%** nel contrastare i sintomi secondari della sinusite	Meno del **10%** dei pazienti riporta effetti collaterali a causa della pseudoefedrina HCl; il **25%** riporta costipazione, disturbi allo stomaco e sonnolenza a causa del fosfato di codeina

Appendice B - Descrizione dei farmaci nello scenario relativo alla vaginite

Farmaco	*Durata del trattamento*	*Possibili effetti collaterali*
Nitrato di butoconazolo 2% (es. Femstat 3)	Crema intravaginale applicata **per tre giorni consecutivi**	Nessun effetto collaterale noto
Tioconazolo 6,5% (es. Vagistat 1)	Crema intravaginale applicata **in una singola dose**	Aggravamento dei sintomi, es.: prurito, gonfiore
Clotrimazolo 1% (es. Miycelex 7) [esca per il butoconazolo]	Crema intravaginale applicata **per sette giorni consecutivi**	Nessun effetto collaterale noto
Fluconazolo (es. Diflucan) [esca per il tioconazolo]	Medicazione orale assunta **in una singola dose** (150 mg)	Il fluconazolo è trasmesso in alte concentrazioni con il latte materno; è quindi controindicato durante l'allattamento

*9. L'ingannevole ricerca di più informazioni**

di *Donald A. Redelmeier, Eldar Shafir, Prince S. Aujla*

1. Introduzione

Tradizionalmente le Facoltà di Medicina formano i medici a seguire un approccio esaustivo nella raccolta dei dati clinici. Una raccomandazione molto comune è, per esempio, «ricostruisci la storia clinica completa e conduci un esame fisico approfondito» (Bates, 1995). Questa massima suggerisce che avere più informazioni – specialmente se sono accessibili con un modesto dispendio di risorse – è sempre un bene. La massima è ulteriormente rinforzata da standard di qualità professionale che raccomandano che si conosca quanto più è possibile, specie se si tratta di informazioni che si possono acquisire in maniera relativamente semplice (Einsenberg, 1986). La strategia di raccogliere numerose informazioni è spesso appropriata a causa della complessità dei disturbi clinici, della possibilità che un paziente soffra di diverse patologie concomitanti e del rischio concreto che la mancata identificazione di un singolo disturbo possa produrre danni irreparabili (Redelmeier, Tan, Booth, 1998).

Per contro, due comuni argomenti giustificano un approccio più cauto alla raccolta delle informazioni. Il primo riguarda l'esigenza economica di limitare le procedure poco efficienti (Ubel, Goold, 1997). Per esempio, le linee-guida scoraggiano l'esecuzione di un'angiografia cerebrale in pazienti adulti con emicrania ma per il resto in salute, perché gli elevati costi degli esami prevalgono ampiamente sulla scarsa probabilità di rilevare una patologia curabile (Pryse-Phillips *et al.*, 1997). Il secondo argomento in favore di un approccio più cauto ha a che fare con il rispetto dell'autonomia del paziente (Asch, Hershey, 1995). Per esempio, i pazienti potrebbero rifiutare l'angiografia cerebrale per il fatto che la procedura è invasiva (Kassirer, 1983). Nessuno di questi due argomenti risulta però convincente per informazioni ricavate attraverso la storia clinica, l'esame fisico o test di semplice esecuzione.

* Traduzione di *The beguiling pursuit of more information*, «Medical decision making», 21 (2001), pp. 376-381. Per gentile concessione dell'editore.

Altri argomenti suggeriscono di astenersi anche dalla ricerca di dati che si possono ottenere facilmente. Talvolta i dati possono rappresentare risultati falsamente positivi o falsamente negativi e la decisione di raccoglierli deve dipendere dalla considerazione delle probabilità di base (o a priori) (Sox, Blatt, Higgins, 1988). Oppure la raccolta dei dati può essere associata a costi significativi per i pazienti; per esempio, un'ecografia dell'addome è un esame inappropriato per un paziente con una emorragia gastrointestinale acuta, perché il tempo passato in radiologia non può essere utilizzato per il monitoraggio dei segni vitali, per l'esecuzione di trasfusioni sanguigne o per altri trattamenti (Goitein, 1990). Infine, i dati possono talvolta indebolire la relazione medico-paziente spingendo i medici (e in certi casi gli stessi pazienti) a concentrarsi sui numeri trascurando altri aspetti più profondi della situazione (Ong *et al*., 1995; Cook, 1997).

Vorremmo sostenere che gli argomenti appena visti sottovalutano un diverso tipo di effetti negativi che emergono dalle debolezze del ragionamento umano. In particolare, la ricerca psicologica ha scoperto incoerenze sistematiche dovute alle imperfette capacità degli individui di distinguere le informazioni rilevanti da quelle ridondanti (Simon, 1959; Tversky, Kahneman, 1974; Nisbett, Ross, 1980; cap. 5 in questo volume). Una delle debolezze del ragionamento umano è la tendenza a cercare informazioni non strumentali – cioè informazioni che potrebbero in generale essere rilevanti, ma non dovrebbero influenzare la decisione corrente (Tversky, Shafir, 1992a; Baron, Beattie, Hershey, 1988). In linea con questa tendenza, le persone talvolta ricercano più informazioni del necessario e, una volta che le hanno ottenute, tendono a ritenere che esse siano cruciali per la decisione (Bastardi, Shafir, 1998). In questo studio indagheremo se anche nelle decisioni cliniche emerge la tendenza a ricercare e impiegare in modo potenzialmente scorretto informazioni mediche non strumentali.

2. Metodi

Abbiamo adattato metodi sviluppati dagli psicologi per valutare gli errori sistematici cui sono soggetti gli individui nelle loro decisioni (Bastardi, Shafir, 1998). In un esperimento, a 539 studenti universitari veniva presentato uno scenario che riguardava domande di iscrizione alle loro università. Per assegnazione casuale, metà di loro riceveva la seguente semplice versione dello scenario, che offriva immediatamente tutte le informazioni.

> Immagina di far parte del comitato di ammissione. Stai valutando la pratica di un candidato che gioca a calcio nella squadra studentesca, ha lettere di raccomandazione positive ed è redattore di un giornale scolastico. Il candidato ha un

punteggio SAT combinato di 1.250 e ha una B come media dei voti della scuola superiore. Decidi di accettare o di respingere il candidato?

Gli altri partecipanti dovevano valutare lo stesso candidato, ma affrontavano una versione differente del problema che comprendeva la possibile ricerca di un elemento di informazione mancante.

> Immagina di far parte del comitato di ammissione. Stai valutando la pratica di un candidato che gioca a calcio nella squadra studentesca, ha lettere di raccomandazione positive ed è redattore di un giornale scolastico. Il candidato ha un punteggio SAT combinato di 1.250 ma ci sono due segnalazioni discordanti riguardo alla media dei voti della scuola superiore. Il rapporto del consulente d'istituto (*guidance couselor*) indica una B, mentre quello dell'ufficio scolastico indica una A. La scuola dello studente ti ha comunicato che è in corso una verifica del punteggio e che verrai informato entro pochi giorni su quale sia il voto corretto. Decidi di accettare il candidato, di respingerlo o di aspettare chiarimenti da parte della sua scuola prima di decidere?
>
> Se decidi di aspettare chiarimenti, rispondi a quanto segue:
>
> La scuola ti informa che la media dei voti del candidato è una B. Decidi di accettare o di respingere il candidato?

I risultati nella versione semplice hanno mostrato che la maggior parte dei partecipanti accettava il candidato sapendo che la media era una B. Naturalmente, la maggior parte lo avrebbe accettato anche se la media fosse stata una A. Quindi, l'incertezza fra una A e una B era non strumentale ai fini di questa decisione per la maggior parte dei partecipanti del primo gruppo. Per contro, nella seconda versione i partecipanti affrontavano una situazione di incertezza riguardo al fatto che la media del candidato fosse una A o una B, e la maggior parte di loro ha scelto di ricercare questa informazione. Questi soggetti ricevevano quindi la stessa informazione compresa nella versione semplice del problema; venivano cioè informati che la media era una B. Ma la ricerca di questa informazione ha alterato le scelte successive. Dopo aver scoperto che la media era una B e non una A, la maggior parte dei soggetti ha deciso di respingere il candidato. Quindi sono stati complessivamente di meno i partecipanti che hanno accettato il candidato nella versione con ricerca di informazioni che nella versione semplice (46% *vs*. 57%; $p = 0{,}020$).

Nel presente studio indaghiamo se anche gli operatori sanitari tendono a prendere decisioni differenti quando ricercano alcune informazioni rispetto a quando ricevono tutte le informazioni allo stesso momento. A questo scopo, abbiamo chiesto ai partecipanti di considerare uno scenario ipotetico in forma scritta che descriveva una situazione medica. Lo scenario era formulato in una di due versioni. La versione *semplice* (*simple version*) dello scenario rendeva immediatamente disponibili tutte le informazioni. La versione *con ricerca* (*search version*) presentava la stessa situazione medica, ma con un elemento

di informazione mancante che, per divenire accessibile, richiedeva una procedura di ricerca (come la semplice attesa o l'esecuzione di un esame). I partecipanti che in questa versione decidevano di svolgere la ricerca ricevevano quindi la stessa informazione presente nella versione semplice, in modo che le due versioni risultassero a questo punto comparabili (Grice, 1975).

La struttura fondamentale di ciascuna delle indagini qui presentate è quella di un controllo randomizzato nel quale vengono confrontate le decisioni di intervento nelle due versioni dello stesso scenario. Per assegnazione casuale, metà dei partecipanti riceveva la versione semplice e l'altra metà la versione con ricerca. I partecipanti sono rimasti all'oscuro riguardo all'intervento sperimentale (e del tutto ignari dell'esistenza di una versione alternativa). La nostra ipotesi principale è che l'aver ricercato una certa informazione spinga a concentrarsi sull'informazione che è stata così ottenuta e porti a decisioni differenti da quelle che si prenderebbero se quella stessa informazione fosse stata disponibile fin dall'inizio (lo studio è stato approvato dal Comitato di etica della ricerca del Sunnybrook Hospital e del Women's College Health Sciences Centre, Toronto).

3. Risultati

3.1. *Infermieri*

La prima indagine ha coinvolto 211 infermieri affiliati a centri di dialisi di Toronto. A tutti loro è stato inviato un questionario di una pagina ed è stato offerto un premio a estrazione di 50 dollari per la partecipazione. In totale, sono stati restituiti 171 questionari, con un tasso di risposte dell'81%. Per assegnazione casuale, metà degli infermieri ricevevano la seguente versione semplice dello scenario.

> Supponi che una tua parente di 68 anni abbia bisogno di un trapianto a causa di una insufficienza renale. Supponi di essere un donatore compatibile. Saresti disposto a donare?

Gli altri infermieri ricevevano la versione con ricerca dello scenario, che conteneva gli stessi dati ma comprendeva la possibilità di una ricerca di informazioni.

> Supponi che una tua parente di 68 anni abbia bisogno di un trapianto a causa di una insufficienza renale. Supponi che non si sappia se tu sei un donatore compatibile. Puoi sottoporti a un test per scoprirlo. Sceglieresti di fare il test?
>
> Se ti sei detto disponibile a sottoporti al test, rispondi a quanto segue:
>
> Supponi che il test sia stato eseguito e che riveli che sei un donatore compatibile. Saresti disposto a donare?

Nella versione semplice, una minoranza degli infermieri ha scelto di donare un rene (44%). Ciononostante, la maggioranza degli infermieri nella versione con ricerca si è detta disponibile a sottoporsi al test per la compatibilità (69%). La disponibilità a fare il test è presumibilmente dovuta al fatto che la decisione di una donazione non è banale per nessuno, alla tentazione di posporla e alla possibilità di evitarla del tutto nel caso di un risultato negativo. Tuttavia, una volta sottoposti al test e identificati come donatori compatibili (come nella versione semplice), la maggior parte degli infermieri che avevano scelto di ricercare questa informazione si sono detti disponibili a donare (93%). Sono stati pertanto di più, nel complesso, gli individui disposti a donare nella versione con ricerca che nella versione semplice (65% *vs.* 44%; $p = 0{,}007$). La discrepanza si è presentata sia fra coloro che hanno risposto per primi al questionario sia fra coloro che hanno risposto più tardi (cioè sia prima sia dopo il periodo mediano di risposta).

3.2. *Urologi*

La seconda indagine coinvolgeva urologi praticanti affiliati alla American Urologic Association. Abbiamo contattato 1.076 di loro offrendo un premio a estrazione di 5.000 dollari per la partecipazione. In totale, 461 hanno completato e rispedito il questionario, con un tasso di risposte del 43%. Per assegnazione casuale, la metà riceveva la seguente versione semplice dello scenario.

> N.F. è un uomo di 69 anni con un cancro alla prostata allo stadio T1C. Il suo livello di antigene prostatico specifico è marginalmente elevato (7.0 ng/ml) e la biopsia rivela un cancro Gleason 6. L'esame dei sistemi non mostra nulla di notevole a parte respiro corto sotto sforzo (2 rampe di scale). Per il resto è in buona salute. Un consulto medico documenta enfisema cronico con ostruzione moderata (FEV1 di 1.000 ml, equivalente a 40% di quello previsto). In questa situazione, raccomanderesti un intervento chirurgico o la terapia radiologica?

Gli altri urologi ricevevano la versione con ricerca dello scenario, che conteneva gli stessi dati ma comprendeva la possibilità di una ricerca di informazioni.

> N.F. è un uomo di 69 anni con un cancro alla prostata allo stadio T1C. Il suo livello di antigene prostatico specifico è marginalmente elevato (7.0 ng/ml) e la biopsia rivela un cancro Gleason 6. L'esame dei sistemi non mostra nulla di notevole a parte respiro corto sotto sforzo (2 rampe di scale). Per il resto è in buona salute. In questa situazione, raccomanderesti un intervento chirurgico, la terapia radiologica o richiederesti un consulto medico prima di decidere?

> Se hai scelto di richiedere un consulto medico, considera quanto segue:
>
> Il consulto medico documenta enfisema cronico con ostruzione moderata (FEV1 di 1.000 ml, equivalente a 40% di quello previsto). In questa situazione, raccomanderesti un intervento chirurgico o la terapia radiologica?

Nella versione semplice, molti degli urologi (42%) hanno scelto di operare. Tuttavia, nella versione con ricerca, una percentuale maggiore (58%) ha scelto di richiedere un consulto medico, probabilmente perché i dati avrebbero potuto essere d'aiuto nella cura del paziente. Tuttavia, dopo aver ricevuto il consulto e scoperto la portata dell'enfisema (identica a quella nella versione semplice), la maggioranza di quelli che avevano richiesto il consulto si sono espressi contro l'intervento chirurgico (92%). Di conseguenza, meno urologi si sono detti favorevoli a operare nella versione con ricerca che nella versione semplice (26% *vs.* 42%; $p = 0{,}001$). La discrepanza si è presentata sia fra coloro che hanno risposto per primi al questionario sia fra coloro che hanno risposto più tardi (sia prima sia dopo il periodo mediano di risposta). Si noti che, diversamente che nel caso presentato agli infermieri, la ricerca di informazioni qui considerata (il consulto medico) non è obbligatoria in vista della decisione se operare oppure no.

3.3. *Accademici*

La nostra terza indagine riguardava medici accademici i cui indirizzi sono stati reperiti su Internet. Un campione di 1.596 medici degli Stati Uniti e del Canada sono stati contattati via email senza ricevere incentivi per la partecipazione. In totale, 574 questionari sono stati completati e restituiti, con un tasso di risposte del 36%. Per assegnazione casuale, la metà dei partecipanti riceveva la seguente versione semplice dello scenario.

> Stai viaggiando su un volo aereo e rispondi alla chiamata «c'è un un medico a bordo?». Un passeggero di 60 anni ha avuto durante il decollo da 15 a 20 minuti di dolore «opprimente» al petto (ora esauritosi). La storia clinica non presenta elementi notevoli. Il paziente appare malato. Il battito cardiaco è a 80 circa. Il kit di primo soccorso contiene un misuratore della pressione. Misuri una pressione sistolica di 120 (l'ambiente è troppo rumoroso per eseguire un'auscultazione). In questa situazione raccomanderesti che il pilota faccia atterrare l'aereo per ragioni mediche o che prosegua il volo come programmato?

Gli altri medici ricevevano la versione con ricerca dello scenario, che conteneva gli stessi dati ma comprendeva la possibilità di una ricerca di informazioni.

> Stai viaggiando su un volo aereo e rispondi alla chiamata «c'è un medico a bordo?». Un passeggero di 60 anni ha avuto durante il decollo da 15 a 20 minuti di dolore «opprimente» al petto (ora esauritosi). La storia clinica non presenta elementi notevoli. Il paziente appare malato. Il battito cardiaco è a 80 circa. Il kit di primo soccorso non contiene un misuratore della pressione, ma

l'assistente di volo sa che c'è un secondo kit in un altro punto dell'aeroplano. Raccomanderesti che il pilota faccia atterrare l'aereo per ragioni mediche, che prosegua il volo come programmato oppure chiederesti di avere il misuratore della pressione prima di prendere una decisione?

Se hai scelto di chiedere il misuratore della pressione, rispondi a quanto segue:

Misuri una pressione sistolica di 120 (l'ambiente è troppo rumoroso per eseguire un'auscultazione). In questa situazione raccomanderesti che il pilota faccia atterrare l'aereo per ragioni mediche o che prosegua il volo come programmato?

Pochi medici nella versione semplice avrebbero raccomandato di proseguire il volo (11%). Per contro, un numero significativo nella versione con ricerca ha richiesto il misuratore della pressione (22%), probabilmente perché il dato avrebbe potuto essere un'utile informazione di riferimento nel caso le condizioni del paziente cambiassero. Una volta ottenuta la misura relativamente rassicurante della pressione sanguigna (la stessa della versione semplice), la maggior parte dei medici che avevano richiesto il misuratore della pressione hanno raccomandato la prosecuzione del volo (85%). Sono quindi stati di più i medici che hanno raccomandato di proseguire il volo nella versione con ricerca che in quella semplice (21% *vs.* 11%; $p = 0{,}001$). La discrepanza è stata osservata sia fra coloro che hanno risposto per primi al questionario sia fra coloro che hanno risposto più tardi (cioè sia prima sia dopo il periodo mediano di risposta). Si noti che, anche in questo caso, la misurazione della pressione non è obbligatoria per questo tipo di decisione.

Tab. 1 - Sintesi dei risultati

					Versione con ricerca				
		Versione semplice			*No ricerca*		*Ricerca*		
Gruppo	*Opzione*	*Accetta*	*Rifiuta*	*% accetta*	*Accetta*	*Rifiuta*	*Accetta*	*Rifiuta*	*% accetta*
Infermieri	Donazione	38	48	44	0	26	55	4	65
Urologi	Intervento chirurgico	97	136	42	49	46	10	123	26
Accademici	Proseguire il volo	32	268	11	6	208	51	9	21

4. Discussione

In medicina, negli affari e in altri ambiti, le persone tendono ad assumere che avere più informazioni non possa nuocere. A fondamento di questa assunzione c'è la credenza che le persone abbiano preferenze chiare e stabili, e che divenire più informati possa soltanto raffinarle ulteriormente (Kassirer, 1989). La ricerca psicologica, d'altra parte, suggerisce che le preferenze degli

individui sono spesso poco chiare e tendono a essere costruite nel corso del processo decisionale (Slovic, 1995; Slovic, Lichtenstein, 1983; Shafir, Tversky, 1995). In conseguenza di ciò, le preferenze possono essere significativamente alterate da vari aspetti del contesto (Ubel, 1999; Shafir, Simonson, Tversky, 1993). Così come talvolta accade nel caso degli incerti confini fra insegnamento e promozione commerciale, è difficile dire a che punto esattamente finisce la chiarificazione e inizia la distorsione. Le priorità degli individui sono malleabili e una piccola differenza nelle condizioni può talvolta alterare le preferenze delle persone e condurre a decisioni divergenti.

Il presente studio suggerisce che uno dei fattori che influenza le decisioni dei medici è il loro stesso comportamento. Usando scenari ipotetici, abbiamo mostrato che la ricerca di un'informazione può indurre medici e infermieri a dare a quell'informazione un peso maggiore di quello che vi avrebbero assegnato se fosse stata disponibile fin dall'inizio. Questi risultati non possono essere attribuiti a mancanza di attenzione o a risposte insincere, che avrebbero prodotto esiti differenti da quelli osservati. Si tratta inoltre di risultati compatibili con studi precedenti sulle decisioni – sia con poste ipotetiche sia con poste reali – riguardo ai comportamenti d'acquisto, alla richiesta di ipoteche e a negoziazioni difficili (Tversky, Shafir, 1992a; Festinger, 1957; Kachelmeier, Shehata, 1992; Camerer *et al*., 1997; Bastardi, Shafir, 2000). Diverse forze psicologiche possono contribuire al fenomeno qui osservato, come l'influenza della dissonanza cognitiva e dei «costi affondati» (*sunk costs*). A nostro parere la nozione più pertinente è però quella di «*self-perception*» (Arkes, Blumer, 1985; Aronson, 1978; Bem, 1978): a quanto pare, i medici e gli infermieri hanno inferito, dal fatto stesso che essi avevano ricercato alcuni dati, che quei dati erano di importanza cruciale. Riteniamo che questo abbia significative implicazioni per la pratica clinica (si veda tab. 2).

Tab. 2 - Sintesi delle raccomandazioni

Raccomandazione	*Violazione*	*Esempio di errore*
Considera se le informazioni mancanti sono rilevanti	Cercare tutti i dati possibili dalla storia clinica e dall'esame fisico	Eseguire un esame rettale in un paziente che sta avendo una dissezione aortica acuta
Elabora un piano d'azione prima che i dati divengano disponibili	Rimandare l'elaborazione di un piano d'azione a dopo la raccolta dei dati	Ignorare la storia clinica del paziente nel momento in cui si agisce a seguito di un risultato inatteso in un test di laboratorio
Chiedi un parere di controllo a colleghi che possono rivedere i dati disponibili in modo imparziale	Trascurare nuovi punti di vista dopo la raccolta dei dati	Accantonare l'opinione di un collega perché «si conosce il paziente da più tempo»

Innanzi tutto, suggeriamo che i medici considerino la rilevanza delle informazioni mancanti prima di ricercarle. I medici capaci raccolgono molte informazioni; tuttavia, in alcuni casi, le informazioni mancanti possono essere irrilevanti e altre volte può essere opportuno acquisirle solo per ragioni indipendenti dal problema di scelta presente (per esempio, per scopi di ricerca) (Asch, Patton, Hershey, 1990). Nei casi in cui l'informazione non è strumentale per la decisione corrente, questo stato di cose deve essere riconosciuto e il processo di raccolta dei dati deve essere adattato di conseguenza (Tversky, Shafir, 1992a). Determinare la rilevanza di un dato ancora mancante può non essere semplice, ma dopo aver compiuto la relativa ricerca ci si può trovare in una prospettiva differente, sentendosi quindi obbligati ad agire in una particolare maniera. Per quanto l'avvertimento di non ricercare informazioni non strumentali possa suonare banale, i dati suggeriscono che di fronte a decisioni difficili può emergere la tendenza a farlo.

In secondo luogo, in situazioni in cui un elemento di informazione è non strumentale ma difficile da ignorare, i medici farebbero meglio a vincolarsi a un corso d'azione specificato in anticipo, cioè prima dell'acquisizione di quella informazione. In altre parole, i medici dovrebbero considerare l'ipotesi di elaborare un piano di contingenza per quando i dati diverranno disponibili. Questa strategia di comportamento incoraggia le persone a far fronte all'incertezza e a identificare gli aspetti che dovrebbero pesare realmente. Per contro, ritardando la decisione fino al momento in cui un'informazione diviene disponibile si rischia di dedicare un'attenzione eccessiva a quell'informazione. In molte situazioni, inoltre, la pianificazione anticipata è essenziale per lasciare ai pazienti e alle famiglie il tempo necessario per contribuire alla decisione. Trovare dei modi per insegnare alle persone a far fronte all'incertezza è una priorità per la futura ricerca.

In terzo luogo, in situazioni in cui le informazioni mancanti sono state ricercate senza un chiaro piano specificato in anticipo, il medico potrebbe consultare un collega che può facilmente esaminare i dati senza l'influenza potenzialmente dannosa di un coinvolgimento diretto nella ricerca di quei dati. È difficile immaginare quale influenza avrebbe avuto l'informazione che è stata ricercata se fosse stata nota fin dall'inizio, ma un collega può offrire un punto di vista libero degli investimenti fatti durante una lunga e complessa ricerca. Questa procedura è analoga a quella per cui una commissione valuta i candidati per un trapianto dopo che ogni paziente è stato complessivamente valutato dal suo medico. I colleghi a cui vengono presentati tutti i dati in una sola volta spesso vedono la situazione in modo differente dal medico che ha ottenuto quegli stessi dati attraverso sforzi graduali.

Un limite della nostra ricerca è il basso tasso di risposte ottenuto nei nostri campioni. Tuttavia è improbabile che questa debolezza dello studio possa spiegare i risultati. La struttura delle indagini presentate valuta le discrepanze

fra due versioni dello stesso scenario, evitando inferenze relative agli orientamenti complessivi di scelta, e le discrepanze non possono essere spiegate attraverso la disponibilità a rispondere. Oltre a ciò, studi precedenti hanno riscontrato discrepanze indipendentemente dal fatto che i tassi di risposta fossero alti o bassi e la discrepanza più ampia che abbiamo rilevato in questa indagine si è manifestata addirittura nel campione con il più alto tasso di risposte. Inoltre i risultati ottenuti sarebbero ancora statisticamente significativi in ciascun caso anche se tutti coloro che non hanno risposto si esprimessero nello stesso identico modo (non importa se in favore di una o dell'altra alternativa di scelta). Infine, occorre osservare che nei nostri questionari l'acquisizione di un dato con un leggero ritardo simulava processi di ricerca delle informazioni che nelle situazioni mediche del mondo reale possono essere ben più intensi e pressanti.

Una caratteristica interessante dei nostri scenari è che persone ragionevoli possono essere in disaccordo su quale sia la decisione corretta. Tuttavia la discrepanza fra le risposte date nelle versioni semplici e nelle versioni con ricerca è preoccupante. Sebbene sia necessario svolgere ulteriori indagini in proposito, riteniamo che le scelte compiute a seguito di una ricerca di informazioni offrano particolari ragioni di preoccupazione. Una di tali ragioni è che è possibile acquisire le stesse informazioni seguendo percorsi diversi che possono dar luogo a decisioni divergenti. La versione semplice, per contro, offre una singola prospettiva coerente per chiunque sia chiamato a decidere. Inoltre la ricerca inopportuna di informazioni ha i tratti di un «errore che si cancella da sé» (Sherman, 1980). In altri termini, le persone possono inizialmente commettere un errore ricercando informazioni non strumentali, ma in seguito compiere scelte che conferiscono a quelle informazioni valore strumentale, «cancellando» così l'errore. In nessun punto di questo processo chi decide è in condizione di riconoscere che un errore è stato commesso.

Donald A. Redelmeier è stato finanziato da un Career Scientist Award del Ministero della Salute dello Stato dell'Ontario, dal de Souza Chair in Trauma presso l'Università di Toronto e da un contributo alla ricerca della Physicians' Services Incorporated Foundation. Prince S. Aujla è stato finanziato da una borsa di studio del Medical Research Council.

*10. Assolutamente relativo: come la presentazione dei risultati della ricerca medica influenza le scelte di trattamento**

di *Lachlan Forrow, William C. Taylor, Robert M. Arnold*

La comunicazione dei risultati della ricerca clinica ai medici rappresenta la base per il progresso della medicina (Comroe, 1978). Se gli studi sono riportati in maniera inappropriata o in modo tale che i risultati possano essere facilmente fraintesi, la comunicazione risulta distorta (Feinstein, 1988). Fra gli studi pubblicati, i più importanti per la medicina clinica sono quelli che documentano gli effetti dei trattamenti. Diversi autori hanno offerto utili suggerimenti per migliorare accuratezza e chiarezza nella diffusione dei risultati di studi di questo tipo (Chalmers *et al.*, 1981; DerSimonian *et al.*, 1982; Feinstein, 1983; Pocock, Hughes, Lee, 1987; MacMahon, Pugh, 1970; Feinstein, 1985; Laupacis, Sackett, Roberts, 1988). Accuratezza e chiarezza, tuttavia, non garantiscono che tali risultati saranno interpretati correttamente; gli errori nell'interpretazione di informazioni di tipo quantitativo sono infatti comuni (Kahneman, Slovic, Tversky, 1982; Casscells, Schoenberger, Graboys, 1978; Berwick, Fineberg, Weinstein, 1981; Haynes, Sackett, Tugwell, 1983; cap. 7 in questo volume)[1]. Molti autori hanno ipotizzato che differenti presentazioni dei risultati delle ricerche possano condurre a diverse scelte di trattamento e hanno espresso la specifica preoccupazione che i resoconti che enfatizzano l'impatto di un trattamento sugli esiti clinici in termini relativi possano indurre i lettori ad attribuire a quel trattamento una efficacia maggiore di quella che vi attribuirebbero se, nel comunicare gli stessi risultati, si fosse riportato l'impatto del trattamento in termini assoluti (Leblond, 1989; Pickering, 1983; Weissler, Miller, Boudoulas, 1989). Il presente studio mira a determinare se tale preoccupazione è giustificata.

* Traduzione di *Absolutely relative: how research results are summarized can affect treatment decisions*, «American journal of medicine», 92 (1992), pp. 121-124, copyright © 1992 Excerpta Medica.

1. Si vedano anche, in questo volume, i capp. 1, 11 e 12 [N.d.C.].

1. Partecipanti e metodi

Abbiamo preparato due questionari, uno relativo al trattamento dell'ipertensione, l'altro relativo al trattamento dell'ipercolesterolemia. Ciascun questionario presentava in sei brevi asserzioni alcune informazioni ricavate da studi clinici pubblicati. Dopo ognuna di queste asserzioni, ai partecipanti veniva chiesto: «In che modo l'aver appreso questa informazione influenzerebbe le tue decisioni nel trattare i tuoi pazienti affetti da ipertensione [ipercolesterolomia]?». I soggetti indicavano la loro risposta su una scala Likert (Likert, 1932), che conteneva sette etichette comprese fra i due estremi «L'adozione di questo trattamento sarebbe indubbiamente più probabile» e «L'adozione di questo trattamento sarebbe indubbiamente meno probabile». Il questionario includeva anche domande relative ad alcuni dati sociodemografici dei partecipanti.

In ogni questionario, in due delle sei domande venivano riportati i risultati di uno stesso studio in due modi differenti (si veda tab. 1). In un caso tali risultati erano descritti nei termini del cambiamento assoluto delle probabilità dei diversi esiti clinici (si vedano i testi HA e CA in tab. 1); nell'altro caso i risultati dello studio erano descritti nei termini del cambiamento relativo di quelle stesse probabilità (si vedano i testi HB e CB in tab. 1). I dati citati provenivano dalle pubblicazioni originali (Hypertension detection and follow-up program cooperative group, 1979; Lipid research clinics program, 1984). L'ordine di presentazione di queste due domande era invertito nella metà dei questionari. Ogni soggetto ha completato il questionario relativo a uno soltanto dei due problemi medici (ipertensione o ipercolesterolomia).

Abbiamo distribuito i questionari nel 1987 durante alcuni incontri di formazione tenutisi in dieci ospedali, quattro dei quali erano ospedali universitari. I questionari sono stati inoltre inviati per posta o distribuiti personalmente ai partecipanti e agli insegnanti di sette programmi di corso biennali relativi all'applicazione dell'epidemiologia clinica e dei metodi di indagine delle scienze sociali allo studio della medicina.

In questo studio abbiamo controllato l'ipotesi che le decisioni dei medici riguardo ai trattamenti differiscano a seconda che le informazioni siano presentate in termini di cambiamenti relativi o assoluti delle probabilità degli esiti clinici. Abbiamo anche indagato se le risposte dei medici differivano in funzione della partecipazione a programmi di formazione avanzata in epidemiologia clinica o altri metodi di ricerca, della loro specialità, del tipo di pratica, del livello di formazione clinica e del genere. Le analisi sono state svolte impiegando il software SAS/STAT (SAS/STAT, 1988). Si è utilizzato il Wilcoxon signed-rank test per confrontare le risposte dei medici alle due coppie di domande in ciascun questionario e il test Mantel-Haenszel χ^2 per analizzare le differenze nelle risposte fra diversi sottogruppi di medici (Moses, Emerson, Hosseini, 1986; SAS/STAT, 1988).

Tab. 1 - Resoconti relativi all'ipertensione e all'ipercolesterolemia

*Resoconti relativi all'ipertensione**

HA. Uno studio randomizzato su più di 6.000 uomini con una forma moderata di ipertensione (pressione sanguigna diastolica: 90-104) ha mostrato che il trattamento farmaceutico sperimentale riduceva la mortalità complessiva in cinque anni dal 7,8% nel gruppo di controllo (sottoposto al trattamento di routine) al 6,3% – una riduzione della mortalità totale statisticamente significativa e pari all'1,5% in cinque anni.

HB. Quando 7.825 uomini sono stati coinvolti in uno studio randomizzato, uno speciale programma di trattamento farmacologico per l'ipertensione moderata (pressione sanguigna diastolica: 90-104) ha ridotto il tasso di mortalità complessivo del 20,3% in confronto a quello nel gruppo di controllo che riceveva il trattamento di routine (la differenza era statisticamente significativa).

A seguito di ciascuno dei resoconti, ai medici veniva chiesto: «In che modo l'aver appreso questa informazione influenzerebbe le tue decisioni nel trattare i tuoi pazienti affetti da ipertensione?»

*Resoconti relativi all'ipercolesterolemia***

CA. Quando in un accurato controllo randomizzato 1.900 uomini a cui era stato somministrato un farmaco per ridurre i livelli di colesterolo sono stati confrontati con 1.906 uomini a cui era stato somministrato un placebo, si è rilevato un tasso di mortalità per patologie coronariche in sette anni del 2,0% nel gruppo placebo e dell'1,6% nel gruppo trattato farmacologicamente, con una riduzione dello 0,4% in sette anni del tasso di mortalità per patologie coronariche (la differenza era statisticamente significativa).

CB. Un accurato studio randomizzato ha confrontato un trattamento farmaceutico con un placebo in quasi 4.000 uomini ipercolesterolemici di mezza età e ha riportato una riduzione relativa (statisticamente significativa) del 24% nel tasso di mortalità per patologie coronariche in sette anni.

A seguito di ciascuno dei resoconti, ai medici veniva chiesto: «In che modo l'aver appreso questa informazione influenzerebbe le tue decisioni nel trattare i tuoi pazienti affetti da ipercolesterolemia?»

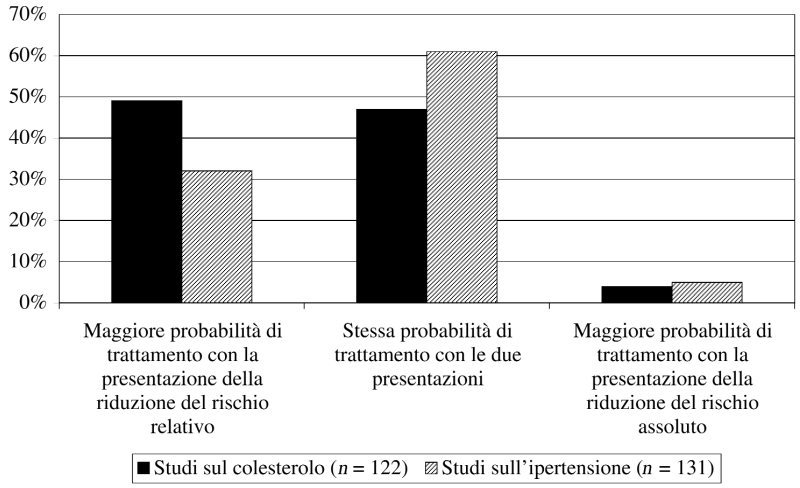
* Adattati da Hypertension detection and follow-up program cooperative group (1979).
** Adattati da Lipid research clinics program (1984).

2. Risultati

Sono stati completati 235 questionari (122 relativi all'ipocolesterolemia, 113 relativi all'ipertensione). Il tasso di risposte è stato variabile: fra il 30% e il 50% in incontri con numerosi partecipanti (127 soggetti) e approssimativamente del 75% in seminari con un numero ridotto di partecipanti (78 soggetti) e fra i questionari distribuiti per posta (30 soggetti). Fra i 196 partecipanti che hanno indicato il loro livello di formazione, 42 (21%) erano specializzandi fino a cinque anni dalla laurea in medicina, 50 (26%) erano medici in corso di specializzazione avanzata fino a nove anni dalla laurea e 104 (53%)

avevano completato interamente la loro formazione medica. Fra i 139 che hanno identificato la loro area di specializzazione, 72 (52%) hanno indicato la medicina interna generale, 25 (18%) erano medici di famiglia e 18 (13%) hanno indicato una sottospecialità di medicina interna. Cinquantadue dei 235 partecipanti erano partecipanti o insegnanti di programmi di corso biennali relativi all'applicazione dell'epidemiologia clinica e dei metodi di ricerca delle scienze sociali allo studio della medicina. Il periodo trascorso dal completamento degli studi universitari era in media di nove anni. L'80% dei partecipanti erano maschi.

Fig. 1 - Confronti fra le risposte dei medici ai resoconti che descrivevano i risultati dello stesso studio in due modi differenti. Si veda tab. 1 per i resoconti che erano stati presentati ai medici



In fig. 1 sono riassunte le risposte che i partecipanti hanno dato alla due coppie di domande. In 122 questionari relativi all'ipercolesterolemia, 57 soggetti (46,7%) hanno dato la stessa risposta alle due domande; 60 (49,2%) hanno indicato che la loro probabilità di adottare il trattamento a fronte della presentazione dei dati in termini relativi sarebbe stata più alta della loro probabilità di adottarlo a fronte della presentazione in termini assoluti degli stessi dati; 5 (4,1%) si sono invece detti più inclini al trattamento a fronte della presentazione in termini assoluti ($p < 0,0001$). Nei 113 questionari relativi all'ipertensione, 70 soggetti (61,9%) hanno risposto nello stesso modo alle due domande; 37 (32,7%) hanno indicato che la loro probabilità di adottare il trattamento a fronte della presentazione dei dati in termini relativi sarebbe stata più alta della loro probabilità di adottarlo a fronte della presentazione in

termini assoluti degli stessi dati; infine, 6 (5,3%) si sono detti più inclini al trattamento a fronte della presentazione in termini assoluti ($p < 0{,}0001$).

Non si sono registrate discrepanze significative fra sottogruppi che differivano rispetto alla formazione avanzata in metodi di ricerca, alla specialità, al livello di formazione clinica (completa o non completa), al tipo di pratica (accademica oppure privata) e al genere.

3. Commenti

Come già notato da Aristotele nella sua *Retorica*, vi sono diversi «modi possibili per persuadere le persone riguardo a un dato argomento». Più recentemente, la ricerca nel campo della psicologia cognitiva ha ampiamente mostrato che il modo in cui informazioni di tipo quantitativo vengono presentate può esercitare una profonda influenza sul modo in cui quelle informazioni verranno probabilmente interpretate (Kahneman, Slovic, Tversky, 1982; Tversky, Kahneman, 1974; Tversky, Kahneman, 1981). I medici, come le persone non esperte, possono essere vittime di errori e incoerenze nell'interpretare informazioni di tipo quantitativo (Casscells, Schoenberger, Graboys, 1978; Berwick, Fineberg, Weinstein, 1981; Haynes, Sackett, Tugwell, 1983; Dawson, Arkes, 1987; cap. 7 in questo volume).

Nel nostro studio, alcune differenze nel modo in cui si possono riportare i risultati di due studi clinici importanti e influenti hanno indotto nei medici giudizi diversi riguardo al trattamento dei pazienti. Cambiando il modo in cui gli esiti di uno studio erano descritti abbiamo osservato che quasi la metà dei partecipanti (108 su 235, cioè il 46%) ha valutato in modo differente la probabilità di adottare il relativo trattamento per i loro pazienti. Fra i soggetti per i quali tale differenza è emersa, l'89,8% (97 su 108) hanno dichiarato che sarebbero stati più inclini ad adottare il trattamento per i loro pazienti a fronte di una presentazione dei dati in termini di riduzione del rischio relativo che a fronte di una presentazione degli stessi dati in termini di riduzione del rischio assoluto. Concludiamo, quindi, che per un'efficace comunicazione dei dati medici non è sufficiente che i resoconti dei risultati delle ricerche siano tecnicamente accurati.

Sebbene si sia mostrato che la forma in cui i risultati delle ricerche cliniche sono presentati può produrre interpretazioni incoerenti fra i medici, il nostro studio non permette di determinare l'entità di questo problema, i suoi effetti diretti sulla cura dei pazienti e le più appropriate strategie correttive. Negli esempi che abbiamo impiegato, le differenze fra gli effetti dei trattamenti in termini assoluti (0,4% e 1,5%) e in termini relativi (24% e 20,3%) era particolarmente significativa. Si tratta di una situazione comune nei casi in cui le probabilità degli esiti clinici di interesse sono espresse da piccole cifre, come

accade in molti interventi di prevenzione. L'estensione dei nostri risultati potrebbe essere meno immediata in situazioni in cui le probabilità degli esiti clinici di interesse è espressa da numeri più alti e in cui, di conseguenza, le differenze fra cambiamenti in termini assoluti e relativi sono meno drammatiche. Ciononostante, se la scelta di una particolare forma di presentazione dei dati piuttosto che un'altra può influenzare anche occasionalmente il giudizio clinico di un medico a seguito della lettura di una pubblicazione scientifica, allora la presentazione dei dati statistici nella letteratura medica dovrebbe essere oggetto di scrupolosa attenzione.

Ci troviamo d'accordo con altri autori nel ritenere che una migliore formazione dei medici nell'analisi critica e nelle tecniche quantitative di elaborazione dei dati è di importanza essenziale (Berwick, Fineberg, Weinstein, 1981; Sackett, Haynes, Tugwell, 1985). Tuttavia nel presente studio abbiamo osservato che i medici con una formazione avanzata in tecniche di ricerca e di analisi hanno reagito in modo simile agli altri partecipanti. Ne concludiamo che questo tipo di formazione, almeno nella sua forma attuale, non rappresenta una risposta adeguata al problema che abbiamo identificato (Redelmeier, 1991).

Perché la letteratura medica serva in modo efficace e affidabile come base per scelte di trattamento razionali, le misure dell'efficacia dei trattamenti dovrebbero essere definite in modo da ridurre o eliminare decisioni incoerenti dovute a differenti presentazioni dei dati statistici. Il modo in cui i risultati vengono riportati dovrebbe essere sottoposto a un'analisi critica più attenta prima della pubblicazione degli studi. Laupacis, Sackett, Roberts (1988) hanno offerto un'utile discussione su vantaggi e svantaggi di diverse misure degli effetti dei trattamenti. Per quanto riguarda i cambiamenti relativi o assoluti nelle probabilità dei diversi esiti clinici, si è osservato che ciascuna delle due forme di presentazione fornisce informazioni potenzialmente utili (MacMahon, Pugh, 1970; Feinstein, 1985). Suggeriamo quindi che gli autori degli studi le riportino entrambe. Nella descrizione dei benefici attesi per pazienti individuali, tuttavia, è opportuna – e dovrebbe pertanto essere richiesta – una maggiore enfasi sull'impatto dei trattamenti in termini di rischio assoluto (MacMahon, Pugh, 1970; Laupacis, Sackett, Roberts, 1988).

Siamo grati a Donald Berwick, Jay J.J. Christensen-Szalanski, Paul Cleary, William Landau, Julia C. Nissen, Thomas Risser, Marylin Spellmayer, Milton Weinstein e i *referee* anonimi di questo articolo per i loro commenti e suggerimenti.

*11. Finanziamenti basati sull'evidenza: la comprensione dei risultati della ricerca clinica**

di *Tom Fahey, Sian Griffiths, Tim J. Peters*

1. Introduzione

Gli studi clinici randomizzati sono considerati il *gold standard* per la valutazione degli interventi medico-sanitari (Sackett *et al.*, 1991). Il loro potere esplicativo permette di concludere, in termini qualitativi, se un particolare trattamento funziona e fornisce indicazioni di tipo quantitativo per valutare in che misura esso funzioni (Sackett, Cook, 1994). Le organizzazioni che gestiscono fondi di finanziamento sono chiamate a effettuare scelte di politica sanitaria basate sull'evidenza scientifica disponibile in relazione all'efficacia clinica degli interventi e gran parte dell'evidenza scientifica rilevante proviene da studi randomizzati (NHS Management Executive, 1993; Ham, Hunter, Robinson, 1995). Le ricerche mostrano che il modo in cui i risultati degli studi vengono presentati influenza in misura significativa la disponibilità dei medici a prescrivere farmaci e la percezione dei pazienti riguardo ai benefici dei trattamenti (Laupacis, Sackett, Roberts, 1990; Naylor, Chen, Strauss, 1992; Farkouh, *et al.*, 1993; Bobbio, Demichelis, Giustetto, 1994; Bucher, Weinbacher, Gyr, 1994; Malenka *et al.*, 1993; cap. 10 in questo volume). Finora non si è però indagato se la presentazione dei risultati degli studi clinici influenzi le decisioni di politica sanitaria.

Vi sono almeno quattro modi diversi per presentare i risultati di uno studio clinico: la riduzione del rischio relativo di un esito negativo, la riduzione del rischio assoluto di un esito negativo, la proporzione di pazienti senza esiti negativi e il numero di pazienti che è necessario trattare per evitare un esito negativo (Laupacis, Sackett, Roberts, 1990). Molto spesso nel riportare i risultati di uno studio si utilizza una sola misura dell'efficacia clinica, solitamente la

* Traduzione di *Evidence based purchasing: understanding results of clinical trials and systematic reviews*, «British medical journal», 311 (1995), pp. 1056-1059. Per gentile concessione di BMJ Publishing Group.

riduzione del rischio relativo. Sfortunatamente, però, i dati riguardanti il rischio relativo non forniscono alcuna informazione sulla diffusione della condizione clinica alla quale il trattamento si rivolge, vale a dire sulla suscettibilità della popolazione generale a quella specifica condizione. Nei casi in cui i benefici relativi sono notevoli e la condizione clinica di interesse è rara, la riduzione del rischio relativo può essere alta, mentre in termini assoluti il valore del trattamento resta modesto. Questo problema viene comunemente trascurato nei commenti e negli editoriali, che discutono i risultati degli studi enfatizzando appunto le differenze relative a discapito degli effetti assoluti. Come ha osservato Feinstein (1992), «i medici vengono impressionati molto di più dai grandi numeri di un risultato espresso in termini relativi che non dalle più modeste dimensioni che lo stesso effetto assume se espresso in termini assoluti».

Con l'avvento della *evidence-based medicine* la comprensione dei metodi quantitativi per la presentazione dei dati degli studi clinici è della massima importanza in vista di decisioni politiche informate (Sackett, Cook, 1994). Il nostro obiettivo è determinare se le decisioni di politica sanitaria siano influenzate dal modo in cui i risultati degli studi clinici e delle rassegne sistematiche vengono presentati.

2. Metodi

Abbiamo spedito un questionario a 182 membri (parte dei quali con responsabilità di dirigenza) di istituzioni, commissioni e servizi di base operanti in ambito sanitario nella regione dell'Anglia e di Oxford. Queste istituzioni sono incaricate di definire le politiche sanitarie nelle loro aree geografiche e sono composte da persone con un'ampia varietà di competenze. I membri con funzioni dirigenziali sono responsabili della gestione delle risorse finanziarie e del personale. I direttori delle strutture hanno una specifica formazione epidemiologica; la valutazione dell'efficacia degli interventi medici e dei bisogni della popolazione locale rientra fra le loro responsabilità. I membri senza funzioni dirigenziali (in maggioranza non medici) sono designati dal presidente dell'autorità sanitaria regionale per conto della Segreteria di Stato e agiscono come rappresentanti pubblici della popolazione locale.

I questionari sono stati costruiti in modo da presentare i risultati di un singolo studio clinico sull'efficacia dello screening mammografico per il cancro al seno (Tabar *et al.*, 1985) e i risultati di una singola rassegna sistematica relativa all'efficacia della riabilitazione cardiaca (Oldridge *et al.*, 1988). Per ognuno dei due interventi i dati sono stati presentati in quattro diverse forme, nell'ordine seguente: riduzione del rischio relativo di morte, riduzione del rischio assoluto di morte, proporzione di pazienti senza esito negativo (morte) e numero di pazienti che è necessario trattare per evitare un esito negativo (morte).

La riduzione del rischio assoluto (o differenza assoluta) equivale a $x - y$, dove x è la proporzione dei pazienti che sperimentano un particolare esito clinico negativo nel gruppo di controllo e y è la proporzione dei pazienti che sperimentano quello stesso esito nel gruppo sottoposto al trattamento che si intende studiare (Guyatt *et al.*, 1994). La diffusione dell'esito clinico considerato nel gruppo di controllo, vale a dire x, riflette l'entità generale del rischio nella popolazione. Il numero di pazienti che è necessario trattare per evitare un esito negativo è il reciproco della differenza assoluta del rischio, cioè $1/(x - y)$, mentre la riduzione del rischio relativo equivale a $[(x - y)/x] \times 100$. Infine, la proporzione di pazienti senza esiti negativi è data dalle percentuali di pazienti che, rispettivamente, nel gruppo di controllo e in quello sperimentale non incorrono nell'esito negativo considerato. La tab. 1 illustra nei dettagli i vari metodi di presentazione dei dati attraverso l'esempio di una rassegna sistematica sull'efficacia dell'introduzione di un *by-pass* arterocoronarico (l'efficacia è misurata nei termini della riduzione della mortalità).

Tab. 1 - Metodi per la presentazione dei risultati di un controllo clinico randomizzato o di una rassegna sistematica (i calcoli sono adattati da Guyatt et al.*, 1994; l'esempio utilizzato è Yusuf* et al.*, 1994)*

	Esito clinico a 10 anni		
Gruppo	*Morte*	*Sopravvivenza*	*Totale pazienti randomizzati in ciascun gruppo*
Introduzione di *by-pass* arterocoronarico	350	974	1.324
Terapia medica	404	921	1.325

Rischio di morte in 10 anni con terapia medica (gruppo di controllo): $x = 404/1325 = 0{,}305$ oppure 30,5%.
Rischio di morte in 10 anni con l'introduzione di *by-pass* arterocoronarico (gruppo sperimentale): $y = 350/1324 = 0{,}264$ oppure 26,4%.
Rischio relativo: $y/x = 0{,}87$.
Riduzione del rischio relativo: $[(x - y)/x] \times 100$ oppure $[1 - y/x] \times 100 = 13{,}3\%$.
Riduzione del rischio assoluto (o differenza del rischio): $x - y = 0{,}04$ oppure 4%.
La proporzione di pazienti senza esiti negativi è in questo caso la percentuale assoluta di individui che, in ciascun gruppo, sopravvive a 10 anni. Nel gruppo di controllo equivale a $(1 - x) \times 100 = 69{,}5\%$; nel gruppo sperimentale equivale a $(1 - y) \times 100 = 73{,}6\%$.
Numero di pazienti che è necessario trattare per evitare una morte: $1/(x - y) = 1/0{,}04 = 25$.

Nel nostro questionario, ciascuna delle diverse forme di presentazione veniva indicata come il risultato di un diverso «studio» (si veda l'Appendice). Ai partecipanti si chiedeva di tracciare una «X» su una scala lineare compresa da «0» («non sosterrei il finanziamento di questo intervento») a «10» («so-

sterrei con forza il finanziamento di questo intervento») relativa alla decisione di finanziare un programma di screening mammografico e alla decisione di finanziare un programma di riabilitazione cardiaca. Non si chiedeva di confrontare fra loro i due programmi. In ciascun questionario si precisava che la decisione di finanziare ciascun programma avrebbe dovuto dipendere esclusivamente dalla forza dei risultati ricavati da ciascuno «studio» e che i costi di tutte le alternative considerate erano identici. Si chiedeva infine ai partecipanti di esprimersi su ciascun programma considerato ignorando le politiche locali o nazionali in vigore al momento.

Abbiamo calcolato i punteggi medi (e l'intervallo di fiducia al 95%) per ogni presentazione alternativa dei risultati. Per ciascun programma di intervento, abbiamo svolto l'analisi della varianza per misurazioni ripetute con la procedura MANOVA di SSPS per Windows (SSPS, 1993). Abbiamo infine controllato le differenze fra le quattro diverse presentazioni relative a ciascun programma attraverso confronti multipli a coppie con la correzione di Bonferroni. Tutte le analisi svolte tengono pienamente conto degli effetti dovuti alle differenze individuali fra i partecipanti.

3. Risultati

In totale, sono stati restituiti 140 questionari, in tre dei quali i partecipanti hanno dichiarato che non erano in grado di rispondere perché ritenevano necessarie ulteriori informazioni. Il tasso di risposte è stato del 77% (79% per i dirigenti e 76% per i non dirigenti), variando fra 68% e 94% nelle diverse istituzioni.

La tab. 2 riporta le statistiche descrittive per i sei confronti a coppie fra le presentazioni alternative. Le decisioni relative al finanziamento sono state significativamente influenzate dal modo in cui i dati venivano presentati (tab. 3). Il test F generale riguardo alle differenze fra i quattro diversi metodi di presentazione dei dati è risultato altamente significativo ($p < 0{,}001$). Per entrambi i programmi, la presentazione in termini di riduzione del rischio relativo ha indotto, in media, una disponibilità a sostenere il finanziamento significativamente più alta di quella ottenuta con ciascuno dei tre restanti metodi di presentazione. Inoltre, la disponibilità media associata al numero necessario per curare è risultata significativamente maggiore di quella associata alla riduzione del rischio assoluto per entrambi i programmi e significativamente maggiore di quella associata alla proporzione di pazienti senza esiti negativi per il programma di riabilitazione cardiaca (tab. 3). Soltanto tre partecipanti (membri non dirigenti che asserivano di non avere una formazione epidemiologica) hanno dichiarato di essersi resi conto che, in relazione a ciascun programma, le quattro alternative si riferivano a risultati identici.

Tab. 2 - Statistiche descrittive per le differenze a coppie fra le quattro presentazioni alternative per ciascun programma

Confronto fra le alternative	*Media (deviazione standard) per la mammografia*	*Media (deviazione standard) per la riabilitazione cardiaca*
RRA-RRR	–4,10 (3,23)	–1,96 (2,78)
PSEN-RRR	–4,12 (3,04)	–4,12 (3,04)
NNC-RRR	–2,84 (2,80)	–1,34 (2,61)
PSEN-RRA	–0,022 (1,57)	–0,33 (1,74)
NNC-RRA	1,26 (2,92)	0,062 (2,40)
NNC-PSEN	0,95 (2,45)	1,28 (2,83)

RRA = riduzione del rischio assoluto; RRR = riduzione del rischio relativo; PSEN = proporzione di pazienti senza esiti negativi; NNC = numero necessario per curare.

Tab. 3 - Punteggio medio (%) (intervallo di confidenza al 95%) assegnato da membri di istituzioni sanitarie di fronte a quattro metodi di presentazione differenti degli stessi dati in due diversi problemi

Metodo di presentazione dei dati	*Riabilitazione cardiaca*	*Mammografia*
Opzione A (RRR)	79 (76-83)† (RRR = 34%)	76 (72-80)† (RRR = 20%)
Opzione B (RRA)	38 (35-42)*,** (RRA = 0,06%)	56 (53-60)* (RRA = 3%)
Opzione C (PSEN)	38 (34-42)*,** (PSEN = da 99,82% a 99,8%)	53 (49-57)*,** (PSEN = da 84% a 87%)
Opzione D (NNC)	51 (47-55)* (NNC = 1.592)	62 (58-66)* (NNC = 31)

† punteggio più alto (massimo sostegno per il finanziamento del programma).
Differenze di Bonferroni fra le medie: * differisce significativamente ($p \leq 0,05$) dalla riduzione del rischio relativo; ** differisce significativamente ($p \leq 0,05$) dal numero necessario per curare.

4. Discussione

Il presente studio mostra che l'interpretazione degli studi clinici e delle rassegne sistematiche è influenzata dal metodo impiegato nel riportare i dati. I responsabili di politiche sanitarie, così come i medici, sono maggiormente impressionati dai metodi di misurazione dell'efficacia clinica in termini relativi che non dagli indicatori dei benefici assoluti. Per entrambi i programmi, il punteggio medio ottenuto con la riduzione del rischio relativo è infatti risultato notevolmente più alto di quello ottenuto con ciascuno degli altri metodi di presentazione. L'entità di tale effetto si è inoltre rivelata maggiore quan-

do la suscettibilità generale della popolazione alla condizione clinica su cui si intendeva intervenire era minore (vale a dire, nel caso della morte da cancro al seno; si veda tab. 3).

I due scenari sono stati deliberatamente scelti in modo da riflettere la considerazione che i benefici di diversi interventi possono differire notevolmente in termini assoluti ma essere simili in termini relativi. Il rischio di morte è 85 volte maggiore in pazienti che sono già stati colpiti da infarto miocardico (16% in tre anni) che in donne di mezza età coinvolte in un programma di screening mammografico (0,19% in sette anni). Quando però vengono riportati solo gli effetti sul rischio relativo, l'ampia differenza dei benefici in termini assoluti non emerge dai risultati.

Come si può vedere, è essenziale che la presentazione dei risultati quantitativi degli studi clinici e delle rassegne sistematiche rispecchi tanto i benefici assoluti quanto quelli relativi (Laupacis, Naylor, Sackett, 1992). Questo è particolarmente importante nell'ambito delle politiche sanitarie, dove gli effetti sul rischio assoluto variano in modo considerevole a seconda che gli interventi si rivolgano a popolazioni in cui una determinata patologia è più o meno diffusa.

Il «numero necessario per curare» (NNC) gode di alcune utili proprietà che ne fanno un metodo versatile di presentazione dei dati. È semplice da calcolare, incorpora le informazioni relative al rischio assoluto e sono ora disponibili specifici metodi per adattarlo all'entità del rischio assoluto in popolazioni differenti (Cook, Sackett, 1995). Il NNC è inoltre utile nel confrontare l'efficacia di diverse forme di intervento, come trattamenti terapeutici, procedure chirurgiche e programmi di screening (Laupacis, Sackett, Roberts, 1988). Con questo metodo i costi e gli effetti collaterali di diverse forme di intervento possono essere convertiti in una stessa «valuta». Infine, in assenza di informazioni relative ai costi associati a programmi di intervento alternativi, il NNC è un indicatore del rapporto fra costi e benefici più realistico della riduzione del rischio relativo.

Alcuni studiosi hanno suggerito che i risultati degli studi randomizzati e delle rassegne sistematiche possano richiedere forme di presentazione differenti a seconda che la valutazione dei benefici avvenga a livello di politiche sanitarie o in relazione a un paziente individuale. Per esempio, una recente rassegna sistematica dei benefici dell'introduzione di *by-pass* arterocoronarici in pazienti con angina instabile riporta i risultati in due modi (Yusuf *et al.*, 1994): innanzi tutto attraverso la media aggregata dell'aumento della sopravvivenza per tutti i pazienti con *by-pass*, equivalente a 4,26 mesi in 10 anni; in secondo luogo attraverso dati disaggregati, secondo i quali il 16% dei pazienti con *by-pass* guadagna fra 1,5 e 3,5 anni di vita mentre il restante 86% non ottiene alcun cambiamento nell'aspettativa di vita (Hux, Naylor, 1994). Anche in questo caso, le ricerche suggeriscono che, sebbene le due presentazio-

ni siano equivalenti, i medici sono maggiormente impressionati dall'aumento dell'aspettativa di vita in forma disaggregata che da un'unica cifra media (Hux, Levington, Naylor, 1994).

Alla luce di tutti gli esempi citati, si dovrebbe riconoscere che decisioni cliniche e scelte politiche realmente informate devono basarsi sulla piena comprensione delle diverse forme di presentazione dei dati e dei loro limiti. Tali decisioni e scelte dipendono infatti dal modo in cui le conoscenze scientifiche vengono descritte ed effettivamente percepite, non dalle probabilità e dalle statistiche in se stesse (Feinstein, 1992).

Le istituzioni responsabili del finanziamento di interventi medico-sanitari vengono esortate a sostenere con forza trattamenti efficaci e a ridurre gli investimenti in trattamenti di cui le ricerche hanno mostrato l'inefficacia (Hayward, 1994). Con la creazione del Programma per lo sviluppo e la ricerca del Servizio Sanitario Nazionale[1] è probabile che i decisori politici si trovino di fronte a una quantità crescente di dati scientifici ricavati da studi clinici e metaanalisi (Ham, Hunter, Robinson, 1995). In particolare, le rassegne sistematiche della Cochrane Collaboration e le pubblicazioni del Centro per le rassegne sistematiche e la loro diffusione (*Centre for Reviews and Dissemination*) del Servizio Sanitario Nazionale richiedono la corretta comprensione dei metodi quantitativi e qualitativi impiegati per riportare i dati scientifici. La consapevolezza dei vantaggi e dei limiti delle diverse forme di presentazione dei dati è fondamentale per l'interpretazione appropriata dei risultati di queste indagini.

Il presente studio mostra la necessità di migliorare la comprensione dei modi di presentazione dei risultati degli studi clinici controllati e delle rassegne sistematiche da parte dei responsabili delle politiche sanitarie. Il Progetto per lo sviluppo di competenze per la valutazione critica fra i finanziatori (*Critical Appraisal Skills for Purchasers*, CASP) mira ad aiutare i responsabili dei finanziamenti e altre persone chiamate a prendere decisioni di tipo medico-sanitario nell'acquisizione delle competenze necessarie per valutare l'evidenza disponibile sull'efficacia dei trattamenti (Institute of Health Sciences, 1994). I risultati di questo studio suggeriscono che le istituzioni sanitarie dovrebbero considerare metodi opportuni per migliorare tali competenze fra i loro membri.

Vorremmo ringraziare per i loro utili suggerimenti su precedenti stesure dell'articolo David Sackett (professore di Epidemiologia, Centre for Evidence-Based Medicine, John Ratcliffe Hospital, University of Oxford) e il Dr. Ruairidh Milne (Consulente di medicina e sanità pubblica, Servizio Sanitario Regionale dell'Anglia e di Oxford).

1. Si intende il National Health Service britannico [N.d.T.].

Appendice

Questionario relativo alla mammografia

Presso la tua istituzione si sta considerando la proposta di offrire un programma di screening mammografico a donne fra i 50 e i 64 anni. Vi sono alcune incertezze riguardo all'efficacia di un programma di questo tipo. Ti presenteremo quattro asserzioni ricavate da quattro studi clinici randomizzati pubblicati in riviste mediche specialistiche. Sulla base di ciascuna asserzione dovresti indicare quanto è probabile che tu ti trovi d'accordo sull'adozione del relativo programma di screening. Assumi che i costi di ciascuno dei diversi programmi siano gli stessi. Ognuno dei risultati riportati si è rivelato statisticamente significativo.

In un periodo di sette anni:

- il programma A ha ridotto il tasso di mortalità da cancro al seno del 34%;
- il programma B ha prodotto una riduzione assoluta delle morti da cancro al seno dello 0,06%;
- il programma C ha aumentato la proporzione di pazienti che sopravvivono al cancro al seno dal 99,82% al 99,88%;
- il programma D ha richiesto che 1.592 donne venissero trattate per evitare una morte da cancro al seno.

Questionario relativo alla riabilitazione cardiaca

Presso la tua istituzione si sta considerando la proposta di offrire un programma di riabilitazione cardiaca a pazienti colpiti da attacco cardiaco (infarto miocardico). Vi sono alcune incertezze riguardo all'efficacia di un programma di questo tipo. Ti presenteremo quattro asserzioni ricavate da quattro studi clinici randomizzati pubblicati in riviste mediche specialistiche. Sulla base di ciascuna asserzione dovresti indicare quanto è probabile che tu ti trovi d'accordo sull'adozione del relativo programma di riabilitazione cardiaca. Assumi che i costi di ciascuno dei diversi programmi siano gli stessi. Ognuno dei risultati riportati si è rivelato statisticamente significativo.

In un periodo di tre anni:

- il programma A ha ridotto il tasso di mortalità del 20%;
- il programma B ha prodotto una riduzione assoluta della mortalità del 3%;
- il programma C ha aumentato la proporzione di pazienti che sopravvivono dall'84% all'87%;
- il programma D ha richiesto che 31 persone venissero trattate per evitare una morte.

III. Valutazioni e consigli

*12. Le consulenze sull'AIDS per persone a basso rischio**

di *Gerd Gigerenzer, Ulrich Hoffrage, Axel Ebert*

Nel 1987, a un convegno sull'AIDS, l'allora senatore della Florida Lawton Chiles riferì che, su ventidue donatori di sangue del suo stato ai quali era stato comunicato che erano risultati HIV-positivi nel test ELISA, sette si erano suicidati. Nella stessa pubblicazione medica che documenta questa tragedia si spiega al lettore che «anche se i risultati dei due esami dell'AIDS, ELISA e Western Blot, sono entrambi positivi, le probabilità che l'individuo sia infetto sono solo del 50%» (Stine, 1996, pp. 333, 338). Questa situazione può verificarsi quando gli individui con esami positivi appartengono a una popolazione a basso rischio, come quello dei donatori di sangue. La discrepanza fra ciò che le persone ritengono un test HIV-positivo significhi e il suo reale significato sembra quindi aver comportato un costo in termini di vite umane che si è aggiunto a quello dovuto alla malattia stessa. Uno degli obiettivi delle consulenze sull'AIDS è spiegare l'effettiva entità dei rischi a chi si sottopone a un controllo. Il presente lavoro tratta delle consulenze relative al significato di un possibile esito positivo che vengono fornite a individui a basso rischio in consultori pubblici della Germania prima dell'esecuzione del test dell'HIV. Ci concentreremo su tre domande: quali sono le informazioni che i consulenti comunicano agli utenti riguardo alla probabilità di una infezione da HIV dato un risultato positivo degli esami? Tali informazioni sono comunicate in modo che gli utenti possano comprenderle? Come si potrebbero migliorare la comunicazione delle informazioni e la loro stessa accuratezza?

1. Le consulenze con individui a basso rischio

Siamo interessati al tipo di consulenze fornite alla popolazione di utenti più numerosa: eterosessuali senza comportamenti a rischio (come l'uso di droghe per via endovenosa). Queste persone si sottopongono al test dell'HIV

* Traduzione di *AIDS counselling for low-risk clients*, «AIDS care», 10 (1998), pp. 197-211, copyright © 1998 Taylor & Francis Ltd. (www.tandf.co.uk/journals).

per varie ragioni: volontariamente, perché desiderano sapere se sono infetti prima di sposarsi, di avere figli, o per altri motivi; oppure perché è richiesto loro di farlo per poter usufruire di assicurazioni sanitarie o sulla vita o in quanto immigrati, militari, donatori di sangue o appartenenti ad altri gruppi a cui la legge impone l'esecuzione dell'esame. Il governo svedese, per esempio, ha incoraggiato a tal punto i controlli volontari che «le persone che hanno meno probabilità di essere infette si sottopongono all'esame in massa» (Månsson, 1990). L'obbligatorietà del test è in diversi paesi uno scenario possibile, a cui le compagnie di assicurazione ricorrono per proteggersi da eventuali perdite. Nel 1990, per esempio, l'allora Governatore dell'Arkansas Bill Clinton ha dovuto sottoporsi al test dell'HIV per ottenere il rinnovo della sua assicurazione sulla vita («Der Tagesspiegel», 9/16/1996, p. 4). Può anche accadere che individui con un basso rischio di infezione siano sottoposti al test dell'HIV non solo involontariamente, ma persino a loro insaputa. Si è per esempio avuta notizia di alcune grandi aziende di Bombay che hanno sottoposto i loro dipendenti a esami del sangue senza informarli che sarebbero stati controllati per l'infezione da HIV; se un dipendente risultava HIV-positivo, veniva licenziato («Der Tagesspiegel», 7/8/1996).

Le consulenze con individui a basso rischio richiedono particolare attenzione per la possibilità di risultati falsamente positivi, cioè per la possibilità che si ottenga un risultato positivo anche se l'individuo non ha effettivamente contratto il virus. La proporzione dei risultati falsamente positivi in una determinata popolazione è tanto maggiore quanto più è ridotta la prevalenza del virus in quella popolazione. In altre parole, se un individuo con comportamenti ad alto rischio risulta positivo al test, la probabilità che egli sia realmente infetto è molto alta, ma se un individuo a basso rischio risulta positivo, tale probabilità può restare al di sotto del 50%, come indicato in precedenza. Se le persone non vengono informate di ciò, tendono a credere che un risultato positivo significhi che sono infette con assoluta certezza. Il caso di un giovane di Dallas, riportato dalla stampa americana, rappresenta un esempio istruttivo. Quest'uomo era risultato HIV-positivo in un controllo di routine. Attraversò un periodo di depressione, rischiando il suicido, e si trasferì in California. Dopo circa diciotto mesi di angoscia, un medico californiano gli fece ripetere il test, che stavolta risultò negativo («Chicago Tribune», 3/5/1993). Se il giovane avesse deciso di suicidarsi, come avevano fatto i donatori di sangue della Florida, non si sarebbe forse mai scoperto che il risultato del suo primo esame era un falso positivo. Come si può vedere, se i consulenti informano adeguatamente le persone dell'eventualità di risultati falsamente positivi, è possibile evitare sofferenze emotive e salvare delle vite[1].

1. Nel loro lavoro dedicato al comportamento suicida e all'infezione da HIV, Catalan e Pugh (1995) concludono che «le idee suicide, l'effettivo compimento del suicidio e l'autolesionismo deliberato non sono infrequenti in persone affette da HIV» (p. 119). Gli stessi Catalan e Pugh osservano d'altra parte che i dati disponibili sono tutt'altro che conclusivi: molti resoconti sono

Non conosciamo alcuno studio che abbia indagato che cosa i consulenti dell'AIDS dicono agli utenti riguardo al significato di un risultato positivo dei loro test. Abbiamo riflettuto a lungo sulla metodologia più appropriata, considerando la possibilità di inviare ai consulenti dei questionari o di chiedere loro di partecipare a un'indagine di laboratorio basata su risposte a domande scritte. Abbiamo però accantonato l'impiego di questionari e altri metodi simili perché è possibile criticarne i risultati suggerendo che sono poco informativi riguardo a come le sessioni di consulenza si svolgono effettivamente. Per esempio, questi metodi sono stati criticati perché non permettono ai medici di formulare autonomamente le loro domande per acquisire informazioni aggiuntive né di far uso delle loro personali stime statistiche invece di quelle fornite dagli sperimentatori, e perché trascurano gli effetti di una reale preoccupazione per il paziente, dal momento che i casi presentati sono del tutto ipotetici oppure sono casi storici risoltisi molti anni prima (Phelps, Shanteau, 1978; Yates, 1990).

Abbiamo quindi deciso di seguire un percorso più diretto. Uno di noi si è presentato in venti consultori, ha partecipato a una serie di sessioni di consulenza ed è stato sottoposto a una serie di test dell'HIV. Eravamo interessati a una specifica questione che i consulenti devono spiegare agli utenti: che cosa significa un risultato positivo del test? Per rispondere a questa domanda, è necessario conoscere (a) il tasso di base, o prevalenza, dell'infezione da HIV fra i maschi eterosessuali senza comportamenti a rischio, (b) la probabilità che il test risulti positivo nel caso che l'individuo sia infetto, detta *sensibilità* del test, e (c) la probabilità che il test sia positivo nel caso che l'individuo non sia infetto, detta *tasso di falsi positivi* (equivalente a: 1 – *specificità* del test). Sulla base di queste informazioni, è possibile stimare che cosa significhi un risultato positivo, vale a dire stimare la probabilità dell'infezione dato un risultato positivo, detta *valore predittivo positivo* (VPP). Ricaviamo innanzi tutto dalla letteratura le migliori stime disponibili di questi valori.

1.1. *Prevalenza*

La Germania ha un numero relativamente basso di casi di AIDS documentati. Alla fine del 1995 la cifra cumulativa era di 13.665 in confronto a circa 30.000 in Italia, 38.000 in Francia e più di 500.000 negli Stati Uniti (WHO Global AIDS Statistics, 1996). Si potrebbe quindi assumere che la prevalenza dell'HIV sia

aneddotici o riguardano piccoli numeri, i risultati variano fra paesi differenti e una serie di problemi metodologici rende difficili confronti controllati fra gruppi diversi (si vedano, per esempio, Pugh, O'Donnell, Catalan, 1993, e Marzuk, Perry, 1993). Un recente studio prospettivo di coorte ha rilevato un incremento di un fattore 1,35 dei suicidi in un gruppo di HIV-positivi in confronto a un gruppo di HIV-negativi, mentre studi precedenti avevano riportato un incremento del rischio di suicidio fra 7 e 36 volte in individui HIV-positivi (Dannenberg *et al*., 1996).

anch'essa relativamente bassa. Nel nostro studio, l'individuo sottoposto al test era un uomo di 27 anni, eterosessuale e senza comportamenti a rischio. Qual è in Germania la prevalenza dell'HIV negli eterosessuali fra 20 e 30 anni senza comportamenti a rischio? Una stima ragionevole è di circa 1 su 10.000, equivalente allo 0,01%[2]. Questa cifra rientra nei margini della prevalenza stimata dell'HIV fra i donatori di sangue degli Stati Uniti (un sottogruppo con una bassa prevalenza nella popolazione americana), che varierebbe fra 1 su 10.000 (Busch, 1994, pp. 229) e 2 su 10.000 (George, Schochetman, 1994, p. 90).

1.2. *Sensibilità e specificità*

In Germania, come nella maggior parte dei paesi occidentali, il test dell'HIV prevede il procedimento seguente. Se un primo test, ELISA, è negativo, si informa l'interessato che egli è HIV-negativo. Se è positivo, si effettua

2. Si tratta di una stima approssimativa, dal momento che non sembra siano disponibili in letteratura dati relativi alla prevalenza dell'HIV in uomini tedeschi fra i 20 e i 30 anni senza comportamenti a rischio. Il valore qui riportato si basa su due approssimazioni. La prima consiste nello stimare la prevalenza ignota sulla base della prevalenza nota fra gli uomini che donano il sangue per la prima volta (non, però, fra i donatori abituali, che rappresentano un gruppo altamente selezionato). La proporzione di HIV-positivi su circa 130.000 uomini alla loro prima donazione è stata di 1,5 su 10.000 (1986-1991, Stato del Baden-Württenberg; si veda Mauer *et al.*, 1993). Per un confronto, si consideri che la proporzione fra i donatori abituali era inferiore di un ordine di grandezza: circa 1,2 su 100.000 (Maurer, Kiehl, Altmann, 1993). Dal momento che il test fornisce talvolta risultati falsamente positivi, si suppone che la proporzione degli uomini realmente infetti sia inferiore a 1,5 su 10.000. Questa stima è approssimativa in diversi sensi; per esempio, non differenzia fra gruppi di età diverse e presuppone che gli uomini con comportamenti a basso rischio siano comparabili a uomini alla prima donazione. Un secondo modo per stimare la prevalenza ignota fa riferimento alla proporzione di individui HIV-positivi che riportano di aver contratto l'infezione per via eterosessuale. Dietz, Seydel, Schwartländer (1994, p. 1998) hanno osservato che il 3,8% dei sieropositivi dichiarava di aver contratto l'infezione per via eterosessuale e non attraverso contatti omosessuali/bisessuali, consumo di droghe per via endovenosa o altre fonti di rischio (per cifre simili, si vedano Glück *et al.*, 1990; Hoffman-Valentin, 1991; Schering, 1992). Nel 1994, quando il presente studio è iniziato, il numero di sieropositivi in Germania era di circa 65.000, dei quali il 29% circa era compreso fra 20 e 30 anni. Se si assume che la cifra del 3,8% valga anche per questo gruppo, si ottiene una stima di 700 sieropositivi in questa fascia di età che riportano di aver contratto l'infezione per via eterosessuale. Poiché nel 1994, in base alle stime disponibili, vi erano in Germania 6.000.000 di uomini fra 20 e 30 anni che non appartenevano a nessun gruppo a rischio noto (un totale di 6.718.500 uomini meno una stima dell'11% appartenenti a uno o più gruppi a rischio noti; si veda Statistiches Bundesamt, 1994), la proporzione di sieropositivi che riportano di aver contratto l'infezione per via eterosessuale può essere stimata intorno a 1,2 su 10.000. Con entrambi i procedimenti descritti, si ottengono cifre fra loro compatibili. Esse dovrebbero comunque essere considerate stime approssimative. Dal momento che non tutti questi individui risultati HIV-positivi hanno realmente il virus (a causa di esiti falsamente positivi), è necessario correggere verso il basso queste cifre. Una prevalenza di 1 su 10.000 sembra quindi essere una stima ragionevole della prevalenza ignota dell'HIV in uomini tedeschi eterosessuali senza comportamenti a rischio fra 20 e 30 anni.

almeno un altro test ELISA (preferibilmente di un diverso produttore). Se il risultato è nuovamente positivo, allora si effettua il test Western Blot, che è più costoso e richiede più tempo. Se il Western Blot è anch'esso positivo, si informa l'interessato che egli è HIV-positivo, e talvolta si ripete il test su un secondo campione sanguigno.

Possono quindi verificarsi due tipi di errori. Primo: un individuo infetto può essere informato che egli è HIV-negativo. La probabilità che si verifichi questo errore (falso negativo) è complementare alla sensibilità del test ELISA. Le stime della sensibilità dell'ELISA solitamente variano fra il 98% e il 99,8% (Eberle, Deinhardt, Habermehl, 1988; George, Schochetman, 1994; Schwartz *et al.*, 1990; Spielberg *et al.*, 1989; Tu *et al.*, 1992; Wilber, 1991). Secondo: un individuo non infetto può essere informato che egli è HIV-positivo. La probabilità di questo secondo errore (falso positivo) è complementare alla specificità combinata del test ELISA e del Western Blot. Sebbene vi sia un accordo generale sul fatto che possano verificarsi falsi positivi, le stime quantitative della loro frequenza variano notevolmente[3]. Ciò è dovuto in parte al fatto che i criteri per la positività di un test Western Blot non sono standardizzati (diverse istituzioni impiegano reagenti, metodi e criteri interpretativi differenti; si veda Stine, 1996, p. 335), che i test ELISA e Western Blot non sono indipendenti (vale a dire che non si può calcolare il tasso combinato di false positività semplicemente moltiplicando fra di loro i singoli tassi di false positività dei due test; si veda Spielberg *et al.*, 1989) e che la specificità del test in una popolazione tende a decrescere al crescere della prevalenza dell'infezione in quella popolazione (Wittkowski, 1989). Per esempio, nel 1990, 20 campioni sanguigni – metà con anticorpi contro l'HIV e metà senza – vennero inviati a ciascuno dei 103 laboratori in sei diverse aree geografiche della Organizzazione Mondiale della Sanità (i laboratori non venivano informati su quali campioni contenessero o non contenessero gli anticorpi) (Snell, Supran, Tamashiro, 1992). Vennero applicate circa 70 diverse combinazioni di test. Fra i campioni senza anticorpi contro l'HIV, l'1,3% venne erroneamente classificato come positivo. Una specificità combinata del 98,7%, come quella osservata in questo studio, è comunque una stima insolitamente bassa. La maggior parte delle stime offerte in letteratura sono notevolmente superiori, di solito maggiori del 99,9% (Burke *et al.*, 1988; Eberle *et al.*, 1988; Peichi-Hoffman,

3. Fra le possibili cause di risultati falsamente positivi rientrano: presenza di anticorpi che danno luogo a reazioni incrociate (Stine, 1996); anticorpi IgM che reagiscono in modo non specifico (Epstein, 1994, p. 56); errori nell'assegnazione dei campioni sanguigni; contaminazione di campioni negativi da parte di campioni positivi adiacenti. Il siero sanguigno può inoltre produrre false positività se trattato termicamente, lipemico o emolizzato; si è riportata un'incidenza del 19% di falsi positivi in pazienti emofiliaci e del 13% in alcolisti affetti da epatite (George, Schochetman, 1994, p. 69). Possono infine risultare falsamente positivi individui affetti da patologie epatiche, individui che hanno ricevuto una trasfusione sanguigna oppure globulina gamma fino a 6 settimane prima dell'esecuzione del test o che hanno ricevuto vaccini antiinfluenzali o contro l'epatite B (Stine, 1996, p. 333).

1991; Tu, Litvak, Pagano, 1992). Per esempio, la Croce Rossa tedesca ha raggiunto una specificità combinata di 99,98% per donatori di sangue al primo prelievo (Wittkowski, 1989). Sulla base delle cifre pubblicate, una stima ragionevole della specificità combinata sembra aggirarsi intorno al 99,99%. Ciò significa che il tasso di falsi positivi è di circa 1 su 10.000. Si tratta di una stima. In future ricerche si potrà forse disporre di cifre più accurate.

1.3. *Valore predittivo positivo*

Ciò che un individuo che si sottopone al test dell'HIV ha bisogno di conoscere è la probabilità che egli sia infetto nel caso che il risultato dell'esame sia positivo. Il valore predittivo positivo (VPP) può essere calcolato sulla base della prevalenza $p(\text{HIV})$, della sensibilità $p(\text{pos}|\text{HIV})$ e del tasso di false positività $p(\text{pos}|\text{no HIV})$:

$$\text{VPP} = \frac{p(\text{pos}|\text{HIV}) \times p(\text{HIV})}{[p(\text{pos}|\text{HIV}) \times p(\text{HIV})] + [p(\text{pos}|\text{no HIV}) \times p(\text{no HIV})]} \qquad (1)$$

dove $p(\text{no HIV})$ equivale a $1 - p(\text{HIV})$. L'equazione (1) è nota come teorema di Bayes. Il teorema esprime il fatto importante che, minore è la prevalenza, minore è la probabilità che l'individuo controllato sia infetto se il test risulta positivo. Qual è il valore predittivo positivo per un tedesco eterosessuale senza comportamenti a rischio fra i 20 e i 30 anni? Introducendo nell'equazione (1) le stime precedentemente indicate – una prevalenza dello 0,01%, una sensibilità del 99,8% e una specificità del 99,99% (con la ripetizione del test ELISA e l'esecuzione di un ulteriore test Western Blot) – si ottiene un VPP di 0,50, o del 50%.

Un VPP stimato del 50% circa per uomini eterosessuali senza comportamenti a rischio è compatibile con le conclusioni della Commissione d'indagine del Parlamento tedesco, che ha stimato il VPP per individui a basso rischio (non solo uomini) come persino «inferiore al 50%» (Deutcher Bundestag, 1990, p. 121).

1.4. *Come comunicare il valore predittivo positivo*

Anche se i consulenti comprendono la formula (1), le persone comuni raramente lo fanno (Gigerenzer, Hoffrage, 1995). Inoltre, studi basati sulla compilazione di questionari e condotti negli Stati Uniti e in Germania indicano che anche i medici esperti incorrono in serie difficoltà quando devono inferire il VPP da informazioni di tipo probabilistico (Casscells, Schoenberger, Grayboys, 1978; Hastie, Dawes, 2001; Hoffrage, Gigerenzer, 1998; Windeler, Köbberling, 1986; cap. 1 in questo volume). Un recente studio condotto a Monaco con 48 medici ha però indicato che le prestazioni dei medici possono migliorare in modo sostanziale, di un fattore superiore a quattro, se le informazioni rilevanti

vengono presentate in termini di *frequenze naturali* anziché come probabilità o percentuali (Gigerenzer, 1996; Hoffrage, Gigerenzer, 1998). Con «frequenze naturali» intendiamo una rappresentazione delle informazioni in termini di frequenze assolute (e non relative), vale a dire la rappresentazione cui un medico sarebbe giunto se egli stesso avesse effettivamente osservato le frequenze esaminando un campione di casi individuali (Gigerenzer, Hoffrage, 1995). Più precisamente, le frequenze naturali sono frequenze che non sono state normalizzate rispetto al tasso di diffusione di base (prevalenza) di una certa condizione. Le frequenze normalizzate, come probabilità e percentuali, sono emerse solo pochi secoli fa come strumenti per rappresentare i gradi di incertezza (Gigerenzer *et al.*, 1989), mentre per gran parte della storia e dell'evoluzione umana la mente ha avuto a che fare soltanto con frequenze naturali.

Come potrebbe un consulente comunicare le informazioni in termini di frequenze naturali? Egli potrebbe illustrare a un paziente il significato di un test positivo nel modo seguente: «Immagini che vengano controllati 10.000 uomini eterosessuali come lei. Uno di loro ha il virus e risulterà positivo al test praticamente con certezza. Anche fra gli uomini rimanenti uno risulterà positivo [in virtù del tasso di false positività dello 0,01%]. Ci aspettiamo quindi che due uomini risultino positivi, e uno solo di loro ha l'HIV. Questa è la situazione nel caso che il suo test risulti positivo; la probabilità di essere infetto è di una su due, o del 50%».

Questo semplice metodo può essere applicato quali che siano i valori numerici rilevanti. Se la prevalenza fosse di 2 su 10.000, il VPP sarebbe di 2 su 3, o del 67%. Se necessario, le cifre possono essere leggermente arrotondate; il punto essenziale è che le informazioni comunicate dal consulente in termini di frequenze naturali anziché in termini probabilistici sono più facilmente comprensibili per coloro che si sottopongono al test. Con una rappresentazione frequentistica è possibile «vedere» in che modo il VPP dipende dalla prevalenza. Se la prevalenza dell'HIV fra gli eterosessuali tedeschi fosse dell'1,5% circa, il consulente potrebbe spiegare: «Pensi a 10.000 uomini eterosessuali come lei. Circa 150 hanno il virus e risulteranno molto probabilmente positivi. Anche fra gli uomini rimanenti uno risulterà positivo. Ci aspettiamo quindi che 151 uomini risulteranno positivi, e 150 di loro hanno l'HIV. Questa è la situazione nel caso che il suo test risulti positivo; la probabilità di essere infetto è di 150 su 151, o del 99,3%».

In generale, il VPP corrisponde al numero di veri positivi (VP) diviso per il numero di veri positivi e di falsi positivi (FP):

$$\mathrm{VPP} = \frac{\mathrm{VP}}{\mathrm{VP} + \mathrm{FP}} \qquad (2)$$

Il confronto fra l'equazione (1) e l'equazione (2) mostra come le frequenze naturali rendano i calcoli mentali più semplici. Le frequenze sono monitorate

in modo accurato e pressocché automatico dagli esseri umani e dagli animali (si vedano, per esempio, Cosmides, Tooby, 1996, e Jonides, Jones, 1992). L'insieme delle ricerche citate indica che sia i medici sia le persone comuni possono comprendere meglio i rischi quando le informazioni rilevanti vengono comunicate come frequenze naturali piuttosto che come probabilità o percentuali.

1.5. *Le consulenze pubbliche sull'HIV*

Circa 300 ambulatori pubblici della Germania (*Gesundheitämter*) offrono gratuitamente al pubblico consulenze sull'AIDS e l'esecuzione del test dell'HIV. Dal 1990 queste istituzioni impiegano 315 consulenti, il 43% dei quali sono medici, il 22% operatori sociali e il 7% psicologi. La formazione professionale dei restanti consulenti è di vario tipo (Fischer, 1990). Come accade in altri paesi, è previsto un incontro di consulenza prima dell'esecuzione del test per far sì che l'utente comprenda la procedura dell'esame, i rischi relativi all'infezione da HIV e il significato di un risultato positivo o negativo (Ward, 1994). Le indicazioni della Commissione d'indagine del Parlamento tedesco prevedono esplicitamente che i consulenti svolgano «una valutazione quantitativa e qualitativa del rischio individuale» e «spieghino qual è l'affidabilità del risultato del test» prima che esso venga eseguito (Deutcher Bundestag, 1990, p. 122). Se un individuo decide di sottoporsi al test, in tutti gli stati tedeschi gli viene garantito l'anonimato (diversamente da quanto avviene negli Stati Uniti, dove in 25 stati si registrano i dati anagrafici; si veda Stine, 1996, p. 346). La consulenza richiede sia un'adeguata sensibilità sociale sia la conoscenza dei margini di incertezza coinvolti nell'esecuzione dell'esame. Il fatto che nel 1990 il 37% degli utenti controllati negli Stati Uniti in cliniche finanziate pubblicamente non si siano presentati per il ritiro dei risultati dei loro esami suggerisce che non sempre le consulenze adempiono al loro scopo (Doll, Kennedy, 1994).

Quali sono le informazioni relative al significato degli esiti del test che i consulenti degli ambulatori pubblici tedeschi forniscono a individui senza comportamenti a rischio? In che modo tali informazioni vengono comunicate (per esempio, in un formato probabilistico o in termini di frequenze naturali)?

2. Lo studio

2.1. *I centri di consulenza*

Uno degli autori ha visitato come utente 20 centri di consulenza sull'AIDS in 20 città tedesche, comprese grandi città come Berlino, Amburgo e Monaco. Due ulteriori centri sono stati visitati in uno studio-pilota condotto per definire i dettagli relativi allo svolgimento del colloquio. I 20 centri di consulenza erano distribuiti in nove stati della ex Germania Occidentale. Dei 20 consulenti, 14 erano medici e 6 erano operatori sociali; 12 erano donne e 8 uomini.

L'utente contattava telefonicamente i centri e prendeva un appuntamento. Poteva visitare due diversi centri di seguito, con un successivo intervallo di almeno due settimane per permettere agli ematomi prodotti dalle iniezioni al braccio di guarire. Questi intervalli erano necessari, altrimenti i segni sulle braccia avrebbero potuto suggerire ai consulenti che l'utente fosse tossicodipendente.

Lo svolgimento di un'indagine sul lavoro dei consulenti sull'AIDS della quale essi non siano al corrente solleva problemi di carattere etico. Abbiamo consultato il Comitato etico dell'Associazione nazionale tedesca di psicologia, che ci ha informato che in questo caso il beneficio atteso dai risultati dello studio potevano giustificare il necessario inganno nei confronti dei consulenti. Le consulenze pubbliche rappresentano comportamenti pubblici, e non privati; ciononostante, in considerazione dell'interpretazione della normativa sulla *privacy* da parte del Comitato etico, abbiamo deciso di non filmare le sessioni di consulenza. Proteggiamo inoltre l'anonimato dei consulenti. Ci scusiamo con tutti i consulenti per aver nascosto loro lo svolgimento della ricerca e i suoi metodi, ma riteniamo che i risultati della nostra indagine giustifichino questo approccio, rivelando in che modo le consulenze sull'AIDS potranno in futuro essere migliorate.

2.2. *L'intervista*

L'utente presentava al consulente le seguenti domande, nell'ordine indicato (a meno che il consulente non fornisse autonomamente qualcuna delle informazioni corrispondenti).

1. *Sensibilità del test dell'HIV*. Se si è infetti, è possibile avere un risultato negativo al test? Qual è l'affidabilità del test nell'identificare il virus se esso è effettivamente presente?
2. *Specificità del test dell'HIV*. Se non si è infetti, è possibile avere un risultato positivo al test? Qual è l'affidabilità del test rispetto a risultati falsamente positivi?
3. *Prevalenza dell'HIV fra uomini eterosessuali*. Quanto è frequente l'infezione da HIV nella popolazione a cui appartengo, cioè fra gli uomini eterosessuali da 20 a 30 anni senza fattori di rischio noti (come l'uso di droghe)?
4. *Valore predittivo di un test positivo*. Qual è la probabilità che gli uomini che appartengono alla mia stessa popolazione abbiano effettivamente l'HIV dopo che hanno ottenuto un risultato positivo al test?
5. «*Periodo-finestra*». Quanto tempo deve passare fra l'infezione e l'esecuzione del test perché gli anticorpi possano essere rilevati?

Lo studio pilota aveva indicato una tendenza dei consulenti a fornire risposte vaghe e non informative, come: «Non si preoccupi, il test è molto affidabile. Si fidi». Aveva anche indicato che se l'utente chiedeva chiarimenti per più di due volte di seguito, era probabile che i consulenti diventassero nervosi e irritati, percependo l'insistenza dell'utente come una violazione delle norme sociali di comunicazione. Sulla base di queste sessioni pilota, nelle interviste per il pre-

sente studio abbiamo stabilito il seguente schema riguardo alle domande di chiarimento. Se il consulente forniva una stima quantitativa (un numero o un intervallo di numeri) o se il consulente dichiarava di non sapere (o di non volere) fornire una risposta più precisa, allora l'utente passava alla domanda successiva. Se la risposta era di tipo qualitativo (es., «abbastanza sicuro») o se il consulente aveva frainteso la domanda o aveva evitato di rispondere, allora l'utente chiedeva ulteriori chiarimenti e, se necessario, ripeteva in seguito la sua richiesta di chiarimento ancora una volta. Se anche il terzo tentativo di ottenere una risposta più precisa falliva, allora l'utente non insisteva oltre e passava alla domanda successiva. Quando l'utente doveva chiedere chiarimenti riguardo alla prevalenza dell'HIV (domanda 3), ripeteva sempre il suo specifico gruppo di appartenenza (uomini eterosessuali fra i 20 e i 30 anni senza fattori di rischio noti); analogamente, quando doveva chiedere chiarimenti riguardo al VPP (domanda 4), faceva sempre riferimento al suo specifico gruppo di appartenenza.

Come si è detto, le informazioni che l'utente forniva spontaneamente ai consulenti erano: niente comportamenti omosessuali, niente consumo di droghe e nessun altro comportamento a rischio noto. Se i consulenti chiedevano ulteriori informazioni – il che è accaduto solo in undici delle venti sessioni – l'utente spiegava di avere 27 anni, di essere monogamo e che né le sue (poche) partner precedenti né quella del momento facevano uso di droghe o praticavano altri comportamenti a rischio. In due di questi undici casi, all'utente è stato presentato un questionario dettagliato per determinare il suo livello di rischio; in uno di questi due casi, il consulente non ha guardato il questionario e l'utente, andandosene, lo aveva ancora con sé.

L'utente è stato addestrato in alcune sessioni simulate a utilizzare un sistema di codifica stenografico (numero della domanda, numero di ripetizioni della domanda, risposta del consulente; es.: «1; 2; 99,9%») che permetteva di prendere nota rapidamente delle informazioni rilevanti durante la sessione di consulenza o, se la sessione era molto breve, di tenerle a mente nel formato stabilito e trascriverle subito dopo il colloquio.

Dopo le sessioni di consulenza, l'utente si è sottoposto al test dell'HIV, tranne che in tre casi (in due casi avrebbe dovuto attendere svariate ore prima di fare l'esame e in un caso il consulente aveva suggerito che l'utente meditasse sulla decisione di fare l'esame lasciando passare una giornata).

3. Risultati

A scopo illustrativo, quattro sessioni di consulenza sono riportate integralmente in tab. 2. Le domande dell'utente sono state abbreviate (es.: «sensibilità?») e le informazioni fornite dal consulente seguono immediatamente le rispettive domande. Le risposte dei consulenti alle richieste di chiarimento dell'utente (successive alla prima formulazione di ciascuna domanda) compaiono di seguito in elenchi puntati.

3.1. *Sensibilità e periodo-finestra*

Diciannove dei venti consulenti hanno fornito all'utente informazioni relative alla sensibilità. (Il ventesimo consulente si è rifiutato di fornire qualsiasi informazione riguardo a sensibilità, specificità e valore predittivo prima dell'esecuzione del test. Ma non li ha forniti neanche quando in seguito l'utente ha ritirato i risultati.) La maggior parte dei consulenti ha fornito informazioni realistiche riguardo alla sensibilità (tab. 1). Tuttavia, 5 dei 19 consulenti hanno erroneamente riferito all'utente che un risultato falsamente negativo era impossibile se non durante il periodo-finestra. Quindici consulenti hanno fornito informazioni relative al periodo-finestra nel momento in cui erano interrogati a proposito della sensibilità. La stima mediana per il periodo-finestra è stata di 12 settimane.

*Tab. 1 - Le informazioni fornite dai consulenti**

	Certezza del 100%	*> 99,9%*	*> 99%*	*> 90%*	*Intervallo*
Sensibilità	5 (su 19)	5	6	3	90-100%
Specificità	13 (su 19)	3	3	0	99,7-100%
Prevalenza	–	–	–	–	0,0075-6%
VPP	10 (su 18)	5	1	2	90-100%

* Non tutti i consulenti hanno fornito stime numeriche. L'espressione verbale «assolutamente certo» è stata qui considerata equivalente a 100%; espressioni verbali come «con certezza quasi assoluta» e «molto, molto sicuro» sono state classificate come > 99% ed espressioni come «molto affidabile» sono state classificate come > 90%.

3.2. *Falsi positivi*

Tredici dei diciannove consulenti hanno erroneamente riferito all'utente che non si danno risultati falsamente positivi (si veda, per esempio, la Sessione 1 in tab. 2). Undici di loro hanno motivato la risposta spiegando che la ripetizione del controllo, prima con il test ELISA e quindi con il Western Blot, elimina del tutto la possibilità di false positività. Cinque dei tredici consulenti hanno riferito all'utente che risultati falsamente positivi erano stati registrati negli anni Ottanta, ma che a oggi non si presentano più, e due consulenti hanno dichiarato che potrebbero darsi risultati falsamente positivi in altri paesi, come in Francia, ma non in Germania. Oltre a questi tredici consulenti, altri cinque hanno suggerito in un primo momento che non si danno falsi positivi, ma si sono mostrati meno sicuri di fronte alla ripetizione della domanda da parte dell'utente, ammettendo questa possibilità (si vedano le Sessioni 2 e 3). Solo i tre consulenti rimanenti hanno prontamente informato l'utente riguardo alla possibilità di risultati falsamente positivi. Uno soltanto (Sessione 4) ha fornito all'utente l'importante informazione che la proporzione dei falsi positivi in confronto ai veri positivi era particolarmente alta in un gruppo come quello cui egli apparteneva.

Tab. 2 - Quattro esempi di sessioni di consulenza. Dopo ogni domanda è riportata la risposta del consulente. Nelle righe successive compaiono le risposte del consulente alle successive richieste di chiarimento in relazione alla stessa domanda

---

Sessione 1. Il consulente è un operatore sociale (donna)

Sensibilità? In realtà, non si danno mai falsi positivi. Anche se, pensando alla letteratura scientifica, alcuni casi del genere sono stati riportati.
- Non so esattamente quanti siano.
- È successo solo una o due volte.

False positività? Non ce ne sono, perché il test viene ripetuto. È assolutamente sicuro.
- Se ci sono gli anticorpi, il test li rileva senza ambiguità e con assoluta certezza.
- No, è assolutamente impossibile che si diano falsi positivi. Grazie alla ripetizione del test, il risultato è assolutamente certo.

Prevalenza? Non glielo so dire con esattezza.
- All'incirca fra 1 su 500 e 1 su 1.000.

Valore predittivo positivo? Come le ho già detto ripetutamente, il test è assolutamente certo.

Sessione 2. Il consulente è un medico (uomo)

Sensibilità? Quando ci sono abbastanza anticorpi il test li identifica in tutti i casi. Per il test si svolgono due esami; il primo è alla quarta generazione ed è calibrato in modo da essere molto specifico e molto sensibile. Identifica comunque più facilmente i casi positivi che quelli negativi.
- Sensibilità e specificità del 99,8%; ma noi ripetiamo il test, e quando risulta positivo è solido come acciaio.

False positività? È sicuro, non ce ne sono; gli eventuali errori sono tutte false negatività, che si presentano quando gli anticorpi non si sono ancora formati.
- Se lei fa l'esame qui, compreso un test di conferma, la certezza è massima. In ogni caso, la specificità è del 99,7%. È una cosa solida come l'acciaio; eliminiamo ogni possibile confusione usando due test.

Prevalenza? Oggi la classificazione delle persone in gruppi a rischio è superata; non si può guardare al problema in questo modo.
- Non ricordo. C'è una tendenza del virus a diffondersi in tutta la popolazione. Le statistiche sono inutili per i casi individuali!

Valore predittivo positivo? Estremamente certo, l'ho già detto: 99,8%.

Sessione 3. Il consulente è un medico (donna)

Sensibilità? Il test è molto, molto affidabile: 99,98%.

False positività? Il test sarà ripetuto. Dopo il primo controllo non si parla di positività ma di «reattività». Quando sono stati effettuati tutti i controlli, allora il risultato è sicuro.

Tab. 2 - Segue

- È difficile dire quanti siano i falsi positivi.
- Quanti precisamente? Dovrei guardare la letteratura per vedere se trovo questa informazione.

Prevalenza? Dipende dalla regione.
- Fra le 67.000 persone infette in Germania, il 9% sono eterosessuali.
- A Monaco abbiamo 10.000 persone infette, cioè l'1% della popolazione. Ma questi sono solo numeri; sul fatto che lei abbia o no il virus non ci dicono nulla.

Valore predittivo positivo? Come le ho già indicato, il risultato è sicuro al 99,98%. Se dovesse avere un risultato positivo, può fidarsi.

Sessione 4. Il consulente è un operatore sociale (donna).

Sensibilità? Molto, molto affidabile.
- No, non è completamente sicuro, niente lo è in medicina, perché è possibile che il virus non venga identificato.
- Vicino al 100%. Non lo so con esattezza.

False positività? Esistono, ma sono estremamente rari.
- Nell'ordine di un decimo di punto percentuale. Probabilmente meno. Comunque, nella popolazione alla quale lei appartiene i falsi positivi sono relativamente più frequenti che in un gruppo ad alto rischio.
- Non so la cifra esatta.

Prevalenza? Con i contatti che lei ha avuto, è improbabile che abbia contratto l'infezione.
- In generale non si può dire. Nel nostro centro, su circa 10.000 test eseguiti in sette anni, ci sono state solo tre o quattro persone eterosessuali non tossicodipendenti, o comunque appartenenti a gruppi non a rischio, che hanno avuto un risultato positivo.

Valore predittivo positivo? Come dicevo, il test non è sicuro al 100%. Se il test confonde gli anticorpi per l'HIV con altri anticorpi, allora la ripetizione del test non è di aiuto. E nel caso di persone come lei, che non hanno reali fattori di rischio, immagino che anche il 5% o il 10% di coloro che risultano positivi avranno ottenuto un esito falsamente positivo.

3.3. *Prevalenza*

La domanda relativa alla prevalenza dell'HIV fra uomini eterosessuali con comportamenti a basso rischio ha prodotto la maggiore incertezza fra i consulenti. Sedici su 20 (tutti i consulenti hanno risposto a questa domanda) hanno espresso incertezza o ignoranza, o hanno sostenuto che la prevalenza fra uomini eterosessuali con comportamenti a basso rischio non può essere determinata (per esempio, per via di casi che non vengono riportati), oppure che tale informazione non sarebbe di nessuna utilità in relazione a casi individuali (si veda, per esempio, la Sessione 2). Diversi consulenti hanno cercato informazioni nelle pubblicazioni per rispondere alla domanda, ma hanno trovato solo dati statistici irrilevanti, come l'alto numero di individui HIV-positivi a Berlino Ovest («Il muro è stato il miglior preservativo per Berlino Est», ha osservato un consulente). Dodici consulenti hanno fornito stime numeriche, con una mediana di 0,1%. La variabilità di queste stime è stata considerevole (tab. 1) e comprendeva la stima estrema per cui, in persone come l'utente, l'infezione da HIV è «meno probabile di vincere tre volte alla lotteria» (non abbiamo incluso questa stima in tab. 1). Quattro consulenti hanno asserito che le informazioni relative alla prevalenza sono di scarsa o di nessuna utilità (es.: «Ma le statistiche non ci aiutano in un caso individuale – e non abbiamo neppure dati precisi»; si vedano anche le Sessioni 2 e 3). Due consulenti hanno dichiarato di avere problemi a ricordare le cifre o a ragionare con i numeri, per esempio: «Ho delle difficoltà a ragionare con le informazioni statistiche. Riguardano i gruppi e trasferirle ai casi singoli è problematico. Mi vengono in mente le lotterie: la probabilità che sei numeri coincidano tutti con la sequenza vincente è molto bassa; però ogni settimana qualcuno vince».

3.4. *Valore predittivo positivo*

Si ricordi che, sulla base delle stime attualmente disponibili, solo il 50% circa degli uomini eterosessuali con comportamenti a basso rischio hanno effettivamente l'infezione da HIV nel caso che siano risultati positivi al test. I consulenti hanno fornito in proposito informazioni decisamente diverse. Metà dei consulenti (10 su 18; due hanno ripetutamente ignorato la domanda) hanno detto all'utente che se l'esame fosse risultato positivo sarebbe stato assolutamente certo (100%) che egli fosse infetto (si vedano tab. 1 e Sessione 1). Cinque consulenti hanno dichiarato che la relativa probabilità sarebbe stata almeno del 99,9% (si veda per esempio la Sessione 3). Quindi, se l'utente avesse ottenuto un risultato positivo e avesse confidato nelle informazioni fornite da questi 15 consulenti, avrebbe potuto effettivamente contemplare la

possibilità del suicidio, come molti altri prima di lui avevano fatto (Stine, 1996).

In che modo i consulenti sono giunti a questa marcata sovrastima del potere predittivo? Sembra che essi abbiano seguito due percorsi di ragionamento. Un totale di otto consulenti ha confuso sensibilità e potere predittivo positivo (una confusione già documentata in letteratura; si vedano Elstein, 1988, e il cap. 1 in questo volume), cioè hanno fornito la stessa cifra per la sensibilità e per il VPP (si vedano per esempio le Sessioni 2 e 3). Tre fra questi otto consulenti hanno spiegato che, con l'eccezione del periodo-finestra, la sensibilità è del 100% e hanno quindi indicato che anche il VPP era del 100%. Altri cinque consulenti hanno ragionato applicando una seconda strategia: hanno assunto (erroneamente) che la ripetizione di diversi test eliminasse del tutto i falsi positivi e ne hanno concluso (correttamente) che il VPP è del 100%. Per entrambi questi gruppi di consulenti, la domanda dell'utente riguardo al VPP deve essere apparsa come una domanda alla quale essi avevano già risposto. In effetti, più della metà dei consulenti (11 su 18) hanno esplicitamente introdotto la loro risposta con formule come «Come ho già detto, …» (si vedano per esempio le Sessioni 1-3). In linea con queste considerazioni, le risposte relative al VPP sono arrivate piuttosto in fretta e l'utente non ha avuto bisogno di chiedere chiarimenti con la stessa frequenza che nelle domande precedenti. La media del numero di domande formulate dall'utente in relazione al VPP è stata soltanto di 1,8, in confronto a 2,4, 2,4 e 2,5 rispettivamente per la sensibilità, la specificità e la prevalenza.

La tab. 1 riporta che due consulenti hanno fornito stime del VPP nella direzione corretta (comprese fra 99% e 90%). Tuttavia, solo una di loro (Sessione 4) è giunta a tale stima ragionando sul fatto che la proporzione dei falsi positivi sul totale dei risultati positivi aumenta al diminuire della prevalenza. Questa consulente è stata anche l'unica a spiegare all'utente che vi sono specifiche ragioni per cui i falsi positivi non possono essere del tutto eliminati con test ripetuti, per esempio la possibilità che i test reagiscano ad anticorpi che vengono erroneamente identificati come anticorpi contro l'HIV. Il secondo di questi due consulenti, invece, ha inizialmente asserito che in base a un risultato positivo l'infezione da HIV è «completamente certa»; ma, quando l'utente ha chiesto che cosa significasse «completamente certa», il medico sembra averci ripensato e ha indicato che il VPP rientra «almeno fra i valori più alti oltre il 90% », dichiarando «non posso essere più preciso».

3.5. *Come viene comunicata l'informazione?*

Come abbiamo già osservato, alcuni studi sperimentali hanno mostrato che le informazioni relative al rischio e all'incertezza sono comunicate e compre-

se più efficacemente se presentata in termini di frequenze naturali che non in termini di probabilità e percentuali. In che modo i consulenti hanno comunicato le informazioni all'utente? Nemmeno uno dei consulenti ha comunicato le informazioni in termini di frequenze naturali, cioè nella rappresentazione che i medici e le persone comuni comprendono meglio. Con l'eccezione della prevalenza dell'HIV, tutte le informazioni numeriche sono state presentate in termini percentuali, come illustrato nelle quattro sessioni in tab. 2. La conseguenza di ciò è che gli utenti probabilmente non comprendono adeguatamente le informazioni ricevute. Del resto, pare che diversi consulenti non comprendano essi stessi le cifre che comunicano. Lo si può inferire dall'osservazione che diversi consulenti hanno fornito all'utente informazioni fra loro incoerenti, ma non sembrano essersene resi conto. Si considerino in proposito i due esempi seguenti.

Un medico ha dichiarato che la prevalenza dell'HIV fra uomini come l'utente era dello 0,1% o leggermente più alta e che la sensibilità, la specificità e il VPP erano ciascuno del 99,9%. Per mostrare che tali informazioni sono incoerenti, le si può rappresentare in termini di frequenze naturali. Immaginate 1.000 uomini che si sottopongono al test per l'HIV. Uno di loro (lo 0,1%) è infetto e risulterà positivo al test praticamente con certezza. Anche uno dei restanti uomini non infetti risulterà positivo al test (assumendo che la specificità sia 99,9%, il tasso di falsi positivi sarà di 0,1%). Quindi due uomini risulteranno positivi al test, e uno solo di loro è infetto. Il che significa che il rapporto di probabilità in favore di un'infezione da HIV è di 1 a 1 e non di 999 a 1 (il 99,9%) (anche se il medico avesse ipotizzato una prevalenza di 0,5%, il rapporto di probabilità sarebbe di 5 a 1 piuttosto che di 999 a 1).

Si considerino ora le informazioni che l'utente ha ricevuto nella Sessione 2. Assumiamo per la prevalenza (che il consulente non ha fornito) la stima mediana indicata dai consulenti, cioè lo 0,1%. Ancora una volta, immaginate 1.000 uomini. Uno di loro ha il virus e risulterà positivo al test praticamente con certezza (sensibilità stimata dal consulente: 99,8%). Anche tre dei restanti uomini non infetti risulteranno positivi al test (specificità stimata dal consulente: 99,7%). Ci aspettiamo quindi che quattro uomini risultino positivi al test, e che uno di loro abbia effettivamente il virus. Ciò significa che la probabilità di essere infetti se il test è positivo equivale al 25% (1 su 4), e non del 99,8%, come il consulente ha indicato all'utente.

Se i consulenti fossero stati formati a rappresentare le informazioni in termini di frequenze naturali, queste incoerenze avrebbero potuto essere facilmente rilevate. Ma, a quanto pare, i consulenti non avevano ricevuto nessuno specifico addestramento sul modo più opportuno per rappresentare e comunicare le informazioni sul rischio. La tab. 3 illustra una ipotetica sessione in cui il consulente utilizza le frequenze naturali.

Tab. 3 - Un consulente «ideale» che comunica le informazioni rilevanti in termini di frequenze naturali. Dal momento che non abbiamo trovato nessun consulente che l'abbia fatto, questa sessione è fittizia

---

Sensibilità? L'esame risulterà positivo per circa 998 persone infette dall'HIV su 1.000. La stima può variare leggermente a seconda delle circostanze, per esempio del particolare tipo di test usato.

False positività? Circa una su 10.000. Possono essere fortemente ridotte, ma non completamente eliminate, con test ripetuti. Fra le possibili cause di false positività ci sono…

Prevalenza? Fra i maschi eterosessuali con comportamenti a basso rischio, è infettato dall'HIV circa uno su 10.000.

Valore predittivo positivo? Pensi a 10.000 uomini a basso rischio come lei. Uno è infetto, ed è praticamente certo che risulterà positivo; ma risulterà positivo anche uno dei 9.999 non infetti. Perciò ci aspettiamo che due uomini risultino positivi e che uno solo fra loro abbia l'HIV. La situazione in cui lei si troverebbe in caso di un esito positivo è questa; la sua probabilità di avere il virus sarebbe grosso modo del 50%.

Gli opuscoli disponibili negli ambulatori – una fonte di informazioni alla quale i consulenti potrebbero attingere – aiutano a comprendere quale sia il significato di un test positivo quando la prevalenza è bassa? Abbiamo studiato 78 diversi opuscoli provenienti dai 20 centri visitati (alcuni erano stati forniti all'utente direttamente dai consulenti), che variavano da pubblicazioni del governo federale fino a documenti prodotti dagli stessi centri locali. Questi opuscoli contenevano una gran quantità di utili elementi di informazione, per esempio relativi alle vie attraverso le quali il virus può o non può essere trasmesso, ma pochissime indicazioni riguardo al significato di un risultato positivo quando la prevalenza è bassa. In particolare, non c'era alcuna informazione relativa alla prevalenza fra uomini e donne senza comportamenti a rischio. Il documento più diffuso era una serie di 10 fascicoli pubblicati dal Centro federale per l'educazione sanitaria (Bundeszentrale für gesundheitliche Aufklärung, 1988-1993). Nel primo fascicolo viene menzionato il problema dei risultati falsamente positivi e in proposito si osserva che il procedimento attraverso test ripetuti è «affidabile». Nel secondo fascicolo il tema è di nuovo menzionato brevemente e compare la ragionevole raccomandazione che persone senza fattori di rischio noti che ciononostante risultino positive al test dovrebbero sottoporsi a un secondo controllo, per esempio in un centro pubblico di consulenza. (Ci si potrebbe chiedere che cosa si sentirebbe dire un persona in questa situazione da quei consulenti che ritengono che un risultato positivo sia assolutamente certo.) Il terzo fascicolo annuncia «nel prossimo futuro» test per gli anticorpi in grado di identificare con certezza le infezioni da HIV-1 e HIV-2 – in contraddizione con le conclusioni della Commissione di indagine del parlamento tedesco secondo le quali non vi sarà mai la

certezza assoluta nella rilevazione dell'HIV, come del resto accade per tutte le altre infezioni virali (Deutscher Bundestag, 1988, p. 79). Neanche i restanti fascicoli forniscono stime quantitative dell'incertezza implicata nell'esecuzione del test. Vengono presentate stime quantitative solo in relazione al periodo-finestra, che è irrilevante riguardo al numero di falsi positivi. In nessun caso si tenta di spiegare al lettore in modo comprensibile la relazione fra prevalenza, sensibilità, false positività e valore predittivo positivo. Sulla base di questi documenti, quindi, né i consulenti né gli utenti possono imparare che cosa significhi un risultato positivo del test quando la prevalenza è bassa.

4. Conclusioni

Questo studio mostra, relativamente a un campione di centri pubblici tedeschi per le consulenze sull'AIDS, che i consulenti non erano preparati a spiegare a un uomo con comportamenti a basso rischio quale sarebbe stato il significato di un risultato positivo nel test dell'HIV. Questo non significa affatto che i consulenti fossero in generale ignoranti in materia; al contrario, diversi fra loro hanno fornito lunghe e sofisticate spiegazioni riguardo ad argomenti quali le tecniche immunodiagnostiche, la natura delle proteine e i percorsi dell'infezione. Ma di fronte al compito di spiegare all'utente il rischio di essere infetto alla luce di un test positivo si è riscontrata tanto una mancanza di informazioni quanto una mancanza di conoscenza riguardo ai modi più opportuni per comunicare il rischio.

I principali problemi identificati in questa ricerca sono i seguenti.

1. Tutti i consulenti hanno comunicato le informazioni in termini di probabilità e percentuali piuttosto che in un formato che agevola la comprensione negli utenti (e in loro stessi), come la rappresentazione in termini di frequenze naturali.
2. Solo uno dei venti consulenti (Sessione 4) ha spiegato che la proporzione di falsi positivi rispetto al totale dei risultati positivi cresce al diminuire della prevalenza.
3. La maggior parte dei consulenti ha erroneamente assicurato all'utente che non si danno mai falsi positivi. I consulenti hanno fornito al riguardo una spiegazione semplice e deterministica: i falsi positivi verrebbero eliminati attraverso la ripetizione di diversi test (analogamente, i falsi negativi scomparirebbero dopo il periodo-finestra).
4. La metà dei consulenti ha erroneamente asserito che se un individuo a basso rischio risulta positivo al test, è assolutamente certo (100%) che egli ha contratto il virus. I consulenti sono giunti a questo giudizio erroneo seguendo due diversi percorsi. Un primo gruppo ha confuso la sensibilità del test e il suo valore predittivo positivo. Un secondo gruppo ha ritenuto che

non vi fossero falsi positivi in virtù della ripetizione di diversi test, il che implica che un risultato positivo indica la presenza dell'infezione con assoluta certezza.

Non sappiamo quanto questi risultati siano rappresentativi delle consulenze sull'AIDS per utenti a basso rischio in altri centri o in altri paesi. Questo studio sembra essere il primo nel suo genere; ma non abbiamo ragioni per escludere che il campione dei centri visitati sia rappresentativo almeno della Germania (per la precisione, dell'ex Germania Occidentale). La lezione da trarre da questa indagine sembra risiedere nell'importanza di insegnare ai consulenti come spiegare agli utenti i rischi rilevanti in termini semplici. A questo scopo, ciò di cui i consulenti hanno bisogno sono innanzi tutto stime approssimative del tasso di falsi positivi, della sensibilità e della prevalenza dell'HIV in diverse popolazioni. Dopodiché è possibile insegnare loro a comunicare queste informazioni in modo comprensibile. Gli studi sperimentali suggeriscono che il modo più semplice ed efficiente è addestrare i consulenti a rappresentare le informazioni rilevanti in termini di frequenze naturali e a comunicarle all'utente nello stesso modo[4]. Questo tipo di addestramento richiede poco tempo e presenta un bilancio molto vantaggioso nel rapporto fra costi e benefici. Per esempio, un software didattico che insegna a costruire rappresentazioni frequentistiche richiede meno di due ore, compresa una fase di valutazione (Sedlmeier, Gigerenzer, 2001). Ancora più importante è il fatto che la mente umana sembra essere programmata per utilizzare il formato frequentistico, dal momento che chi viene addestrato in questo modo non mostra, nel corso del tempo, il comune decadimento dell'apprendimento.

È chiaro che saper spiegare in forma semplice che cosa significhi un risultato positivo non esaurisce le competenze necessarie per un consulente; ma costituisce una parte importante. La comprensione appropriata di informazioni corrette può prevenire reazioni autodistruttive da parte degli utenti. Tali reazioni comportano un costo in termini di vite umane che potrebbe essere evitato, ma che altrimenti si somma a quello dovuto alla malattia stessa.

Questa indagine è stata sostenuta finanziariamente dai Fonds zur Förderung der Wissenshaften (P8842-Med) e dalla Deutsche Foschungsgemeinshaft (Ho 1847/1-1). Ringraziamo anche Christian Fiebach, Adam Goodie, Andreas Stoltze e Anita Todd per il loro aiuto e i loro utili commenti su stesure precedenti dell'articolo.

4. Esistono anche dati sperimentali che mostrano come la presenza dello specifico errore commesso con più frequenza dai consulenti dell'AIDS – vale a dire la confusione fra sensibilità e valore predittivo positivo – risulta marcatamente ridotta (dal 19% al 5% di tutte le inferenze diagnostiche) quando le informazioni rilevanti vengono rappresentate in termini frequentistici anziché probabilistici (Gigerenzer, Hoffrage, 1997).

13. Giudizio clinico vs. *giudizio statistico**

di *Robyn M. Dawes, David Faust, Paul E. Meehl*

Un paziente psichiatrico manifesta sintomi ambigui. È preferibile trattare la sua condizione con la sola psicoterapia o è opportuno l'impiego di farmaci antipsicotici che occasionalmente danno luogo a pericolosi effetti collaterali? Un uomo anziano lamenta una perdita di memoria, ma le indagini neurologiche e i riscontri diagnostici forniscono indicazioni equivoche. Il neuropsicologo deve somministrare e interpretare una serie di test psicologici che permettano di escludere una patologia cerebrale progressiva. Un controllo clinico conferma i peggiori timori di un paziente: ha un cancro in fase terminale. Il paziente chiede al suo medico quanto tempo gli resta per mettere ordine nella sua vita.

Questi tre brevi scenari illustrano alcune delle molte situazioni in cui gli esperti sono chiamati a diagnosticare o prevedere una condizione o un evento. Una pianificazione ottimale delle cure spesso dipende dall'accuratezza dei giudizi degli esperti che offrono la loro consulenza. Che si tratti di medici, psichiatri o psicologi, questi esperti svolgono due funzioni fondamentali per la decisione: raccolgono informazioni e ne forniscono un'interpretazione. In questo saggio siamo interessati alla seconda di queste due funzioni e in particolare al valore relativo dell'interpretazione dei dati basata sul metodo clinico e su quello statistico.

1. Il confronto fra diversi metodi di giudizio

Nell'impiego del metodo clinico, un individuo combina o processa un insieme di informazioni «a mente». Il metodo statistico, per contro, mette da parte il giudizio umano e produce una determinata conclusione esclusivamen-

* Traduzione di *Clinical* vs. *actuarial judgment*, «Science», 243 (1989), pp. 1668-1673, copyright © 1989 American Association for the Advancement of Science.

te in base ai rapporti, stabiliti su basi empiriche, fra i dati disponibili e l'evento che si intende prevedere. Per esempio, un assicuratore fa uso del metodo clinico se combina i fattori di rischio rilevanti per l'aspettativa di vita attraverso un giudizio personale; usa invece il metodo statistico se inserisce i dati in una formula o in tabelle che contengono informazioni empiriche sul rapporto fra tali dati e l'aspettativa di vita.

Come procedura per elaborare giudizi, il metodo clinico non è confinato alla pratica clinica. Un individuo che affronta problemi clinici in ambito medico o psichiatrico può far uso tanto del metodo clinico quanto di quello statistico. Analogamente, il metodo statistico non dovrebbe essere semplicemente identificato con l'impiego di regole automatiche di decisione. Anche il giudizio clinico, infatti, può essere riprodotto con strumenti automatizzati. È possibile, per esempio, programmare un computer affinché individui «tratti di dipendenza» come farebbe un esperto di fronte a certi risultati di un test psicologico. Per essere genuinamente statistica, l'interpretazione dei dati deve essere sia automatica (vale a dire ottenuta attraverso procedure predefinite e standardizzate) sia fondata su relazioni empiriche stabilite.

È possibile trattare con strumenti statistici praticamente qualsiasi tipo di dati. Per esempio, le osservazioni che emergono da un colloquio possono essere codificate in termini quantitativi (il paziente appare remissivo: [1] sì, [2] no). Si possono quindi integrare in una procedura statistica sia osservazioni qualitative sia dati quantitativi. Inoltre, è possibile ricavare attraverso il metodo statistico praticamente qualsiasi tipo di diagnosi, descrizione o previsione.

Alcune ricerche su giudizio clinico e giudizio statistico sono apparse già prima della metà degli anni Cinquanta. Ma è stata la pubblicazione del lavoro di Meehl (1954) a portare questo tema all'attenzione di numerosi studiosi interessati al comportamento umano e a stimolare lo sviluppo di un intero settore di studi. In particolare, si deve a Meehl la definizione delle condizioni necessarie per una equilibrata valutazione comparativa dell'efficacia dei due metodi.

Innanzi tutto, i giudizi da confrontare dovrebbero essere stati ricavati attraverso i due diversi metodi a partire dagli stessi dati. Questa condizione non esclude che, in una fase precedente al confronto, le regole o strategie decisionali impiegate nei due approcci possano essere state sviluppate sulla base di insiemi di dati differenti. Solitamente le strategie decisionali coinvolte nel giudizio clinico di un esperto dipendono da precedenti esperienze e conoscenze. Lo sviluppo di metodi statistici, d'altra parte, richiede una serie iniziale di casi nei quali l'esito (per esempio, la diagnosi) sia noto. Le strategie cliniche e quelle statistiche possono quindi essere ricavate da insiemi di dati separati o solo parzialmente sovrapposti ed è anche possibile che il numero di casi utilizzati in un caso e nell'altro sia differente. Per esempio, un esperto potrebbe aver interpretato 1.000 test psicologici per la rilevazione di disfun-

zioni cerebrali e potrebbe essere al corrente della diagnosi definitiva (documentata da successivi esami radiologici) di alcuni di questi casi. Una procedura statistica potrebbe invece essere stata sviluppata sulla base del sottoinsieme di quei 1.000 casi per i quali la diagnosi definitiva è nota.

In secondo luogo, è necessario evitare condizioni che possano incrementare artificialmente l'accuratezza del metodo statistico. Per esempio, in virtù delle procedure matematiche solitamente utilizzate per costruirle (come l'analisi di regressione o l'analisi discriminante), le regole statistiche possono risultare avvantaggiate dall'identificazione di relazioni fra variabili che si presentano in un particolare insieme di casi a causa di fluttuazioni casuali. Per questo motivo, la derivazione di regole statistiche dovrebbe essere seguita da una fase di «validazione incrociata» (*cross-validation*), cioè dall'applicazione a un nuovo insieme di casi, oppure da una stima statistica standard del probabile esito dell'applicazione a un nuovo insieme di casi. La validazione incrociata evita una valutazione troppo benevola dell'accuratezza e rende possibile una misurazione realistica delle prestazioni del metodo esaminato. Si tratta di un momento essenziale dell'indagine, perché la validità di una procedura di previsione dovrebbe essere controllata là dove serve, cioè in casi nei quali l'esito finale è inizialmente ignoto. Se la procedura in questione è destinata a un impiego limitato nel contesto in cui è stata sviluppata, è possibile suddividere un campione rappresentativo di casi in due insiemi e utilizzarne uno per la costruzione della regola statistica e uno per la validazione incrociata della stessa. Se il campo di applicazione inteso è più ampio, sarà necessario ricorrere a un insieme di nuovi casi rappresentativi degli ambiti e delle popolazioni di interesse.

2. I risultati degli studi comparativi

I tre scenari delineati all'inizio di questo saggio esemplificano tre importanti studi comparativi. In particolare, Goldberg (1965, 1968) ha analizzato la distinzione fra neurosi e psicosi sulla base del Minnesota Multiphasic Personality Inventory (MMPI), un test psicologico comunemente utilizzato per questo compito diagnostico. Si tratta di un giudizio diagnostico che può avere importanti conseguenze di carattere pratico (per esempio, una diagnosi di psicosi può implicare un trattamento medico rischioso o l'impossibilità di sottoscrivere alcune polizze assicurative). Goldberg ha derivato diverse regole di decisione analizzando statisticamente i punteggi ottenuti da una serie di pazienti in 11 prove comprese nel MMPI e le diagnosi finali di quegli stessi pazienti. La regola più efficace per la distinzione delle due condizioni cliniche, detta *regola di Goldberg*, era piuttosto semplice: somma i punteggi ottenuti in tre specifiche prove e sottrai quelli ottenuti in altre due; se il totale è infe-

riore a 45, al paziente viene assegnata la diagnosi di neurosi, se è maggiore o uguale a 45, la diagnosi di psicosi.

Goldberg ha quindi raccolto un totale di 861 nuovi MMPI di pazienti seguìti in sette diverse strutture di degenza o ambulatoriali. Tali strutture facevano parte di ospedali universitari, privati o di assistenza ai veterani (*Veterans Administration*) negli stati della California, del Minnesota e dell'Ohio. L'accuratezza delle diverse regole di decisione in questo campione di nuovi casi è stata confrontata con quella di 29 soggetti che hanno analizzato le stesse informazioni ed elaborato la stessa distinzione diagnostica su basi cliniche. Alcuni dei partecipanti erano psicologi esperti specializzati nell'interpretazione del MMPI; altri avevano poca o nessuna precedente esperienza con questo test.

Nei sette diversi ambiti da cui provenivano i casi, la media dell'indice di accuratezza dei partecipanti (cioè del coefficiente di correlazione fra giudizi e diagnosi finali) variava fra 0,15 e 0,43. Per la totalità dei casi il coefficiente complessivo era di 0,28, corrispondente al 62% di diagnosi corrette. Il singolo partecipante con le migliori prestazioni ha ottenuto un coefficiente complessivo di 0,39, corrispondente al 67% di diagnosi corrette. In ciascuno dei sette insiemi di casi, diverse regole di decisione hanno superato il livello medio di accuratezza dei partecipanti. La regola di Goldberg ha ottenuto un livello di accuratezza comparabile a quello dei partecipanti in tre dei sette insiemi di casi; in quattro di essi ha prodotto un incremento, talvolta notevole, dell'accuratezza dei giudizi, con uno scarto che variava fra 0,16 e 0,31. In relazione al campione totale di 861 casi, la regola di Goldberg otteneva un coefficiente di accuratezza di 0,45, corrispondente al 70% di diagnosi corrette, superando quindi sia l'accuratezza media dei 29 partecipanti sia quella del migliore fra loro.

Rorer, Goldberg (1965) hanno indagato se questi risultati potessero essere modificati dando modo ai partecipanti di esercitarsi nello specifico compito diagnostico studiato. I partecipanti ricevevano dei blocchi contenenti il profilo MMPI di 300 pazienti, ognuno dei quali riportava sul retro la diagnosi finale (neurosi o psicosi). In tal modo i partecipanti potevano controllare in una serie di prove l'accuratezza di ciascuna delle diagnosi che avevano elaborato e potevano eventualmente modificare le loro strategie di giudizio. Lo svolgimento di ripetute sessioni di addestramento culminava con un insieme di 4.000 giudizi diagnostici. Ma anche in queste condizioni nessuno dei partecipanti dello studio è arrivato a eguagliare il tasso di accuratezza della regola di Goldberg. Infine, in un'ulteriore variante della loro indagine, Rorer e Goldberg informavano un sottogruppo dei partecipanti (che comprendeva tutti gli psicologi esperti) delle diagnosi ricavate dalla regola di Goldberg. I partecipanti erano al corrente del livello generale di efficacia della regola e potevano in ciascun caso, a loro discrezione, farne uso. Le prestazioni dei partecipanti mostrarono un modesto miglioramento, ma ancora una volta nessuno rag-

giunse il livello di accuratezza della regola di Goldberg. Ciascun partecipante avrebbe quindi giudicato meglio se avesse seguito la regola in tutti i casi.

In un ulteriore studio, sulla base degli stessi 861 test MMPI utilizzati in indagini precedenti, Goldberg (1970) ha costruito modelli matematici (lineari) che riproducessero il più fedelmente possibile i giudizi di ciascuno dei 29 partecipanti che avevano valutato quei casi. Per la costruzione di modelli matematici che colgano le strategie decisionali degli individui non è necessaria la conoscenza degli esiti effettivi dei casi considerati. Ciò che si richiede è invece l'analisi dei rapporti fra le informazioni disponibili a ciascun individuo e i giudizi che egli ha formulato. In linea di principio, se un soggetto pesasse le diverse variabili in modo completamente coerente e affidabile (vale a dire se giudicasse sempre nello stesso modo insiemi di dati fra loro identici), un modello matematico potrebbe riprodurre perfettamente ciascuno dei suoi giudizi. In pratica, gli esseri umani non giudicano in modo del tutto coerente. Di conseguenza, il soggetto e il corrispondente modello matematico saranno talvolta in disaccordo. Goldberg ha scoperto che, in casi di disaccordo, l'individuo dai cui giudizi era stato precedentemente ricavato un particolare modello forniva meno risposte corrette di quello stesso modello. La loro perfetta affidabilità spiega le migliori prestazioni dei modelli matematici in questo e in altri studi simili (Dawes, Corrigan, 1974).

Leli e Filskov hanno studiato la diagnosi di una disfunzione cerebrale progressiva sulla base di test relativi alle capacità intellettuali (Leli, Filskov, 1984). Una regola statistica di decisione, derivata da un insieme preliminare di casi, è stata applicata a un nuovo campione e ha identificato la diagnosi corretta l'83% delle volte. Sulla base degli stessi dati, un gruppo di clinici senza una specifica esperienza in questo compito valutò correttamente il 63% dei casi, mentre un gruppo di clinici esperti ne valutò correttamente il 58%. In una seconda parte dello studio, i partecipanti venivano informati dei risultati dell'applicazione della regola statistica. In questa condizione, sia i clinici esperti sia quelli non esperti mostravano un miglioramento (rispettivamente: 75% e 68% di diagnosi corrette), ma nessuno dei due gruppi raggiungeva il livello di accuratezza dell'83%. Il miglioramento dei giudizi dei partecipanti sembrava dipendere dalla frequenza con cui facevano uso della regola.

Infine Einhorn (1972) ha studiato la previsione del periodo di sopravvivenza sulla base delle biopsie di pazienti affetti da linfoma di Hodgkin. Al momento in cui lo studio è stato condotto, il periodo di sopravvivenza era negativamente correlato con la severità della patologia (oggi il linfoma di Hodgkin può essere efficacemente trattato). Tutti i 193 pazienti coinvolti nello studio in seguito morirono, fornendo così informazioni tragicamente oggettive per il controllo delle previsioni. Tre patologi (uno dei quali era un'autorità internazionalmente riconosciuta) hanno valutato le immagini della prima biopsia di ciascun paziente in relazione a nove indicatori istologici che ri-

tenevano rilevanti per determinare la severità della condizione; quindi, per ogni paziente, hanno fornito una valutazione complessiva della severità. L'analisi statistica dei rapporti fra le valutazioni dei nove indicatori istologici da parte dei patologi e gli effettivi tempi di sopravvivenza dei primi 100 pazienti considerati ha permesso di elaborare diverse formule statistiche per la previsione della sopravvivenza. I 93 casi restanti sono quindi stati utilizzati per la validazione incrociata e il confronto fra i diversi metodi di previsione. I giudizi dei patologi non hanno mostrato praticamente nessuna correlazione con la sopravvivenza dei pazienti, mentre diversi metodi statistici di previsione hanno rivelato una effettiva capacità predittiva, seppur modesta. Lo studio di Einhorn non evidenzia soltanto un vantaggio del metodo statistico; esso mostra anche che le valutazioni dei patologi riguardo ai nove indicatori istologici considerati contenevano informazioni potenzialmente rilevanti, ma che solo il metodo statistico riusciva a metterne a frutto il valore predittivo.

3. Ulteriori ricerche

I tre studi appena visti illustrano gli aspetti fondamentali dell'ampia letteratura scientifica riguardo ai rapporti fra giudizio clinico e giudizio statistico. Queste tre indagini rispettano infatti le condizioni precedentemente specificate per un confronto equilibrato fra i due metodi. Inoltre i loro risultati sono rappresentativi dell'insieme delle ricerche disponibili. Una volta eliminati gli studi nei quali non sono state prese sufficienti precauzioni per evitare che le prestazioni del metodo statistico risultino artificialmente alte, restano circa 100 studi comparativi nell'ambito delle scienze umane. Nella quasi totalità di questi studi, il metodo statistico ha eguagliato o superato quello clinico, con uno scarto talvolta modesto e talvolta notevole (Meehl, 1986; Sawyer, 1966). Per esempio, nello studio di Watley, Vance (1974) sulla previsione dei risultati scolastici di studenti universitari, l'efficacia dei due metodi è risultata comparabile; nello studio di Carroll *et al*. (1982) sull'infrazione delle condizioni di libertà vigilata, il metodo statistico ha ottenuto un leggero vantaggio; e nello studio di Wittman (1941) sulla previsione di una risposta positiva alla terapia con elettroshock il metodo statistico ha fornito circa il doppio di giudizi corretti in confronto al metodo clinico.

La validità e la generalizzabilità dei primi studi comparativi sono state oggetto di diverse critiche. Si è sostenuto, per esempio, che tali studi davano una rappresentazione deformata del metodo clinico per il fatto che i partecipanti non avevano accesso a cruciali fonti di informazione (come colloqui personali con i pazienti), perché erano impegnati in compiti artificiali e non corrispondenti al loro specifico settore di specializzazione o perché l'esperienza e l'abilità dei soggetti coinvolti negli studi era discutibile.

L'evidenza scientifica accumulata negli anni ha permesso di far fronte a queste critiche. Numerosi studi hanno esaminato compiti di giudizio per nulla artificiali, ma comuni nella pratica quotidiana, e per i quali molti esperti offrono la loro consulenza professionale. Gli esempi comprendono i tre studi descritti nel paragrafo precedente, che riguardano la diagnosi differenziale di due disturbi psichiatrici di diversa gravità, la rilevazione di danni cerebrali e la previsione dei tempi di sopravvivenza in pazienti affetti da una patologia grave. Altri studi hanno analizzato la distinzione fra disturbi di tipo medico e disturbi di tipo specificamente psichiatrico (Oskamp, 1962), la descrizione della personalità (Halbower, 1955), la previsione dei risultati di un trattamento (Barron, 1953), la previsione della durata di un periodo di ospedalizzazione per motivi psichiatrici (Dunham, Meltzer, 1946) e del verificarsi di episodi di comportamento violento (Werner, Rose, Yesavage, 1983). Si tratta di problemi e decisioni che praticanti e specialisti affrontano spesso e in molti di questi studi i ricercatori non hanno sottoposto ai partecipanti nuovi problemi definiti appositamente, ma hanno piuttosto esaminato decisioni che erano già state prese nella pratica clinica ordinaria.

In altre indagini i partecipanti potevano accedere alle fonti di informazione che essi prediligevano. In particolare, già nel 1966, Sawyer ha identificato 17 confronti fra metodo clinico e metodo statistico basati sui risultati di test psicologici e di colloqui con i pazienti. In alcuni studi, inoltre, i partecipanti hanno potuto scegliere liberamente quali dati raccogliere e in che modo raccoglierli. Per esempio, nell'indagine di Carroll *et al.* (1982) sulla previsione del comportamento di individui in libertà vigilata le procedure solitamente in uso per la raccolta delle informazioni considerate rilevanti non sono state alterate in alcun modo. E nello studio di Dawes (1971) sulla previsione dei risultati scolastici di studenti universitari, il comitato di ammissione disponeva delle stesse informazioni usualmente utilizzate per la selezione dei candidati. Nessuno dei 17 studi discussi da Sawyer né quelli di Carroll e collaboratori e di Dawes hanno rilevato una qualche superiorità del metodo clinico rispetto a quello statistico.

Infine, gli esiti delle ricerche non risultano alterati dalla considerazione di diversi livelli di esperienza e di abilità all'interno di uno stesso studio o fra studi differenti. Nell'indagine di Goldberg (1965) il giudizio clinico di esperti e non esperti ha dato luogo a prestazioni simili e nessuno dei due gruppi ha superato il livello di accuratezza del metodo statistico – analogamente a quanto osservato da Leli, Filskov (1984) nel problema della rilevazione di danni cerebrali. Ulteriori indagini sull'identificazione e la localizzazione di danni cerebrali hanno dato risultati simili (Leli, Filskov, 1981; Wedding, 1983). Per esempio, Wedding (1983) ha osservato che né i giudizi clinici di soggetti con un'ampia esperienza nell'interpretazione dei test psicologici rilevanti né quelli di un neuropsicologo noto a livello nazionale superavano l'ac-

curatezza di metodi statistici che valutano la presenza, la collocazione e le possibili cause di un danno cerebrale.

Gli studi comparativi spesso non permettono di trarre conclusioni generali sulla superiorità di alcune particolari regole statistiche rispetto ad altre. Inoltre, sebbene alcuni studi (come quelli di Goldberg) abbiamo affrontato il problema dell'applicazione in contesti fra loro differenti, molte delle ricerche hanno fatto uso di campioni limitati. In generale, gli studiosi sono stati meno interessati a determinare l'ambito di applicabilità di specifiche procedure statistiche che non a svolgere ulteriori indagini sul rapporto fra metodo clinico e metodo statistico, estendendo in tal modo il campo degli studi comparativi.

I diversi studi comparativi possono essere visti come una serie di prove ripetute all'interno di un ampio insieme di compiti di giudizio e di previsione rilevanti per lo studio del comportamento umano. In mancanza di una conoscenza completa degli elementi che compongono tale ampio insieme di compiti, la rappresentatività dei risultati ottenuti non può essere valutata con precisione. Tuttavia, con un campione di circa 100 studi e uno stesso risultato in pressocché tutti i casi, è ragionevole concludere che la superiorità del metodo statistico non costituisca un'eccezione, ma sia piuttosto un fenomeno generale che probabilmente si estende a molti dei compiti di giudizio che non sono stati direttamente studiati. In altri termini, l'evidenza disponibile sostiene con forza una risposta positiva alla domanda: «una procedura statistica sviluppata per uno specifico compito (per esempio, la previsione del successo scolastico in una certa università) si rivelerebbe altrettanto efficace, o migliore, del metodo clinico?». Con le parole di Meehl (1986): «non esiste altra controversia nelle scienze umane in cui una così ampia mole di studi qualitativamente diversi dia luogo a risultati tanto concordi» (p. 373).

4. Possibili eccezioni

Se gli studi comparativi generalmente favoriscono il metodo statistico, si potrebbe invertire il punto di vista e indagare in quali specifiche circostanze il giudizio clinico possa rivelarsi migliore di quello statistico. È possibile, per esempio, che il giudizio clinico risulti superiore se può contare su un certo margine di vantaggio informativo? Supponiamo che non vi sia una formula per elaborare gli esiti di un colloquio e che un esperto debba scegliere quale privilegiare fra due conclusioni fra loro divergenti: il giudizio clinico fondato sia su un colloquio con il paziente sia sui punteggi ottenuti dal paziente in una serie di test psicologici, e l'analisi dei soli punteggi dei test psicologici attraverso una regola statistica. I risultati delle ricerche disponibili che affrontano questo particolare problema indicano senza eccezioni che, anche con un certo margine di vantaggio informativo, il giudizio clinico non riesce a supe-

rare l'accuratezza del metodo statistico. E spesso la possibilità di accedere a ulteriori elementi di informazione non influisce affatto sull'entità dello scarto fra i due metodi (Sawyer, 1966; Wiggins, 1981).

Non è difficile ipotizzare altre circostanze in cui il giudizio clinico potrebbe superare il metodo statistico. (1) Un primo caso è quello di giudizi basati su teorie; giudizi del genere sono infatti difficili o impossibili da riprodurre a partire da semplici relazioni statistiche fra variabili empiriche. (2) Un secondo caso è quello in cui certe conclusioni ricavate con procedure statistiche possono essere opportunamente corrette in base all'osservazione di eventi rari o a considerazioni di utilità che non trovano posto in quelle procedure. (3) Infine è possibile che fra le variabili predittive e la condizione da giudicare sussistano complesse relazioni dette «configurali», nelle quali il peso di una variabile dipende dal valore di una o più ulteriori variabili (Meehl, 1957, 1959, 1973).

La potenziale superiorità di giudizi basati su teorie rispetto a conclusioni ottenute dalla considerazione di semplici frequenze empiriche può sembrare un'ovvietà a chi opera nelle scienze «dure». Le previsioni basate su teorie si rivelano efficaci nei casi in cui uno scienziato è a conoscenza dei principali fattori causali rilevanti, possiede accurati strumenti di misurazione per valutarne l'influenza e si affida a principi ben controllati per mettere in relazione teoria e dati di fatto. Per questi motivi, sebbene la maggior parte delle ricerche comparative in medicina favoriscano il metodo statistico, le indagini che indicano un leggero vantaggio del giudizio clinico riguardano ambiti in cui è possibile fondare tale metodo di giudizio su solide basi teoriche (Martin, Apostolakos, Roazen, 1960). Ma, nelle discipline che si occupano degli esseri umani e del loro comportamento, le teorie che guidano le previsioni degli esperti raramente soddisfano queste condizioni. Per esempio, le previsioni relative alla possibile risposta di un paziente alla psicoterapia o al verificarsi di atti di violenza possono fondarsi su teorie psicodinamiche che ammettono conclusioni fra loro opposte e non si avvalgono di tecniche formali di misurazione. È possibile che in psicologia e nelle altre scienze umane i giudizi basati su teorie arrivino a superare in affidabilità il metodo statistico, ma le condizioni necessarie perché ciò si verifichi rappresentano a oggi un obiettivo, o una speranza, ben lontani.

Il giudizio clinico può in certi casi trarre vantaggio dalla considerazione di eventi che, per la loro rarità, non sono inclusi nelle formule statistiche e sono in contrasto con le conclusioni da esse ricavate. Questa osservazione indica la possibilità di un approccio «clinico-statistico», in cui si considerano gli esiti di entrambi i metodi e quindi si decide caso per caso se correggere la valutazione ottenuta con procedure statistiche. In psicologia, la questione è nota come il «problema della gamba rotta». La denominazione è dovuta all'esempio di una ipotetica formula statistica per prevedere se un particolare individuo si

recherà a una proiezione cinematografica settimanale; per quanto efficace, tale formula dovrebbe essere messa da parte se si è scoperto che l'individuo in questione è attualmente ingessato con una frattura al femore (Meehl, 1954, 1957). In un caso del genere, il giudizio clinico permette di correggere il metodo statistico, a condizione di essere in grado di identificare il verificarsi di un evento raro e di decidere di conseguenza. In linea di principio, sarebbe possibile affrontare il problema della rilevanza di eventi rari attraverso il metodo statistico; gli ostacoli pratici sono però enormi (in particolare, l'insieme dei possibili fattori rilevanti è indefinitamente ampio).

La questione appena descritta può essere facilmente indagata fornendo ai partecipanti di uno studio comparativo sia i dati disponibili in ciascun caso sia la conclusione che se ne ricava con il metodo statistico e quindi lasciando che siano gli stessi partecipanti a decidere se adottare o respingere tale conclusione. Sebbene in numero limitato, tutte le indagini su questo tema indicano che il livello più alto di accuratezza si ottiene affidandosi senza eccezioni al metodo statistico e non utilizzandolo in maniera discrezionale (Goldberg, 1968; Sawyer, 1966). A quanto pare, il giudizio clinico tende a identificare troppe «eccezioni». In altri termini, le conclusioni del metodo statistico che vengono respinte risultano essere, per la maggior parte, corrette. Adottando un atteggiamento più conservativo verso le indicazioni del metodo statistico si potrebbero forse ottenere risultati migliori, ma questa congettura deve ancora essere adeguatamente studiata.

La considerazione del valore di utilità delle previsioni solleva un problema analogo. A seconda del compito, certi errori di giudizio possono essere più seri di altri. Per esempio, la mancata identificazione di una condizione patologica destinata a scomparire spontaneamente potrebbe essere meno grave della diagnosi erronea di una condizione per la quale verrà prescritto un trattamento rischioso. Se vi è una discrepanza fra le conseguenze di diverse forme di errore (per esempio, fra falsi negativi e falsi positivi) può essere legittimo modificare una regola di decisione al fine di ridurre in modo selettivo gli errori più gravi, anche se tale modifica può dar luogo a una diminuzione dell'accuratezza complessiva della procedura. In alcuni casi la considerazione dei valori di utilità potrebbe quindi suggerire, su basi razionali, l'adozione del giudizio clinico di un esperto che contraddice il metodo statistico, anche se tale giudizio ha meno probabilità di essere corretto (per esempio, una probabilità di 0,6 contro 0,8). Da un punto di vista formale, questo approccio è equivalente all'inclusione del giudizio clinico dell'esperto all'interno dell'equazione statistica (sotto forma di una nuova variabile). Una valutazione adeguata dell'impatto di questa procedura richiede nuove e specifiche indagini. Non si può però dare per scontato che l'inclusione del giudizio clinico fra le variabili sia d'aiuto. I dati disponibili suggeriscono che tale inclusione non migliora l'accuratezza, né necessariamente l'utilità, dei giudizi stati-

stici. Inoltre i tentativi di adattamento del giudizio statistico su basi informali o soggettive possono facilmente produrre più danni che benefici (Sawyer, 1966).

La possibilità che il giudizio clinico possa avvantaggiarsi della presenza di relazioni configurali fra diversi fattori predittivi solleva due questioni collegate ma distinte, che esamineremo separatamente: la capacità di riconoscere tali relazioni e la capacità di servirsene in vista della diagnosi o della previsione. Nel riconoscimento di schemi e strutture vi sono alcune prestazioni umane che non è ancora possibile riprodurre con strumenti artificiali. Il problema del riconoscimento di schemi visivi ha messo alla prova un'intera generazione di ricercatori nel campo dell'intelligenza artificiale. Gli esseri umani conservano una significativa superiorità, per esempio, nel riconoscimento delle espressioni facciali (si osserva un'analoga superiorità in relazione a prestazioni quali la traduzione linguistica e l'invenzione di teorie con una profonda struttura concettuale). Soltanto un osservatore umano può quindi riconoscere una particolare espressione facciale o un certo modo di muoversi dotati di genuino valore predittivo (come la camminata «fluttuante» di alcuni pazienti schizofrenici). Queste capacità osservative rendono disponibili informazioni utili che altrimenti non sarebbe possibile raccogliere.

Il possesso di capacità osservative uniche implica che l'interazione umana è spesso necessaria per raggiungere la massima accuratezza predittiva (e per la scoperta di variabili potenzialmente utili). Occorre però guardarsi dal trarre frettolosamente un'ulteriore conclusione, molto più dubbia. Una capacità di osservazione unica non equivale a una capacità unica di prevedere attraverso l'integrazione delle osservazioni effettuate. Come abbiamo notato in precedenza, praticamente qualunque osservazione può essere codificata in termini quantitativi ed essere quindi coinvolta in un'analisi statistica. Il già citato studio di Einhorn (1972) e altre indagini simili mostrano che è possibile raggiungere una maggiore accuratezza se gli osservatori esperti intervengono con le loro capacità nella raccolta delle informazioni e quindi fanno un passo indietro, lasciando che l'interpretazione dei dati (osservativi e di altro tipo) sia guidata dal metodo statistico.

5. I fattori alla base della superiorità del metodo statistico

Per spiegare le differenti prestazioni delle procedure statistiche e del giudizio clinico è utile considerare le diverse proprietà dei due metodi. Innanzi tutto, come si è detto, le procedure statistiche conducono sempre alla medesima conclusione in presenza di uno stesso insieme di dati, diversamente da quanto accade nel giudizio umano. In uno studio, per esempio, diversi reumatologi e radiologi hanno rivisto una serie di casi che avevano già valutato in preceden-

za, fornendo in molti casi opinioni differenti (Fries *et al.*, 1986). Fattori quali stanchezza, esperienze recenti, piccole differenze nell'ordine di presentazione delle informazioni, nella formulazione del problema o del compito da svolgere possono produrre significative fluttuazioni nei giudizi degli individui (Hammond, Summers, 1965; Kahneman, Tversky, 1984). Tali fluttuazioni riducono l'affidabilità dei giudizi e di conseguenza la loro accuratezza. In particolare, l'accuratezza risulta evidentemente ridotta dal fatto che gli stessi dati possono produrre un giudizio corretto in un caso e un giudizio differente, ed erroneo, in un altro.

Ancora più importanti sono forse le specifiche proprietà matematiche di un metodo statistico di giudizio, una volta che sia stato ricavato in maniera appropriata. Tale metodo garantisce infatti che le variabili contribuiscano al raggiungimento di una conclusione sulla base del loro effettivo valore predittivo e delle relazioni che intrattengono con la condizione che si intende giudicare. In particolare, le regole statistiche derivate attraverso tecniche di regressione multipla includono soltanto variabili pertinenti, escludendo quelle prive di valore predittivo e assegnano a ciascuna variabile un peso conforme al suo specifico contributo all'accuratezza dei giudizi. Queste peculiari caratteristiche emergono attraverso procedure essenzialmente automatiche nella previsione statistica, ma presentano ostacoli formidabili per la mente umana.

Le ricerche mostrano infatti che gli individui distinguono con difficoltà variabili rilevanti e non rilevanti e spesso sviluppano credenze erronee riguardo all'associazione fra diverse variabili (Chapman, Chapman, 1967, 1969). Occorre inoltre notare che in molti casi psicologi e psichiatri dispongono di informazioni scarse o nulle circa l'accuratezza delle loro diagnosi e previsioni. Per esempio, i consulenti cui si chiede di predire eventuali episodi di violenza possono non venire mai a sapere se le loro previsioni erano corrette. Pertanto, in relazione alla correttezza dei loro giudizi, questi esperti raramente ricevono riscontri immediati e affidabili (per esempio, come quelli che i medici ottengono con l'intervento del patologo alla fine di una conferenza clinicopatologica) (Meehl, 1973). E in mancanza di informazioni chiare riguardo all'accuratezza dei propri giudizi risulta difficile determinare se, e in che misura, le variabili di cui si fa uso sono effettivamente utili. Lo stesso problema può presentarsi se si applicano i metodi statistici «alla cieca», vale a dire a nuove situazioni o in nuovi ambienti senza un controllo preliminare delle loro prestazioni.

Un altro problema consiste nel fatto che, in alcune circostanze, i giudizi clinici possono produrre «profezie che si auto-avverano». La previsione di un certo evento può infatti condurre a decisioni che in seguito influiscono sul verificarsi di quell'evento (Einhorn, Hogarth, 1978). Un aneddoto illustra questo problema. In un processo per omicidio, uno psichiatra prevede che l'imputato in futuro agirà in modo pericoloso. L'imputato è giudicato colpe-

vole e, in considerazione della sua pericolosità, è condannato alla pena capitale. Nel periodo di reclusione che precede l'esecuzione il condannato mostra comportamenti violenti, e questo sembra sostenere la validità delle previsioni dello psichiatra. Tuttavia, una volta condannato a morte, questo individuo ha ben poco da perdere; si sarebbe potuto comportare in modo diverso se la valutazione dello psichiatra, e quindi la sentenza, fossero state differenti.

Occorre anche aggiungere che, quando un evento è noto, esso appare più prevedibile di quanto non fosse in realtà (Fischhoff, 1975) e che le previsioni formulate nel passato vengono spesso erroneamente ricordate come compatibili con il successivo svolgimento dei fatti (Fischhoff, 1980; Fischhoff, Beyth-Marom, 1975). Per esempio, Arkes e collaboratori hanno presentato a diversi gruppi di medici uno stesso caso clinico e hanno chiesto ai partecipanti di assegnare dei valori di probabilità ad alcune diagnosi alternative. Nel gruppo di controllo, le probabilità assegnate alle diverse diagnosi erano fra loro simili. In ciascuno dei gruppi sperimentali i medici venivano informati che una in particolare fra le diagnosi considerate era stata in seguito confermata in modo definitivo. Qualunque fosse la diagnosi che era indicata come corretta, i medici tendevano a dichiarare che quella era la diagnosi a cui avrebbero assegnato la probabilità più alta se avessero dovuto affrontare il caso (si veda il cap. 6 in questo volume). Se i propri giudizi vengono inconsapevolmente deformati per accordarsi con qualunque cosa si sia verificata, allora la conoscenza degli effettivi sviluppi di una situazione che si era chiamati a prevedere può avere un valore correttivo modesto o nullo.

È da notare inoltre che il personale clinico è spesso esposto a campioni di individui in condizioni anomale. Ma in mancanza di campioni rappresentativi può essere difficile, se non impossibile, valutare in modo affidabile le relazioni fra variabili diverse. Immaginate per esempio che la metà circa degli adolescenti che vengono seguiti per una storia di delinquenza giovanile mostri piccole anomalie dell'elettroencefalogramma. Sulla base della compresenza di questi due fattori si potrebbe giungere a considerare le anomalie dell'elettroencefalogramma un indicatore della delinquenza fra gli adolescenti, o a ipotizzare che la delinquenza giovanile sia associata a disfunzioni cerebrali. In effetti, i clinici hanno spesso postulato relazioni di questo tipo (Spreen, 1981). Ma per valutare l'esistenza di una correlazione fra i due fattori è necessario sapere se l'indicatore considerato si presenta più spesso fra i soggetti interessati dalla condizione considerata che fra quelli *non* interessati da quella stessa condizione. Per esempio, per valutare una possibile associazione fra anomalie dell'elettroencefalogramma e delinquenza è necessario sapere anche con quale frequenza gli adolescenti *senza* una storia di delinquenza mostrano o non mostrano anomalie dell'elettroencefalogramma. Inoltre, anche nel caso che una correlazione esista, non è possibile stabilire l'utilità predittiva dell'indicatore a meno che non si sappia anche: (1) di quanto è maggiore

la frequenza dell'indicatore quando la condizione da prevedere è presente rispetto a quando la condizione è assente, e (2) la frequenza generale della condizione[1]. Un indicatore che, in presenza della condizione di interesse, è solo leggermente più frequente potrebbe avere un'utilità molto limitata. Se la condizione è rara, allora i giudizi basati sull'indicatore in questione potrebbero essere in gran parte parte errati, perché nella maggioranza degli individui in cui l'indicatore è presente la condizione non lo è. Per esempio, supponete che in un test psicologico una particolare risposta ricorra nel 10% degli individui con un danno cerebrale e solo nel 5% degli individui sani. Se 9 su 10 pazienti controllati non hanno danni cerebrali, allora la maggior parte dei pazienti che forniscono quella particolare risposta non ha danni cerebrali.

Nella pratica clinica è molto più probabile che vengano valutati i casi in cui una o più condizioni problematiche sono presenti che non i casi in cui esse sono assenti, e l'esposizione a campioni anomali di individui ostacola lo svolgimento di tutti i confronti necessari. Di fatto, l'indagine empirica mostra che le «anomalie» dell'elettroencefalogramma sono piuttosto comuni fra i giovani e suggerisce inoltre che la delinquenza giovanile non è più diffusa fra individui con disturbi neurologici in confronto a individui senza tali disturbi (Capute, Neidermeyer, Richardson, 1968; Spreen, 1981). La diffusione di false credenze in problemi di questo tipo è ulteriormente alimentata da una forte tendenza a dare eccessivo peso alle informazioni coerenti con le proprie ipotesi a discapito di informazioni in contrasto con esse (Greenwald *et al.*, 1986). Con il risultato che credenze o conclusioni erronee, una volta formatesi, tendono a sopravvivere al presentarsi di evidenza contraria.

Numerosi studi sull'accuratezza delle diagnosi e delle previsioni documentano la difficoltà di separare variabili rilevanti e non rilevanti sulla base dell'esperienza clinica (Faust, 1984). Alcune ricerche mostrano per esempio che i giudizi clinici basati su colloqui e interviste raggiungono livelli di validità trascurabili (Carroll *et al.*, 1982). Altre indagini indicano che la validità di giudizi clinici basati sui risultati di test psicologici può essere modesta (Leli e Filskov, 1981, 1984; Wedding, 1983; Werner, Rose, Yesavage, 1983). Sebbene i colloqui e i test psicologici possano fornire valide indicazioni diagnostiche, il metodo clinico spesso non permette di distinguere fra le informazioni utili e quelle che utili non sono.

Un vantaggio del metodo statistico la cui importanza viene talvolta sopravvalutata è l'attribuzione a ciascuna variabile di un peso ottimale. Di fatto, procedure basate su pesi unitari per tutte le variabili (cioè pesi fra loro equivalenti) forniscono previsioni altamente correlate con quelle ottenute integrando le stesse informazioni in base a pesi ottimali (purché, bisogna precisa-

1. In proposito si vedano, in questo volume, i capp. 1, 3, 5 e 12 [N.d.C.].

re, la direzione in cui ciascun fattore è correlato alla condizione da prevedere possa essere specificata in anticipo, e purché i diversi fattori predittivi non siano negativamente correlati fra loro; si vedano Dawes, Corrigan, 1974; Einhorn, Hogarth, 1975; Wainer, 1978; Wilks, 1938). Inoltre, i pesi ottimali delle variabili sono specifici della popolazione di casi da cui sono stati derivati; qualsiasi vantaggio ottenuto in un particolare contesto potrebbe quindi andare perduto applicando lo stesso metodo in un contesto differente. È da notare comunque che, nei casi in cui i pesi ottimali incrementano in modo significativo l'accuratezza predittiva, il giudizio umano si trova in una situazione di forte svantaggio. Con un esempio di Meehl (1986, p. 372).

> Tutti sappiamo bene che il cervello umano è poco efficiente nel pesare e calcolare diversi fattori. All'uscita dal supermercato non accade che diamo una rapida occhiata alla borsa della spesa e diciamo: «Beh, direi che sono circa 17 dollari; lei che ne pensa?». Il cassiere esegue la somma. Non ci sono argomenti solidi […], basati su studi empirici […], per ritenere che gli esseri umani, su basi soggettive, sappiano assegnare pesi ottimali a diversi fattori; né che essi giudichino applicando in modo coerente tali pesi.

Si potrebbe obiettare che questa analogia (la cui funzione è illustrativa e non probatoria) fa riferimento a un modello additivo e per questo motivo potrebbe essere respinta da chi sostiene che il giudizio umano si basa su strategie più sofisticate (come regole configurali). Supponete però che la regola per calcolare i prezzi del supermercato sia la seguente: «nel caso in cui siano state acquistate sia carne bovina sia verdure fresche, moltiplica il logaritmo in base 0,78 del prezzo della carne per la radice quadrata del doppio del prezzo delle verdure»; forse in queste condizioni il cliente e il cassiere calcolerebbero meglio il conto a mente? Quasi certamente, farebbero peggio. Se gli esseri umani hanno scarse prestazioni dovendo stimare e combinare i termini di una semplice funzione matematica additiva, non c'è ragione di aspettarsi che facciano meglio dovendo eseguire un'integrazione complessa degli stessi fattori.

6. Scarsa influenza e fonti di resistenza

La ricerca su giudizio clinico e giudizio statistico ha avuto un modesto impatto sulle pratiche decisionali, a partire dal settore nel quale essa ha avuto origine, vale a dire la psicologia clinica. Secondo la rassegna di Guilmette *et al*. (1990) la maggior parte degli psicologi che si specializzano nella valutazione di danni cerebrali preferiscono procedure per le quali non sono stati sviluppati metodi statistici in confronto a quelle per cui tali metodi sono disponibili. Lo svolgimento di un colloquio è ancora considerato una condizio-

ne imprescindibile per l'accesso a qualsiasi programma di formazione nel campo della salute mentale e in molti stati è esplicitamente richiesto per l'abilitazione alla pratica clinica. Ma, a dispetto dei risultati degli studi che mostrano come l'interpretazione clinica dei colloqui abbia un valore predittivo scarso o nullo, l'interpretazione dei colloqui con strumenti statistici (il cui valore è ben documentato) viene usata di rado o non è usata affatto.

Tale scarsa influenza è in parte dovuta a una scarsa familiarità con l'evidenza scientifica disponibile. In alcuni casi, manca la consapevolezza dei risultati degli studi comparativi e persino dell'esistenza del problema. In altri casi, facendo riferimento alle prime indagini pubblicate, si argomenta che il giudizio clinico è stato studiato in condizioni svantaggiose, senza considerare che ricerche successive hanno reso superati tali argomenti. Anche quando l'evidenza scientifica è nota, è possibile che venga trascurata in modo tendenzioso o pregiudiziale. In particolare, la formazione, l'addestramento, gli orientamenti teorici e i valori personali dei professionisti che operano nel campo della salute mentale possono ostacolare il riconoscimento della superiorità del metodo statistico. Alcuni psicologi ritengono per esempio che l'uso di equazioni «deumanizzi» i loro pazienti. Questo punto di vista trascura completamente i costi umani associati agli errori dovuti al mancato impiego delle migliori procedure disponibili.

Un argomento, o pregiudizio, comune contro il metodo statistico consiste nel sostenere che le statistiche non si applicano a singoli individui o a singoli eventi. Questo argomento vìola aspetti elementari della nozione di probabilità: sebbene individui ed eventi possano mostrare alcune caratteristiche uniche, essi tipicamente condividono numerose caratteristiche con altri individui o eventi e ciò rende possibile generalizzazioni dotate di potere predittivo. Un fautore di questo argomento antistatistico sarebbe tenuto a sostenere, per coerenza logica, che se si fosse costretti a giocare una sola volta alla roulette russa, e si avesse la possibilità di scegliere fra una pistola con cinque pallottole e una con una sola pallottola, l'unicità dell'evento renderebbe la scelta del tutto arbitraria.

Infine, una fiducia infondata nell'accuratezza del giudizio clinico e la falsa impressione che il metodo statistico sia inferiore possono essere alimentate da valutazioni soggettive. La derivazione e il controllo empirico di un metodo statistico forniscono informazioni oggettive sulla qualità delle prestazioni (Dawes, 1979). Ma anche considerando le ricerche che indicano, per esempio, che la regola di Gooldberg per l'interpretazione del MMPI ha raggiunto un'accuratezza del 70%, superando il giudizio clinico dei 29 partecipanti, è ancora possibile restare convinti che la propria capacità di giudizio sia superiore. Gli aspetti più immediati e salienti dell'esperienza clinica possono favorire errori di valutazione di questo tipo. È infatti possibile che si ricordino con facilità casi drammatici in cui le proprie interpretazioni si sono rivelate

corrette o in cui si è evitato un errore discostandosi da una procedura statistica, senza riconoscere, o senza considerare a sufficienza, esempi di segno contrario.

In definitiva, la scelta è fra le proprie osservazioni e impressioni personali e l'evidenza scientifica disponibile sull'efficacia relativa del metodo clinico e di quello statistico. I fattori che ostacolano una valutazione affidabile dell'accuratezza del proprio giudizio clinico sono precisamente quelli che le comuni procedure scientifiche – come le tecniche di campionamento, la manipolazione sperimentale delle variabili e la valutazione «in cieco» degli esiti – mirano a contrastare. La mancata accettazione dei risultati di una mole di evidenza scientifica estesa e convergente in virtù di osservazioni personali di dubbia validità può essere descritta come un normale errore umano; oppure, nel caso di professionisti che si identificano in un approccio scientifico, come una forma di pura e semplice irrazionalità.

7. Applicazioni del metodo statistico: limiti, benefici e implicazioni

Le ricerche analizzate in questo articolo indicano che i metodi statistici, se sviluppati e applicati in modo appropriato, sono in grado aiutare nella diagnosi e nella previsione di eventi che riguardano gli esseri umani e il loro comportamento. Tali metodi risultano altrettanto efficaci, e spesso migliori, del giudizio clinico, anche in casi in cui quest'ultimo si fonda su un insieme di informazioni altrettanto esteso o addirittura più ampio. Le ricerche che documentano la superiorità dell'approccio statistico dovrebbero però essere integrate dalla consapevolezza dei limiti di tale approccio e della necessità di opportune procedure di controllo.

Innanzi tutto, per quanto superiori al metodo clinico, le procedure statistiche non sono affatto infallibili e in alcuni casi le loro prestazioni restano modeste. In secondo luogo, anche una specifica procedura che risulti efficace in un particolare ambito dovrebbe essere periodicamente ricontrollata e non dovrebbe essere estesa a contesti differenti senza le necessarie precauzioni. Sebbene considerazioni teoriche ed empiriche suggeriscano che la scelta delle variabili rilevanti è spesso più importante dello specifico peso che si attribuisce loro, è possibile utilizzare particolari tecniche statistiche per determinare il peso delle variabili in modo da ottimizzare l'accuratezza di una procedura nell'applicazione a nuovi casi all'interno di una stessa popolazione. È inoltre possibile monitorare l'accuratezza via via che si procede a elaborare previsioni e modificare in meglio i metodi disponibili a fronte di eventuali cambiamenti del contesto o della popolazione di riferimento. Infine è possibile controllare se l'accuratezza delle procedure può essere migliorata con l'aggiunta di nuove variabili.

Metodi statistici opportunamente costruiti possono produrre diversi benefici. Anche nei casi in cui essi si limitano a eguagliare l'accuratezza del metodo clinico, il loro utilizzo può ridurre il dispendio di tempo e di risorse. Per esempio, ogni anno si impiegano milioni di dollari e il prezioso tempo di numerosi specialisti nel tentativo di prevedere il comportamento violento. La predizione del comportamento violento su basi statistiche è una procedura molto più economica e permetterebbe di utilizzare meglio le risorse (per esempio per esigenze terapeutiche che al momento restano insoddisfatte). Inoltre, anche in situazioni in cui le procedure statistiche non vengono utilizzate come unico strumento decisionale, è possibile impiegarle per escludere preliminarmente possibilità o opzioni che non sarebbero comunque state scelte, rendendo più efficiente il processo di valutazione.

Nei casi in cui i metodi statistici risultano più accurati del giudizio clinico i benefici per gli individui e la società sono evidenti. I vantaggi di una maggiore accuratezza nella previsione del comportamento violento o della violazione della libertà vigilata, così come nella diagnosi dei disturbi o nell'identificazione di trattamenti efficaci sarebbero ovviamente notevoli. Oltre a ciò, la conoscenza – fondata su basi oggettive – dei limiti conoscitivi e predittivi delle procedure impiegate può in certi casi evitare il verificarsi di conseguenze indesiderate. Se, per esempio, una diagnosi erronea di Alzheimer formulata con grande sicurezza su basi cliniche lasciasse il posto a una valutazione più cauta (o, meglio ancora, alla diagnosi corretta) sarebbe possibile evitare molte sofferenze non necessarie.

Occorre poi osservare che i metodi statistici si basano su procedure esplicite, in contrasto con il giudizio clinico che coinvolge processi mentali spesso difficili da specificare. Le procedure esplicite rendono possibile la critica informata e sono accessibili ad altri membri della comunità scientifica che intendano replicarne o estenderne le applicazioni.

Infine, i metodi statistici – almeno negli ambiti discussi in questo articolo – rivelano i limiti superiori delle nostre attuali capacità predittive. La consapevolezza dei risultati talvolta modesti che è possibile raggiungere anche con i migliori metodi disponibili corregge una fede irrealistica nelle nostre capacità di comprensione e di previsione. Sostituire a una fiducia eccessiva una sobria incertezza può essere di grande importanza se incoraggia l'apertura a nuovi approcci e alla considerazione di nuovi fattori in grado di aumentare l'efficacia delle nostre spiegazioni e previsioni.

L'argomento per cui in molte decisioni cliniche non sono attualmente disponibili metodi statistici non spiega perché quelli disponibili non vengano utilizzati e sottovaluta la relativa facilità con la quale è possibile sviluppare procedure di questo tipo in ambiti specifici. Anche in casi in cui è difficile ottenere informazioni oggettive sugli eventi o le condizioni che si intende prevedere, è possibile modellare statisticamente il giudizio clinico degli

esperti: le procedure così ottenute probabilmente supereranno in accuratezza quegli stessi esperti (Dawes, Corrigan, 1974; Goldberg, 1970). Ciò di cui vi è bisogno è un maggiore sviluppo di procedure statistiche e di programmi di valutazione che tengano sotto controllo i due metodi di giudizio e i loro livelli di accuratezza nei diversi ambiti, rendendo così possibile una scelta informata fra procedure alternative. Mettersi al lavoro su questi obiettivi si rivelerà molto più produttivo che ignorare l'evidenza scientifica o lamentare la mancanza di strumenti già pronti per l'uso.

Nota dei curatori. Nella riedizione del presente saggio apparsa in Gilovich, Griffin, Kahneman (2002) gli autori hanno aggiunto un Poscritto nel quale si osserva: «Gli argomenti proposti in questo articolo sono stati notevolmente rafforzati da due pubblicazioni apparse in seguito. La prima meta-analisi relativa al confronto fra previsione clinica e previsione basata su procedure standardizzate (Grove *et al.*, 2000) ha rivelato che, su 136 studi comparativi considerati, tutti tranne 8 documentavano la superiorità predittiva delle procedure standardizzate o la loro equivalenza con il metodo clinico. Questi 8 studi (il 6% del totale) non avevano nulla in comune, se non il fatto che il metodo clinico era avvantaggiato dall'accesso a un maggior numero di informazioni, e potrebbero rappresentare risultati non replicabili dovuti a semplici errori di campionamento. Grove, Meehl (1996) hanno inoltre esaminato le comuni e persistenti obiezioni all'adozione di metodi statistici e hanno mostrato come esse siano in contrasto con l'evidenza empirica, logicamente o matematicamente fallaci, oppure infondate da un punto di vista etico o economico».

*14. Ricordi di una colonoscopia: uno studio randomizzato**

di *Donald A. Redelmeier, Joel Katz, Daniel Kahneman*

1. Introduzione

Il cancro colorettale è una delle principali cause di morte negli Stati Uniti, con circa 150.000 nuovi casi e 60.000 morti l'anno (Silverberg, Boring, Squires, 1990). Per una persona di cinquant'anni il rischio di sviluppare un cancro colorettale entro gli ottant'anni è del 5% e il rischio di morirne è del 2,5% (Seidman *et al*., 1985). Nel tentativo di ridurre la diffusione e la mortalità associate al cancro colorettale, le autorità sanitarie raccomandano controlli periodici per le persone a rischio (Winawer *et al*., 1997; Burt, 2000; Imperiale *et al*., 2000). La colonoscopia è la procedura di controllo associata ai maggiori benefici clinici, ai più alti costi immediati e al più esteso dibattito fra gli studiosi (Goldman, 1989; Lieberman *et al*., 2000). Come accade per molte altre procedure di controllo sicure e accurate, il contributo della colonoscopia nel ridurre il rischio di morte da cancro colorettale dipende fortemente dalla disponibilità dei pazienti a sottoporsi al trattamento (Redelmeier, 1995; Frazier *et al*., 2000).

Il ricordo che i pazienti conservano di procedure mediche spiacevoli influenza le loro decisioni riguardo a futuri trattamenti (Erskine *et al*., 1990). Il 20% circa delle donne che rifiuta la mammografia cita il dolore provato durante esami precedenti come l'unica ragione alla base della decisione di non aderire a successive raccomandazioni di sottoporsi al test (Baines, To, Wall, 1990). Quasi il 10% degli individui che evita una visita dentistica fa riferimento al disagio causato da precedenti interventi nel motivare il suo rifiuto (Kent, 1985). Circa il 40% dei pazienti che rimane mentalmente competente dopo essere sopravvissuto a un arresto cardiaco sceglie di evitare, per il futuro, tentativi di rianimazione, in parte a causa del ricordo del trattamento che

* Traduzione di *Memories of colonoscopy: a randomized trial*, «Pain», 104 (2003), pp. 187-194. Per gentile concessione dell'International Association for the Study of Pain.

ha subito (Bedell *et al.*, 1983). Queste osservazioni indicano che un'esperienza avversa, anche solo momentanea, può avere effetti duraturi sul comportamento dei pazienti.

I risultati della ricerca psicologica suggeriscono che la memoria è una facoltà cognitiva imperfetta e soggetta a distorsioni sistematiche. In particolare, le esperienze dolorose sembrano essere rappresentate nella memoria dalle caratteristiche di alcuni loro momenti specifici e la valutazione complessiva di un certo episodio è in gran parte determinata dalle caratteristiche dei particolari momenti che sono stati in tal modo selezionati (Varey, Kahneman, 1992; Frederickson, Kahneman, 1993; Kahneman *et al.*, 1993; Stone *et al.*, 2000; Diener, Wirtz, Oishi, 2001). La durata di un certo episodio spiacevole ha un impatto modesto sul modo in cui viene in seguito giudicato, mentre la sua parte peggiore e l'entità del dolore che precede immediatamente la sua conclusione esercitano un peso notevole sull'impressione finale. Abbiamo indagato se queste particolari caratteristiche del funzionamento della memoria possono offrire strumenti di tipo non farmacologico per modificare il ricordo che i pazienti conservano di una procedura medica dolorosa, aumentando la probabilità che gli individui decidano di tornare a sottoporsi a successivi controlli colonoscopici.

2. Metodi

2.1. *Pazienti*

Abbiamo reclutato pazienti ambulatoriali cui era stata consigliata una colonoscopia e che si sono presentati a svolgere l'esame fra il 7 novembre 1994 e il 12 dicembre 1995. Nessuno degli interventi è stato effettuato esclusivamente ai fini della ricerca. Sono stati esclusi i pazienti che non parlavano inglese, che soffrivano di seri disturbi delle funzioni cognitive o di severe patologie concomitanti. La struttura ospedaliera permetteva di svolgere simultaneamente esami colonoscopici in tre diverse stanze; il primo paziente di volta in volta disponibile è stato selezionato per lo studio. Ciascun partecipante ha fornito il suo consenso informato in forma scritta e il protocollo dello studio è stato approvato dallo Wellesley Hospital Human Ethics Committee (compreso l'impiego di valutazioni in tempo reale per controllare l'attenzione e mantenere l'occultamento dell'intervento sperimentale nei confronti dei pazienti). La programmazione degli interventi è stata eseguita da personale amministrativo dell'unità del medico responsabile. Il personale coinvolto non era a conoscenza del protocollo sperimentale.

2.2. *Randomizzazione e occultamento dell'intervento sperimentale*

Diversamente da quanto accade in molti altri studi, questa indagine richiedeva lo svolgimento di alcuni compiti in tempo reale durante la procedura. Attraverso un processo di randomizzazione determinato in anticipo, computerizzato e ignoto a tutti gli individui coinvolti sono stati preparati involucri sigillati, opachi e numerati contenenti le istruzioni relative all'intervento su ciascun paziente individuale (procedura convenzionale o procedura modificata). Gli involucri sono stati associati a ciascun paziente all'inizio della colonoscopia ma sono stati aperti soltanto nella fase finale della procedura, cioè in quella in cui l'estremità del colonoscopio era entrata nel *cecum* o in cui il medico presente riteneva che un ulteriore inserimento del colonoscopio non fosse più praticabile. Rimandando la comunicazione delle istruzioni relative all'intervento sperimentale si è garantito che i medici non alterassero consapevolmente o inconsapevolmente le fasi precedenti della procedura. I pazienti non sono stati informati in alcun modo del fatto che erano stati assegnati a un particolare gruppo o del fatto che erano stati sottoposti a un qualche intervento sperimentale.

2.3. *Descrizione dell'intervento sperimentale*

La procedura modificata consisteva in un intervento di tipo non farmacologico volto a indebolire il ricordo del dolore provato durante la colonoscopia. L'obiettivo era quello di ridurre al minimo il livello di dolore fisico durante gli ultimi minuti della procedura, permettendo in tal modo ai pazienti di conservare un ricordo migliore dell'esperienza. A questo scopo, prima della rimozione definitiva del colonoscopio, la sua estremità veniva lasciata immobile nel retto per un periodo fino a tre minuti (in questo lasso di tempo non aveva luogo alcuna operazione di aspirazione, di pressione o di rilascio di sostanze anestetiche aggiuntive). La procedura modificata produceva quindi un'estensione della durata dell'intervento, ma faceva sì che i momenti conclusivi risultassero meno dolorosi. La nostra ipotesi era che questo intervento potesse indebolire il ricordo del dolore e permettere ai pazienti di conservare un'impressione più favorevole (meno sfavorevole) dell'esperienza. Per far sì che i pazienti rimanessero all'oscuro della modifica introdotta nella procedura, abbiamo chiesto a tutti gli operatori coinvolti di mantenere la propria posizione e di non modificare il proprio comportamento fino alla fine dell'esame.

2.4. *Valutazioni in tempo reale*

Abbiamo impiegato l'approccio elaborato da Gottman, Levenson (1985) per determinare il livello di dolore durante procedure mediche invasive attraverso l'elicitazione di resoconti momento per momento per l'intero intervallo di tempo fino al completamento dell'intervento. In particolare, ai pazienti veniva fornito un dispositivo manuale che controllava la posizione di un indicatore sullo schermo di un computer. Il computer presentava un'immagine simile a quella di un termometro, con le due estremità associate alle etichette «nessun dolore» e «dolore estremo». Muovendo il dispositivo (o indicando a un assistente coinvolto nella ricerca come farlo) il paziente determinava la posizione dell'indicatore in modo da esprimere il livello di dolore provato al momento. Il computer era programmato per iniziare 30 secondi dopo l'inserimento del colonoscopio e richiedeva una valutazione del dolore da parte del paziente a intervalli di 60 secondi fino a che il colonoscopio non veniva definitivamente rimosso.

2.5. *Definizione dei termini*

Il computer registrava le posizioni dell'indicatore e le convertiva su una scala che variava fra 0 e 10, con il valore «0» corrispondente a «nessun dolore» e «10» a «dolore estremo». Cinque dati statistici sono stati calcolati per ciascun paziente. Con il termine «durata» indicheremo il tempo, in minuti, trascorso fra l'introduzione del colonoscopio e la sua rimozione. Con «dolore medio» indicheremo l'intensità media durante l'intera procedura. Con «picco di dolore» indicheremo l'intensità del dolore nel momento peggiore. Con «dolore iniziale», «dolore intermedio» e «dolore finale» indicheremo rispettivamente l'intensità media durante i 3 minuti iniziali, centrali e finali della procedura. Con l'eccezione della durata, tutti i valori delle quantità descritte variano fra 0 e 10. Un valore più alto indica una maggiore intensità di dolore.

2.6. *Valutazioni retrospettive*

Dopo che la colonoscopia si era conclusa e gli effetti di meperidina, midazolam o altri anestetici si erano esauriti, i pazienti erano chiamati a riflettere sulla loro esperienza e a compilare un questionario. Per esprimere la valutazione globale dell'esperienza sono state utilizzate due misure. In primo luogo, ai pazienti era richiesto di quantificare «il disagio complessivo causato dalla procedura» attraverso una scala visiva di 10 cm con le due estremità etichettate rispettivamente con «nessun disagio» e «enorme disagio» (Katz,

Melzack, 1999). In seguito, si chiedeva ai pazienti di ordinare gli elementi di una lista di otto eventi personalmente spiacevoli (come «una visita media dal dentista» e «due giorni a letto con l'influenza») nella quale era compresa «la colonoscopia cui ti sei appena sottoposto» (Chapman *et al.*, 1985). Studi precedenti suggeriscono che queste due misure forniscono valutazioni valide e affidabili dei ricordi a breve termine dei pazienti riguardo al dolore provato durante la colonoscopia (Redelmeier, Kahneman, 1996).

2.7. *Successivi controlli colonoscopici*

Abbiamo indagato in che misura i pazienti si sono sottoposti a successivi controlli analizzando i dati del Ministero della Salute nell'intervallo fra l'1 novembre 1994 e l'1 agosto 2000. Nello stato dell'Ontario, la colonoscopia è effettuata esclusivamente da medici, i medici vengono retribuiti soltanto se la relativa fattura viene inviata al Ministero e i data-base ministeriali registrano la ricezione delle fatture indipendentemente dal modo in cui avviene la retribuzione. Abbiamo svolto una ricerca sui data-base attraverso informazioni identificative dei pazienti e raccolto i dati relativi alla prima colonoscopia successiva, indipendentemente dalla struttura presso la quale era stata svolta. Sono stati così identificati i pazienti che in qualunque momento si sono sottoposti a una successiva colonoscopia, indipendentemente dal fatto che siano ritornati nella stessa struttura ospedaliera e dalle raccomandazioni cliniche che possono aver ricevuto.

2.8. *Analisi statistica*

L'analisi principale è rappresentata dal confronto delle proporzioni di pazienti che si sono sottoposti a una successiva colonoscopia fra quelli che hanno ricevuto il trattamento convenzionale e fra quelli che hanno ricevuto il trattamento modificato. Si è determinata l'ampiezza del campione in modo da poter identificare una differenza del 10% con una probabilità dell'80% (Hulley, Cummings, 1988). La regressione logistica è stata impiegata per controllare l'eventuale interferenza dovuta ai seguenti fattori: età, genere, peso, altezza, indice di massa corporea, precedenti colonoscopie, livello di istruzione, indicazioni specifiche, uso di analgesici (di qualsiasi tipo), anormalità emerse dall'esame (presenti oppure assenti) e interventi effettuati durante l'esame (presenti oppure assenti). Le proporzioni di pazienti sottoposti a ulteriori colonoscopie sono state inoltre analizzate in relazione alle valutazioni retrospettive (per controllare la robustezza di tali valutazioni, quelle ottenute attraverso la scala visiva da 0 a 10 e quelle ottenute attraverso l'ordinamento delle diverse esperienze spiacevoli sono state analizzate separatamente). Tutti i valori p so-

no a due code, sono stati calcolati in base ad analisi *intent-to-treat* e impiegando i protocolli di sicurezza dello Institute for Clinical Evaluative Sciences.

3. Risultati

3.1. *Presentazione generale dei dati descrittivi*

Abbiamo contattato in totale 733 pazienti e 682 hanno acconsentito a partecipare allo studio. Fra i pazienti che hanno declinato la proposta l'età era più alta (in media di 62 anni *vs*. 57, $p = 0{,}025$) e la probabilità di avere un livello di istruzione universitario era inferiore (40% *vs*. 60%, $p = 0{,}005$). Per il resto, questo gruppo presentava una distribuzione di genere e di tempi di attesa simile a quella del gruppo dei partecipanti.

Tab. 1 - Caratteristiche dei pazienti

		Procedura convenzionale (n = 345)	*Procedura modificata (n = 337)*
Età (anni)	20-39	13	12
	40-59	40	41
	60-79	40	39
	80-99	7	7
Genere (% maschi)		48	55
Peso	40-59	23	17
	60-79	45	49
	80-99	27	29
	≥ 100	5	5
Altezza (cm)	140-154	7	7
	155-169	42	35
	170-184	44	52
	185-200	6	7
Indice di massa corporea (kg/m^2)	15-19	11	10
	20-24	42	44
	25-29	36	33
	≥ 30	11	13
Educazione (livello più alto)	Media inferiore	10	12
	Media superiore	29	29
	Università	42	42
	Post-laurea	19	17
Precedenti colonoscopie?		47	41

I dati riportati sono % in ciascun gruppo. A causa di possibili approssimazioni, la somma per alcune caratteristiche può non essere esattamente equivalente a 100.

Il gruppo dei pazienti assegnati casualmente alla procedura clinica modificata era simile, dal punto di vista delle caratteristiche demografiche, al gruppo di pazienti assegnati casualmente alla procedura convenzionale (tab. 1). Inoltre i due gruppi di partecipanti sono risultati simili rispetto alle indicazioni cliniche, all'uso di analgesici, ai risultati ottenuti attraverso l'esame colonoscopico e all'eventuale esecuzione di altri interventi durante la colonoscopia (tab. 2). Si sono verificate complicazioni in quattro pazienti (due per ogni gruppo) rappresentate da sanguinamento associato a un intervento di polipectomia (un caso), ipotensione (un caso), ipossiemia (un caso) e malfunzionamento dell'attrezzatura (un caso).

*Tab. 2 - Descrizione delle procedure**

		Procedura convenzionale (n = 345)	*Procedura modificata (n = 337)*
Indicazione clinica	Dolore	19	22
	Sangue nelle feci	11	12
	Anemia	5	6
	Modificazioni nella defecazione	14	16
	Screening	19	16
	Controllo	27	22
	Altro**	5	5
Merepedina	Nessuna dose (0 mg)	18	15
	Dosi basse (fino a 50 mg)	65	69
	Dosi medie (fra 50 e 100 mg)	15	15
	Dosi alte (più di 100 mg)	1	0
Midazolan	Nessuna dose (0 mg)	12	11
	Dosi basse (fino a 3 mg)	54	53
	Dosi medie (fra 3 e 6 mg)	31	32
	Dosi alte (più di 6 mg)	3	3
Altri anestetici?		3	5
Nessun anestetico		12	10
Risultati	Normali	46	54
	Polipo singolo	16	15
	Polipi multipli	17	15
	Colite/proctite	9	7
	Angiodisplasia	1	1
	Diverticolite	5	4
	Emorroidi	1	2
	Lesione	1	1
	Altro***	3	2
Intervento	Biopsia	2	2
	Polipectomia con laccio (*snare*)	12	12
	Altro****	1	1

* I dati riportati sono % in ciascun gruppo.
** Comprende: perdita di peso, accertamento relativo a colite (colite di Crohn o colite ulcerosa), risultati radiologici anomali e disturbi non specifici.
*** Comprende: *melanosis coli*, varici, sarcoma di Kaposi, granuloma, stenosi, elmintiasi, risultati sconosciuti, esame inadeguato.
**** Comprende: dilatazione, cauterizzazione.

3.2. *Valutazioni in tempo reale*

Non si è trovata alcuna differenza significativa fra i due gruppi rispetto al livello di dolore registrato nella parte iniziale della procedura, nel momento intermedio e in quello di dolore più intenso (tab. 3). Questo mostra che gli accorgimenti per l'occultamento dell'intervento sperimentale sono risultati efficaci. Come previsto, il livello di dolore durante la parte finale della procedura è stato in media inferiore nel gruppo dei pazienti sottoposti al trattamento modificato (1,7 *vs*. 2,5, $p < 0{,}001$). La durata complessiva della colonoscopia è stata superiore di circa un minuto per il gruppo di pazienti sottoposti al trattamento modificato (27,6 *vs*. 26,8, $p > 0{,}20$). Non si è trovata alcuna correlazione significativa fra l'intensità media del dolore durante la procedura e la sua durata complessiva ($r = 0{,}01$, $p > 0{,}20$), il che suggerisce che procedure più estese nel tempo hanno normalmente implicato più dolore totale.

*Tab. 3 - Valutazioni in tempo reale dell'esperienza**

	Procedura convenzionale	*Procedura modificata*
Dolore iniziale	3,7 ± 2,5	3,4 ± 2,4
Dolore intermedio	3,8 ± 3,0	3,6 ± 2,8
Dolore finale**	2,5 ± 2,4	1,7 ± 2,0
Picco di dolore	8,1 ± 2,3	8,0 ± 2,2
Dolore medio	3,7 ± 2,0	3,5 ± 1,8
Durata (min)	27 ± 15	28 ± 14

* I dati riportati sono medie e deviazioni standard per ciascun gruppo. Il dolore varia fra 0 e 10. Valori più alti indicano una maggiore intensità.
** $p < 0{,}001$ per la differenza fra i due gruppi. Per tutte le altre differenze $p > 0{,}10$.

Si sono riscontrate associazioni significative fra alcuni fattori e una maggiore intensità o durata della procedura. L'intensità del dolore durante la colonoscopia è stata in media più alta per le donne che per gli uomini (4,0 *vs*. 3,3, $p < 0{,}001$), in accordo con precedenti ricerche (Fillingim, 2000). A parte la stratificazione per genere, non sono state trovate altre differenze rispetto ai fattori presenti in tab. 1. I pazienti che hanno ricevuto analgesici non mostravano una tendenza significativa verso un più alto livello di dolore medio (3,7 *vs*. 3,2, $p > 0{,}20$), forse a causa dello stesso processo di selezione di questo gruppo. In media, la procedura ha richiesto circa 2,6 minuti in più fra i pazienti oltre i sessant'anni (28,6 *vs*. 26,0, $p = 0{,}022$) e 2,5 minuti in meno fra i pazienti con un livello di istruzione universitario (26,1 *vs*. 28,6, $p = 0{,}031$). Genere, altezza, peso e uso di analgesici non sono risultati associati con una durata più estesa della procedura.

*Tab. 4 - Valutazioni retrospettive**

	Procedura convenzionale	*Procedura modificata*	*p*
*Valutazioni sulla scala visiva***			
Analisi sui dati grezzi	4,9 ± 2,6	4,4 ± 2,5	0,006
Aggiustata per dati demografici***	4,9 ± 2,5	4,4 ± 2,4	0,006
Aggiustata per esami precedenti	4,9 ± 2,6	4,4 ± 2,5	0,002
Aggiustata per uso di anestetici	4,9 ± 2,6	4,4 ± 2,5	0,006
Aggiustata per tutte e tre le classi di fattori	4,9 ± 2,5	4,4 ± 2,4	0,002
*Classificazione rispetto ad altre esperienze spiacevoli****			
Analisi sui dati grezzi	4,6 ± 2,1	4,1 ± 2,1	0,002
Aggiustata per dati demografici***	4,6 ± 2,1	4,1 ± 2,1	0,002
Aggiustata per esami precedenti	4,6 ± 2,1	4,1 ± 2,1	0,002
Aggiustata per uso di anestetici	4,6 ± 2,1	4,1 ± 2,1	0,002
Aggiustata per tutte e tre le classi di fattori	4,6 ± 2,0	4,1 ± 2,1	0,002

* I dati riportati sono medie e deviazioni standard in ciascun gruppo.
** Le misure variano fra 0 e 10. Valori più alti indicano un dolore più intenso.
*** I dati demografici sono: età, genere, peso, altezza ed educazione.
**** Le misure variano fra 1 e 8. Valori più alti indicano un dolore più intenso.

3.3. *Valutazioni retrospettive*

Vi è stata una variabilità sostanziale fra i pazienti nel ricordo del dolore totale provato. A dispetto di quanto si è appena visto riguardo alle valutazioni in tempo reale, i pazienti sottoposti alla procedura modificata, d'accordo con la nostra ipotesi iniziale, hanno ricordato meno dolore totale, come è indicato dalle valutazioni sulla scala visiva, che in questo gruppo sono in media inferiori del 10% (tab. 4). Inoltre, i pazienti che hanno ricevuto il trattamento modificato hanno generalmente classificato la procedura come meno spiacevole nel confronto con le altre sette esperienze negative. Risultati simili sono stati osservati dopo aver controllato la possibile influenza dei fattori in tab. 1 e 2. Questi risultati sono corroborati dalle proporzioni di pazienti che hanno valutato la procedura indicando l'estremo negativo («enorme disagio») della scala visiva (rispettivamente: 3,3% nel gruppo sperimentale *vs*. 6,7% in quello di controllo, $p = 0{,}039$) o che hanno classificato la colonoscopia come la peggiore fra tutte le esperienze personali negative considerate (8,3% *vs*. 14,0%, $p = 0{,}018$).

I ricordi del dolore della colonoscopia sono risultati correlati con alcuni aspetti delle valutazioni in tempo reale ($p < 0{,}05$). Sia le valutazioni quantitative sulla scala visiva sia la classificazione rispetto ad altre esperienze spiace-

voli correlano con il picco di dolore (rispettivamente, $r = 0{,}44$ e $r = 0{,}26$), con il dolore iniziale (rispettivamente, $r = 0{,}32$ e $r = 0{,}19$), con il dolore intermedio (rispettivamente, $r = 0{,}40$ e $r = 0{,}24$), con il dolore finale (rispettivamente, $r = 0{,}33$ e $r = 0{,}22$) e con il livello medio di dolore (rispettivamente, $r = 0{,}56$ e $r = 0{,}33$). La durata della colonoscopia non permetteva di prevedere in modo accurato come la procedura sarebbe stata ricordata dai pazienti (rispettivamente, $r = 0{,}10$ e $r = 0{,}09$) – un risultato che è stato osservato anche in analisi relative ai soli pazienti che non avevano ricevuto anestetici (rispettivamente, $r = 0{,}01$ e $r = 0{,}12$).

3.4. *Esecuzione di una successiva colonoscopia*

Il periodo di osservazione dopo l'esperimento è stato in media di 5,3 anni e circa la metà dei pazienti sono tornati a sottoporsi a una colonoscopia (tab. 5). L'intervento sperimentale non ha avuto un effetto ampio e generale nell'incrementare la proporzione di successive colonoscopie (53% *vs*. 48%, $p > 0{,}20$). La regressione logistica ha indicato che questo equivale a un incremento relativo del 18% del rapporto di probabilità in favore di una successiva colonoscopia (intervallo di fiducia al 95%: 13-59, $p > 0{,}20$). Tenendo conto della storia clinica dei pazienti, delle indicazioni specifiche e di esiti anomali nell'esame (i tre predittori risultati significativi in una regressione *step-wise*) si ottiene un incremento relativo del 41% del rapporto di probabilità in favore di una successiva colonoscopia (intervallo di fiducia al 95%: 2-96, $p = 0{,}038$) dovuto alla procedura colonoscopica modificata. Fra i pazienti che avevano «sanguinamento» come indicazione per la colonoscopia si è registrato il più ampio incremento delle proporzioni di una successiva colonoscopia (rapporto incrociato [*odds ratio*]: 3,01; intervallo di fiducia al 95%: 1,2-7,39, $p = 0{,}014$) indipendentemente da precedenti colonoscopie e da eventuali risultati anomali dell'esame.

Tre fattori clinici predicevano efficacemente una successiva colonoscopia. Sia l'esecuzione di precedenti esami, sia l'indicazione clinica che la procedura era «di controllo», sia, infine, la scoperta di condizioni anomale durante l'esame permettevano di prevedere un aumento nella probabilità di una successiva colonoscopia (tab. 5). Nessuna altra caratteristica demografica (tab. 1), nessun fattore relativo alla procedura (tab. 2) o all'intensità del dolore percepito in tempo reale (tab. 3) permetteva di predire in misura significativa la probabilità di una successiva colonoscopia. Come previsto, le valutazioni retrospettive sulla scala visiva e la classificazione rispetto ad altre esperienze spiacevoli (tab. 4) si sono rivelate correlate con una diminuzione delle probabilità di una successiva colonoscopia ($p < 0{,}02$ in entrambi i casi). Per esempio, coloro che hanno classificato la colonoscopia come la peggiore fra tutte

le esperienze spiacevoli considerate avevano una probabilità di ripeterla significativamente inferiore rispetto a coloro che l'avevano classificata come la migliore fra tutte le esperienze spiacevoli considerate (45% *vs*. 64%, $p = 0{,}016$).

*Tab. 5 - Esecuzione di una successiva colonoscopia**

	Procedura convenzionale	*Procedura modificata*	*Numero necessario per trattare*
Precedenti colonoscopie			
Sì	67	67	N/A**
No	32	42	10
Indicazione per la colonoscopia			
Dolore	32	39	15
Sangue nelle feci	36	63	4
Anemia	39	58	5
Modificazioni nella defecazione	33	38	21
Screening	49	49	N/A**
Controllo	73	73	N/A**
Altro***	59	47	–9
Risultati della colonoscopia			
Normali	35	44	12
Anomali***	60	63	33

* I dati riportati sono % di pazienti sottoposti a una successiva colonoscopia in ciascun gruppo.
** N/A indica che il valore è al di fuori dell'intervallo fra –200 e +200.
*** Categorie definite come in tab. 2.

Attraverso analisi stratificate, abbiamo esaminato i ricordi e il comportamento dei soggetti in relazione alla propensione a ripetere la colonoscopia. I pazienti che presentavano due o tre dei predittori significativi precedentemente indicati sono stati classificati come molto propensi a ripetere l'esame ($n = 256$). I pazienti che presentavano uno o nessun predittore sono stati classificati come poco propensi a ripetere l'esame ($n = 426$). Si è potuto riscontrare che l'entità dell'effetto dell'intervento sperimentale rispetto alle valutazioni retrospettive sulla scala visiva era simile fra i pazienti con un'alta propensione (3,9 *vs*. 4,7, $p = 0{,}010$) e fra quelli con una bassa propensione a ripetere l'esame (4,6 *vs*. 5,1, $p = 0{,}065$). Analogamente gli effetti dell'intervento sperimentale sulla classificazione retrospettiva rispetto ad altre esperienze spiacevoli sono risultati coerenti nei due sottogruppi. Infine, l'effetto sulla probabilità di una

successiva colonoscopia è stato trascurabile fra coloro che avevano un'alta propensione a ripetere l'esame (72% *vs.* 71%, $p > 0{,}20$) e moderato fra coloro che avevano una bassa propensione a ripetere l'esame (43% *vs.* 32%, $p = 0{,}023$).

4. Discussione

Nel presente studio abbiamo messo alla prova un metodo non farmacologico per modificare i ricordi che i pazienti conservano a seguito di una colonoscopia. Abbiamo riscontrato che la colonoscopia è stata in media spiacevole, che ha prodotto ricordi negativi a breve termine e che soltanto la metà circa dei pazienti è tornata a ripetere l'esame nei cinque anni successivi. In accordo con le nostre ipotesi iniziali, l'aggiunta di un breve intervallo di disagio minimo al termine della procedura ha indotto i pazienti a conservare un ricordo complessivamente più favorevole (meno negativo) dell'esperienza. Questo intervento ha prodotto una diminuzione relativa del 10% circa nel ricordo del dolore complessivo e un aumento relativo del 10% nel numero di pazienti che sono tornati a sottoporsi a un ulteriore controllo, suggerendo allo stesso tempo che nella pratica clinica sono necessari strumenti di intervento più efficaci.

Nel considerare i limiti del presente studio, va segnalato che si tratta di un'indagine relativa alla psicologia del ricordo. Di conseguenza, non sappiamo quali sarebbero i risultati del tentativo di considerare una gamma più estesa delle percezioni del paziente. La colonoscopia prevede l'eventualità di episodi potenzialmente umilianti e di vari altri possibili inconvenienti; pertanto, il ricordo non è l'unico fattore che può influenzare il comportamento del paziente. La ripetizione della colonoscopia non sempre è necessaria, non sappiamo che cosa sia stato detto a ciascun paziente a seguito della procedura, ed è stato possibile trovare differenze significative nella probabilità di un successivo esame solo dopo aver controllato l'influenza di importanti aspetti clinici (storia, indicazioni, anormalità). Lo studio di interventi più impegnativi rimane una prospettiva aperta per la ricerca futura; tuttavia, ulteriori indagini in questo senso dovranno affrontare serie difficoltà nel garantire l'occultamento dell'intervento sperimentale nei confronti dei pazienti. Per il momento, abbiamo potuto riscontrare che la metà circa dei pazienti sottoposti a colonoscopia non ritorna per ulteriori controlli nei cinque anni successivi e che, fra coloro che hanno una bassa propensione a farlo, l'intervento qui considerato induce i pazienti a sottoporsi a un esame successivo, che altrimenti avrebbero evitato, in un caso ogni nove individui trattati.

La fallibilità della memoria, persino entro intervalli di tempo ristretti, è documentata da decenni di ricerca psicologica. Ricerche precedenti indicano che le distorsioni del ricordo non sono del tutto casuali; al contrario, nella maggior parte delle persone ricorrono errori sistematici, che è possibile prevedere. I nostri risultati forniscono sostegno a ricerche precedenti secondo le

quali il ricordo complessivo di un episodio è costituito attraverso il richiamo selettivo di alcuni suoi specifici momenti piuttosto che ripercorrendo fedelmente l'esperienza nel suo insieme. La durata di un episodio ha un'influenza relativamente modesta sul ricordo, a meno che non sia molto saliente (per esempio, l'attesa per un intervento chirurgico) o correlata con l'intensità (per esempio, la durata di un lavoro faticoso). Le impressioni finali (*last*) possono spesso risultare le più durature (*lasting*) quando le persone riflettono sulle esperienze della loro vita passata.

In campo medico, i pazienti associano i loro ricordi alle esperienze passate e può talvolta essere opportuno rendere migliori tali ricordi, per il fatto che influenzeranno decisioni future o semplicemente perché sono spiacevoli in se stessi. Il presente studio suggerisce un modo in cui la ricerca psicologica di base può orientare la pratica clinica. Le possibili applicazioni e le future ricerche dovranno tenere conto del rapporto fra costi e benefici dei trattamenti, della disponibilità di strumenti tecnologici più avanzati e della tensione etica fra il riconoscimento della fallibilità del ragionamento umano e l'imperativo in favore dell'autonomia del paziente (Kahneman, 2000a,b,c; Schreiber, Kahneman, 2000). Tuttavia, nessuna di queste questioni specifiche autorizza a trascurare i principi della psicologia umana, dal momento che il punto di vista del paziente sarà sempre un elemento essenziale della pratica clinica.

I ricordi sono imperfetti e talvolta le cose sembrano migliori di come sono. I pazienti sottoposti a colonoscopia sono più inclini a ricordare l'esperienza come spiacevole se la procedura si conclude poco dopo il suo momento peggiore. I pazienti sottoposti a un esame simile, ma con una fase conclusiva prolungata che permetta al dolore di diminuire gradualmente, potrebbero conservare un ricordo più favorevole (meno sfavorevole) della procedura. Il medico dovrebbe quindi preoccuparsi di gestire con particolare delicatezza e attenzione le fasi conclusive di una procedura dolorosa effettuata su un paziente cosciente, soprattutto se il paziente non è stato sottoposto in precedenza alla stessa procedura e se l'esame non ha evidenziato condizioni cliniche anomale. Far ciò è spesso possibile, senza compromettere gli aspetti tecnici dell'intervento, e aumentando il numero di pazienti che si sottoporranno in futuro a ulteriori controlli.

Donald A. Redelmeier riceve finanziamenti per la ricerca da un Career Scientist Award del Ministero della Salute dello Stato dell'Ontario e dal de Souza Chair in Trauma presso l'Università di Toronto. Joel Katz riceve finanziamenti per la ricerca da un Investigator Award dei Canadian Institutes of Health Research. Daniel Kahneman riceve finanziamenti per la ricerca dallo Eugene Higgins Chair in Psychology presso l'Università di Princeton e dalla National Science Foundation. Questa indagine è stata finanziata con un contributo del National Cancer Institute del Canada. Ringraziamo Susan Campbell per l'assistenza amministrativa, Bill Redelmeier per la programmazione informatica e Chaim Bell, David Juurlink, Gabor Kandel, Micheal Schull, Matthew Stanbrook, John-Paul Szalai e Robert Tibshirani per i loro commenti su precedenti stesure di questo lavoro.

*15. Effetto di incorniciamento e protezione solare: i vantaggi di una presentazione in termini di guadagni**

di *Jerusha B. Detweiler, Brian T. Bedell, Peter Salovey, Emily Pronin, Alexander J. Rothman*

1. Introduzione

Il 40% circa di tutti i casi di cancro colpisce la pelle e approssimativamente un americano su cinque sviluppa un cancro della pelle durante la vita. Dal momento che l'esposizione a raggi solari ultravioletti è responsabile di più del 90% dei casi di cancro della pelle (Skin Cancer Foundation, 1995), è possibile prevenire la malattia con comportamenti in grado di ridurre gli effetti dell'esposizione solare, quali l'impiego di creme solari resistenti all'acqua con un fattore di protezione uguale o superiore a 15, la loro applicazione ripetuta e l'utilizzo di indumenti protettivi, come alcuni tipi di cappello (Council on Scientific Affairs, 1989). Nel presente studio ci concentreremo sull'uso di creme solari con un fattore di protezione uguale o superiore a 15. Molte osservazioni documentano l'efficacia dell'applicazione regolare di creme solari protettive. Thompson, Jolley, Marks (1993) hanno rilevato che individui di 40 anni o più che usavano quotidianamente creme solari con un fattore di protezione 17 nel corso di un'intera estate mostravano un minor numero di nuove lesioni cutanee e un maggior numero di regressioni di lesioni cutanee preesistenti rispetto agli individui assegnati a un gruppo di controllo. Inoltre l'uso costante di creme solari con un fattore di protezione uguale o superiore a 15 durante l'infanzia e l'adolescenza può ridurre fino al 78% l'incidenza di carcinomi delle cellule basali e squamose nell'arco della vita (Stern, Weinstein, Baker, 1986).

Nonostante l'importanza di comportamenti in grado di proteggere dall'esposizione solare, la loro attuazione non è uniformemente diffusa (si vedano, per esempio, Keesling, Friedman, 1987, e Wichstrom, 1994). Come

* Traduzione di *Message framing and sun-screen use: gain-framed messages motivate beach-goers*, «Health psychology», 18 (1999), pp. 189-196, copyright © 1999 American Psychological Association.

si possono persuadere gli individui a conformarsi con costanza a comportamenti protettivi? Un primo passo consisterebbe nel presentare messaggi il più possibile convincenti e adatti a sollecitare questo tipo di comportamenti. Le ricerche sulla cognizione sociale suggeriscono diverse strategie per mettere a punto messaggi persuasivi (si vedano Eagly, Chaiken, 1993, Petty, Wegener, 1998, Salovey, Rothman, Rodin, 1998). Nella presente indagine assumeremo la *prospect theory* (Tversky, Kahneman, 1981) come punto di riferimento per l'elaborazione di messaggi efficaci.

I principi della *prospect theory* prevedono che le persone rispondano diversamente alla presentazione delle informazioni a seconda del modo in cui tali informazioni sono «incorniciate» (*framed*)[1]. Le informazioni di carattere medico possono essere incorniciate sia in termini di possibili guadagni (vantaggi o benefici) sia in termini di possibili perdite (svantaggi o costi). Un esempio di presentazione in termini di guadagni è «se segui le indicazioni fornite dal Surgeon General incrementerai la tua probabilità di vivere a lungo e in salute». Per contro, una presentazione in termini di perdite potrebbe asserire: «se non segui le indicazioni del Surgeon General incrementerai la tua probabilità di morire prematuramente». La *prospect theory* suggerisce che le persone si mostrano avverse al rischio (tendono, cioè, a evitare alternative di azione rischiose) nei casi in cui i guadagni risultano salienti, ma sono propense al rischio (tendono, cioè, a scegliere alternative di azione rischiose) nei casi in cui risultano salienti le perdite. Anche quando le informazioni presentate sono di fatto equivalenti, l'inclinazione a correre dei rischi per promuovere un esito desiderato o per evitare un esito indesiderato dipende del modo in cui il messaggio viene incorniciato (Kahneman, Tversky, 1979, 1982 e 1984).

Una delle variabili che hanno contribuito a chiarire l'influenza dell'incorniciamento delle informazioni sui comportamenti rilevanti per la salute è il tipo di comportamento che si intende promuovere – se un'azione preventiva o di controllo (Rothman, Salovey, 1997; Salovey, Rothman, Rodin, 1998). Una differenza decisiva fra un'azione preventiva e una di controllo risiede nel livello di rischio imminente percepito. Azioni di controllo come la mammografia o il PAP test vengono generalmente percepite come rischiose nel momento in cui le si mette in atto, perché c'è la possibilità che si scopra che qualcosa non va nelle proprie condizioni di salute (Banks *et al.*, 1995; Meyerowitz, Chaiken, 1987). La *prospect theory* suggerisce che le persone tendono a preferire opzioni rischiose quando si trovano a considerare possibili perdite; l'inclinazione a intraprendere azioni di controllo dovrebbe quindi essere favorita dalla presentazione delle informazioni rilevanti in termini di possibili perdite. Al contrario, le azioni preventive sono normalmente percepite come meno ri-

1. Si vedano in proposito, in questo volume, Introduzione e cap. 7 [N.d.C.].

schiose degli interventi di controllo. Per esempio, l'impiego di creme protettive riduce direttamente il rischio futuro di cancro, comportando rischi immediati nulli o minimi. Dal punto di vista della *prospect theory*, la messa in atto di un comportamento preventivo dovrebbe quindi corrispondere a un attitudine di avversione al rischio, e gli individui tendono a preferire opzioni sicure o poco rischiose quando si trovano a considerare i possibili guadagni o benefici a partire da un punto di riferimento identificato come neutrale. Ci si dovrebbe quindi aspettare che l'inclinazione a intraprendere azioni preventive sia favorita dalla presentazione delle informazioni rilevanti in termini di possibili guadagni. In studi precedenti abbiamo indagato l'impatto dell'incorniciamento delle informazioni sull'esecuzione di azioni di controllo (si veda, per esempio, Banks *et al.*, 1995). Nel presente studio ci concentreremo sulla protezione dall'esposizione solare, vale a dire un comportamento preventivo.

Un primo risultato a sostegno di un vantaggio delle presentazioni in termini di guadagni nel promuovere azioni preventive è stato ottenuto da Rothman *et al.* (1993) in uno studio su un campione di studenti universitari riguardo agli effetti di differenti presentazioni sull'intenzione di proteggersi dall'esposizione solare. Le donne cui erano state presentate informazioni incorniciate in termini di guadagni avevano una maggiore probabilità di richiedere una crema solare con un fattore di protezione appropriato rispetto a quelle che avevano letto un opuscolo in cui le informazioni erano state presentate in termini di perdite. In questo studio, però, gli uomini non hanno reagito in modo significativamente differente alle due diverse presentazioni. L'obiettivo primario della presente indagine è estendere i risultati di Rothman *et al.* (1993) a una popolazione più eterogenea e in uno studio sul campo. Più specificamente, intendiamo controllare se il vantaggio delle presentazioni in termini di guadagni in relazione a comportamenti per la prevenzione del cancro alla pelle, osservato fra le studentesse universitarie, possa essere replicato in individui di entrambi i sessi in un contesto in cui tutti dovrebbero essere particolarmente sensibili al problema: la spiaggia. Ci aspettiamo un risultato più significativo fra le persone che si recano in spiaggia senza l'intenzione di utilizzare creme protettive. Ipotizziamo infatti che la probabilità di modificare il comportamento di questi individui meno «preparati» sia maggiore, in particolare rispetto a coloro che si recano in spiaggia con la loro crema protettiva e chiaramente intenzionati a usarla.

Il nostro secondo obiettivo è esplorare alcuni possibili fattori all'opera nella costruzione di rappresentazioni in termini di guadagni e in termini di perdite. Alcuni studiosi hanno osservato che una presentazione in termini di guadagni può essere costruita in diversi modi, e lo stesso vale per una presentazione in termini di perdite. In particolare, Petty, Wegener (1991) hanno rilevato che una presentazione persuasiva può indicare una conseguenza positiva che probabilmente si verificherà se il comportamento raccomandato verrà

adottato, oppure indicare una conseguenza negativa che, adottando quello stesso comportamento, si riuscirà probabilmente a evitare. In alternativa, una presentazione persuasiva potrebbe sottolineare una conseguenza negativa che probabilmente si verificherà se il comportamento raccomandato non verrà adottato o una conseguenza positiva che, non adottando quel comportamento, è improbabile che si verifichi. Brendl, Higgins, Lemm (1995) hanno avanzato osservazioni analoghe, suggerendo che vi sono quattro modi per descrivere le possibili conseguenze di una scelta: come un guadagno, una perdita, un mancato guadagno o una perdita evitata.

Nel presente studio faremo riferimento alla descrizione di queste quattro forme di presentazione offerta da Rothman, Salovey (1997), che si fonda sull'identificazione di due diversi attributi: il tipo di conseguenza di un dato comportamento (avverarsi *vs*. non avverarsi di un esito) e le caratteristiche dell'esito considerato (desiderabile *vs*. indesiderabile). Le presentazioni in termini di guadagni possono sottolineare l'eventualità che si avveri un esito desiderabile o che non si avveri un esito indesiderabile. Per esempio, una presentazione in termini di guadagni potrebbe asserire: «l'uso di creme solari protettive ti aiuta a conservare la pelle in salute»; oppure: «l'uso di creme solari protettive diminuisce il tuo rischio di sviluppare un cancro della pelle». Le presentazioni in termini di perdite possono a loro volta assumere due forme: possono indicare l'eventualità che si verifichi un esito indesiderabile o che non si verifichi un esito desiderabile. Da questo punto di vista, sia «se non usi creme protettive aumenti il tuo rischio di sviluppare un cancro della pelle» sia «se non usi creme protettive non puoi difendere la salute della tua pelle» costituiscono presentazioni in termini di perdite. Il nostro interesse per una esplicita identificazione delle dimensioni che determinano i diversi tipi di presentazione (guadagni *vs*. perdite) è dovuto al sospetto che in precedenti studi l'operazionalizzazione di queste nozioni possa aver risentito di confusioni fra i diversi modi in cui è possibile incorniciare le informazioni sia in termini di guadagni sia in termini di perdite.

Alcuni studiosi hanno sostenuto che ciascuno dei quattro diversi modi di presentazione ha un impatto differente (si vedano, per esempio, Higgins *et al*., 1994, e Higgins, Tykocinski, 1992), mentre altri ritengono che sia soltanto la distinzione fra presentazioni in termini di guadagni e di perdite a svolgere un ruolo critico (si veda, per esempio, Petty, Wegener, 1991). Noi ci riconosciamo nella seconda posizione, ma crediamo che le due dimensioni relative al tipo di conseguenza del comportamento e alle caratteristiche dell'esito considerato dovrebbero essere indagate e, comunque, tenute sotto controllo nell'elaborazione delle procedure sperimentali. Per questa ragione, nella presente ricerca indagheremo se l'impatto di informazioni presentate in termini di guadagni o di perdite varia in funzione di queste dimensioni, confrontando l'efficacia di diverse forme di presentazione nel persuadere i bagnanti a uti-

lizzare creme solari protettive. Prevediamo che, in confronto alle due presentazioni in termini di perdite, entrambe le presentazioni in termini di guadagni favoriscano l'intenzione di far uso di creme protettive e inducano un maggior numero di partecipanti a richiederne un campione gratuito. Ci aspettiamo inoltre di non riscontrare differenze significative né fra le due presentazioni in termini di guadagni né fra le due presentazioni in termini di perdite.

2. Metodi

2.1. *Partecipanti*

Duecentodiciassette bagnanti di 18 anni o più sono stati reclutati durante il mese di agosto in una spiaggia pubblica del sud del New England frequentata da una popolazione in cui sono rappresentate età, etnie e condizioni socioeconomiche differenti, con una maggioranza di bianchi di classe media. Il campione rifletteva adeguatamente questa varietà. La maggioranza dei partecipanti erano donne (76% *vs*. 24% di uomini). L'età variava fra i 18 e i 79 anni, con una media di 38,7. Il 90% circa delle persone interpellate ha acconsentito a partecipare e più del 95% degli opuscoli distribuiti sono stati compilati e restituiti. Il processo di raccolta dei dati non ha permesso di tenere un conteggio dettagliato degli opuscoli distribuiti. Le percentuali qui riportate sono quindi stime.

2.2. *Opuscoli*

A ciascuno dei partecipanti è stato consegnato un opuscolo stampato in un formato standard e ripiegato in tre ante. Il titolo era *Beach Survey 1996*. Sulla prima pagina esterna si trovavano alcune istruzioni e domande preliminari. Dopo aver risposto a queste domande preliminari, aprendo l'opuscolo, i partecipanti leggevano alcune informazioni generali riguardo al cancro della pelle e una serie di asserzioni relative alla protezione solare. Queste asserzioni comparivano in una fra quattro diverse forme di presentazione, corrispondenti alle combinazioni delle due dimensioni che determinano il possibile incorniciamento di uno stesso messaggio: vale a dire, nei termini (a) dei possibili benefici della protezione dall'esposizione solare, (b) degli esiti indesiderati che si possono evitare proteggendosi dall'esposizione solare, (c) dei mancati benefici dell'esposizione solare senza protezione, e (d) degli esiti indesiderati in cui si può incorrere non proteggendosi dall'esposizione solare. La fig. 1 riporta alcuni esempi delle asserzioni contenute in ciascuna delle quattro varianti.

Fig. 1 - I contenuti critici degli opuscoli: messaggi incorniciati in quattro forme diverse

Guadagni: avverarsi di un esito desiderabile	*Guadagni: non avverarsi di un esito indesiderabile*	*Perdite: avverarsi di un esito indesiderabile*	*Perdite: non avverarsi di un esito desiderabile*
Proteggerti dal sole ti aiuterà a restare in salute	Proteggerti dal sole non metterà a rischio la tua salute	Esporti al sole senza protezione metterà a rischio la tua salute	Esporti al sole senza protezione non ti aiuterà a restare in salute
Se utilizzi una crema solare con un fattore di protezione uguale o superiore a 15, aumenti le tue probabilità di mantenere la tua pelle sana e vivere vivere a lungo	Se utilizzi una crema solare con un fattore di protezione uguale o superiore a 15, diminuisci le tue probabilità di danneggiare la tua pelle e morire prematuramente	Se non utilizzi una crema solare con un fattore di protezione uguale o superiore a 15, aumenti le tue probabilità di danneggiare la tua pelle e morire prematuramente	Se non utilizzi una crema solare con un fattore di protezione uguale o superiore a 15, diminuisci le tue probabilità di mantenere la tua pelle sana e vivere a lungo
Usare una crema protettiva aumenta le tue probabilità di mantenere una pelle sana e giovane	Usare una crema protettiva diminuisce le tue probabilità di sviluppare il cancro e il prematuro invecchiamento della pelle	Non usare una crema protettiva aumenta le tue probabilità di sviluppare il cancro e il prematuro invecchiamento della pelle	Non usare una crema protettiva diminuisce le tue probabilità di mantenere una pelle sana e giovane
Più è alto il fattore di protezione, più sarai protetto dai raggi solari	Più è alto il fattore di protezione, meno i raggi solari ti nuoceranno	Più è basso il fattore di protezione, più i raggi solari ti nuoceranno	Più è basso il fattore di protezione, meno sarai protetto dai raggi solari
Proteggerti dal sole è il modo più sicuro per prevenire il cancro della pelle	Proteggerti dal sole è il modo più sicuro per evitare di sviluppare il cancro della pelle	Non proteggerti dal sole è il modo più sicuro per sviluppare il cancro della pelle	Non proteggerti da sole è il maggiore impedimento nel prevenire il cancro della pelle
In questo preciso momento ti trovi esposto al sole – ti stai proteggendo, assicurandoti così che la tua pelle resti sana?	In questo preciso momento ti trovi esposto al sole – ti stai proteggendo, evitando così danni alla tua pelle?	In questo preciso momento ti trovi esposto al sole – non ti stai proteggendo, provocando così danni alla tua pelle?	In questo preciso momento ti trovi esposto al sole – non ti stai proteggendo, non assicurandoti così che la tua pelle resti sana?

La prima parte dell'opuscolo conteneva anche alcune informazioni (il cui modo di presentazione non variava fra i diversi partecipanti) riguardo alla prevalenza e alla severità del cancro della pelle e del melanoma; conteneva inoltre le raccomandazioni per la protezione dall'esposizione solare stabilite dalla American Cancer Society.

Per passare alla parte finale dell'opuscolo, i partecipanti dovevano sbloccare un sigillo che la teneva separata dal resto. La parte finale conteneva alcune domande conclusive e alcune instruzioni che ricordavano ai partecipanti di riconsegnare l'opuscolo compilato per ricevere gratuitamente un biglietto della lotteria.

2.3. *Misure*

Domande precedenti la manipolazione sperimentale. Tre blocchi di domande permettevano di valutare le intenzioni iniziali dei partecipanti.

(1) *Intenzioni iniziali di impiego di creme protettive*. I partecipanti indicavano, innanzi tutto, se per quel giorno avevano previsto di usare creme solari protettive, rispondendo «sì» o «no».

(2) *Fattore di protezione previsto*. Se i partecipanti avevano previsto di usare creme solari protettive, si richiedeva loro di indicare quale fattore di protezione avevano previsto di impiegare. Nel caso che non intendessero usare creme solari protettive, il fattore di protezione solare è stato identificato con «0». Le risposte sono state codificate in modo dicotomico a seconda che il fattore di protezione che si prevedeva di impiegare fosse o non fosse conforme a quello raccomandato dalla American Cancer Society (uguale o superiore a 15).

(3) *Livello di rischio per il cancro della pelle*. Abbiamo incluso tre domande utili a determinare il livello di rischio oggettivo e soggettivo di cancro della pelle: si chiedeva di indicare il colore naturale dei capelli e della pelle (che sono stati combinati in un singolo indice continuo di rischio oggettivo) e il rischio percepito di sviluppare il cancro della pelle (le risposte sono state codificate in modo dicotomico: livello di rischio percepito «alto» *vs*. «basso»).

Domande successive alla manipolazione sperimentale. Vi erano quattro blocchi di domande conclusive.

(1) *Reazioni emotive immediate*. Si chiedeva ai partecipanti di indicare il loro livello di ansia a seguito della lettura dell'opuscolo su una scala discreta da «1» («non sono affatto in ansia») a «7» («sono estremamente in ansia»). Si chiedeva inoltre di indicare quanta paura avessero di incorrere in un cancro della pelle, in un prematuro invecchiamento della pelle o in entrambi gli eventi su una scala discreta che variava fra «1» («non ne ho affatto paura») e

«7» («ne ho una paura estrema»). Abbiamo calcolato un indice delle reazioni emotive immediate attraverso una media fra le due risposte per ciascun partecipante ($r(213) = 0{,}62$, $p < 0{,}0001$). Questa misura è stata impiegata per controllare se i messaggi presentati in termini di perdite avevano sollecitato reazioni di paura e di ansia, influenzando gli effetti delle diverse presentazioni (Higbee, 1969; Leventhal, 1970; Robbertson, Rogers, 1988).

(2) *Efficacia percepita di comportamenti protettivi*. Abbiamo incluso due domande sulle credenze dei partecipanti riguardo ai comportamenti per la protezione dall'esposizione solare. Nella prima domanda si chiedeva di giudicare in che modo la probabilità di sviluppare un cancro della pelle cambia «se ti proteggi dall'esposizione solare». Nella seconda domanda si chiedeva invece: «se *non* ti proteggi dall'esposizione solare, in che modo cambia la tua probabilità di sviluppare un cancro della pelle?». In entrambi i casi le risposte venivano espresse su una scala discreta che variava fra «1» («resta identica») a «7» («aumenta/diminuisce in modo drammatico»). Abbiamo definito un indice dell'efficacia percepita di comportamenti protettivi calcolando la media fra queste due risposte per ciascun partecipante ($r(210) = 0{,}56$, $p < 0{,}0001$).

(3) *Anticipazione delle reazioni emotive*. Abbiamo incluso quattro domande per valutare le previsioni dei partecipanti sul modo in cui si sarebbero sentiti utilizzando o non utilizzando creme solari protettive. Le prime due domande riguardavano la previsione delle loro sensazioni dopo avere applicato una crema protettiva: si chiedeva ai partecipanti quanto si sarebbero sentiti «sollevati» e quanto si sarebbero sentiti «infastiditi» (su una scala discreta a 7 punti che variava da «per nulla» a «estremamente»). La seconda coppia di domande riguardava le reazioni dei partecipanti alla scoperta di un impedimento all'utilizzo di una crema protettiva (per esempio: la crema era stata dimenticata a casa, o il negozio aveva esaurito il prodotto). Si chiedeva ai partecipanti quanto si sarebbero sentiti «dispiaciuti» e quanto si sarebbero sentiti «ansiosi» in questa situazione (ancora una volta, su una scala discreta a 7 punti che variava da «per nulla» a «estremamente»). Abbiamo definito un indice dell'anticipazione delle reazioni emotive invertendo il valore della risposta dei partecipanti su quanto si sarebbero sentiti «infastiditi» e calcolando la media delle quattro risposte per ciascun partecipante (α di Cronbach = 0,76).

(4) *Intenzioni finali di impiego di creme protettive*. Abbiamo misurato in due modi le intenzioni relative all'uso di creme protettive a seguito della lettura degli opuscoli: (a) abbiamo innanzi tutto considerato le intenzioni riportate dai partecipanti di utilizzare in futuro creme protettive; (b) ci siamo inoltre serviti, come indicatore delle intenzioni dei partecipanti, di una misura comportamentale: la richiesta di un campione gratuito di crema solare protettiva.

(a) *Intenzioni riportate dai partecipanti*. Le intenzioni finali dei partecipanti riguardo all'impiego di creme solari protettive sono state misurate attra-

verso quattro domande. Si chiedeva innanzi tutto ai partecipanti di indicare con quale frequenza erano intenzionati a utilizzare creme solari protettive in spiaggia per quella estate. In secondo luogo, si chiedeva loro di indicare con quale frequenza erano intenzionati a impiegare creme solari protettive nelle loro attività quotidiane. Le risposte venivano espresse su una scala discreta che variava fra «1» («mai») e «7» («sempre»). Si chiedeva quindi ai partecipanti di indicare quante volte avrebbero applicato una crema solare protettiva nel corso di un'intera giornata in spiaggia (dalle ore 10 alle 16). Infine, si chiedeva di indicare il tipo di prodotto per la pelle che i partecipanti intendevano impiegare per quella estate («nessuno», «olio abbronzante», «crema protettiva»); se la risposta era «crema protettiva», si doveva specificare quale fattore di protezione si prevedeva di utilizzare. Le risposte a quest'ultima domanda sono state codificate in modo dicotomico a seconda che si prevedesse di utilizzare un fattore di protezione uguale/superiore a 15 oppure inferiore[2].

b) *Misura comportamentale*. Dopo aver completato il questionario, i partecipanti ricevevano un buono con il quale, in un momento successivo della giornata, avrebbero potuto ritirare un campione gratuito di crema solare protettiva. L'effettivo o il mancato ritiro del campione costituiva la principale misura comportamentale dello studio.

2.4. *Procedura*

Due degli autori hanno avvicinato i bagnanti e hanno chiesto loro se intendevano partecipare a uno studio riguardo alla «frequentazione delle spiagge». I bagnanti contattati erano informati che in cambio della loro partecipazione (la lettura di un sintetico opuscolo e la compilazione di un breve questionario) avrebbero ricevuto gratuitamente un biglietto della lotteria dello stato del Connecticut. Ai partecipanti veniva detto che avrebbero ricevuto il biglietto riconsegnando l'opuscolo agli sperimentatori presso una postazione collocata nelle vicinanze. I primi quattro gruppi che hanno accettato di partecipare (persone che si trovavano in spiaggia insieme, oppure individui singoli) sono stati assegnati alle quattro diverse condizioni sperimentali e i successivi gruppi sono stati rispettivamente assegnati alle quattro condizioni in modo ciclico.

2. Fra le domande precedenti la manipolazione sperimentale ve n'erano anche tre per la valutazione del *need for cognition* (Cacioppo, Petty, 1982) e una relativa alle credenze soggettive dei partecipanti riguardo agli ostacoli all'uso di creme solari protettive (in particolare, ai costi percepiti associati a tale uso). Erano inoltre incluse una domanda relativa al tono del messaggio presentato e una sui precedenti contatti con persone affette da cancro della pelle. Nessuna di queste variabili è stata inclusa nelle analisi qui presentate. In questo articolo non verranno quindi ulteriormente discusse.

Dal momento che gli sperimentatori hanno contattato tutti gli individui in condizione di partecipare e che non erano a conoscenza di quale fosse la specifica condizione sperimentale corrispondente all'opuscolo che stavano distribuendo, la procedura adottata produce gli stessi effetti di un'assegnazione casuale dei partecipanti ai quattro gruppi.

Una volta consegnati i questionari presso la postazione prestabilita, i partecipanti ricevevano un buono per un campione gratuito di crema solare con un fattore di protezione 15 e venivano informati che avrebbero potuto utilizzare il buono presentandosi presso una diversa postazione circa mezz'ora più tardi (veniva indicata un'ora precisa). Si diceva ai partecipanti che ciascuno poteva presentare il buono che aveva personalmente ricevuto. Un numero identificativo di ciascun partecipante veniva copiato dal questionario sul buono omaggio corrispondente, in modo che fosse possibile individuare i partecipanti che avevano ritirato il campione gratuito. I partecipanti che utilizzavano il loro buono ricevevano un piccola confezione di crema solare protettiva da uno sperimentatore, che si metteva a loro disposizione per eventuali domande relative allo studio.

3. Risultati

3.1. *Varianti delle presentazioni in termini di guadagni e di perdite*

Non si sono registrati effetti significativi riconducibili al tipo di conseguenze (avverarsi *vs*. non avverarsi di un esito) e alle caratteristiche degli esiti considerati (desiderabile *vs*. non desiderabile) né rispetto alla richiesta del campione di crema protettiva né rispetto alle altre variabili associate alle intenzioni finali dei partecipanti. I nostri risultati mostrano invece un'interazione incrociata fra queste due dimensioni, che rappresenta un effetto semplice dovuto alla presentazione in termini di guadagni o in termini di perdite. Ciò significa che le due varianti della presentazione in termini di guadagni (avverarsi di un esito desiderabile, non avverarsi di un esito non desiderabile) hanno avuto effetti simili, così come le due varianti della presentazione in termini di perdite (avverarsi di un esito non desiderabile, non avverarsi di un esito desiderabile). Abbiamo quindi aggregato i dati in modo da concentrarci sul confronto fra incorniciamento in termini di guadagni e in termini di perdite. Analogamente, poiché non si è osservata alcuna interazione fra genere e incorniciamento rispetto a nessuna delle variabili dipendenti, abbiamo omesso la considerazione del genere dalle analisi riportate di seguito.

3.2. *Effetti dell'incorniciamento*

Nella nostra analisi principale abbiamo considerato la percentuale dei bagnanti che hanno riconsegnato il loro buono per ricevere il campione gratuito di crema solare protettiva. Nelle due condizioni di incorniciamento in termini di guadagni, il 71% dei partecipanti ha presentato il buono omaggio, mentre solo il 53% lo ha fatto fra i partecipanti nelle due condizioni di incorniciamento in termini di perdite. In una regressione logistica, la considerazione dell'incorniciamento ha prodotto un accordo significativamente migliore con il comportamento dei partecipanti in confronto al solo rischio oggettivo e soggettivo di cancro della pelle: Wald, χ^2 (1, n = 217) = 7,37, p < 0,01. Tale effetto è inoltre rimasto significativo anche dopo che erano state incluse fra i fattori considerati le intenzioni iniziali di utilizzare creme solari protettive durante la giornata e le intenzioni iniziali di usare un fattore di protezione uguale o superiore a 15: Wald, χ^2 (1, n = 217) = 6,66, p < 0,01 (b = 0,75, SE = 0,29, rapporto incrociato [*odds ratio*] = 2,11). Questi risultati documentano il previsto vantaggio di una presentazione in termini di guadagni nel promuovere un importante comportamento volto alla protezione dall'esposizione solare, vale a dire l'impiego di creme protettive[3].

I risultati relativi alle intenzioni riportate dai partecipanti di usare creme protettive in seguito alla lettura degli opuscoli sono presentati in tab. 1. La regressione lineare ha rivelato che il rischio di cancro della pelle era correlato sia con l'intenzione di usare creme protettive in spiaggia sia con l'intenzione di applicarle ripetutamente nel corso della giornata. In entrambi i casi, il vantaggio della presentazione in termini di guadagni emerge attraverso l'interazione fra il tipo di presentazione e le intenzioni iniziali dei partecipanti. L'interazione osservata è conforme alle nostre predizioni. In particolare, si è rilevato un vantaggio significativo della presentazione in termini di guadagni nel promuovere le intenzioni di quei bagnanti che non erano inizialmente intenzionati a utilizzare creme protettive. I partecipanti che si erano recati in spiaggia con la chiara intenzione di utilizzare creme protettive non sono stati invece influenzati in modo differente da messaggi formulati in termini di

3. Anche adottando un approccio particolarmente cauto e assumendo una completa dipendenza fra i membri dei gruppi (dal momento che abbiamo assegnato tutti i membri di uno stesso gruppo alla medesima condizione, in modo da separare il più possibile fra loro le manipolazioni sperimentali), il controllo dell'effetto principale dovuto all'incorniciamento sulle richieste di campioni di crema solare (d di Cohen = 0,36) con 108 gradi di libertà (217 partecipanti diviso per 2, il numero medio stimato di partecipanti in uno stesso gruppo) risulta ancora marginalmente significativo (p < 0,10, due code). Lo stesso vale per l'interazione fra incorniciamento e intenzioni iniziali rispetto alle intenzioni di applicare ripetutamente creme solari protettive (d di Cohen = 0,36; p < 0,10, due code).

guadagni o di perdite (ciò potrebbe essere dovuto a un «effetto soffitto»: fra questi soggetti la media delle intenzioni iniziali di utilizzare creme protettive era in entrambe le condizioni di 6,33 su una scala da 1 a 7). Il diverso incorniciamento ha quindi influenzato soprattutto quei bagnanti che era più importante convincere, cioè quelli che per quel giorno non avevano previsto di utilizzare creme protettive. È interessante notare che questo effetto non sembra estendersi alle intenzioni di usare creme protettive in attività quotidiane al di fuori della spiaggia, forse perché le informazioni presentate negli opuscoli non erano sufficientemente salienti o specifiche per avere un impatto in questo contesto.

Tab. 1 - Le intenzioni finali di far uso di creme protettive e i potenziali mediatori dell'effetto di incorniciamento: medie aggiustate ed errori standard (SE)

	Intenzione iniziale				*No intenzione iniziale*				
	Guadagni (n = 83)		*Perdite (n = 75)*		*Guadagni (n = 26)*		*Perdite (n = 33)*		p^b
	Media	*SE*	*Media*	*SE*	*Media*[a]	*SE*	*Media*	*SE*	
Intenzioni finali (utilizzo di creme protettive)									
In spiaggia	6,33	0,13	6,33	0,14	3,84**	0,23	3,18	0,21	0,07
In attività quotidiane	3,72	0,19	3,69	0,20	1,94	0,35	2,17	0,35	*n.s.*
N. applicazioni/giorno in spiaggia	2,80	0,16	2,99	0,17	2,47**	0,28	1,50	0,26	< 0,01
Potenziali mediatori									
Reazioni emotive negative	3,62	0,16	3,60	0,17	3,79	0,29	3,50	0,26	*n.s.*
Efficacia percepita	6,01	0,15	6,04	0,16	5,53	0,26	5,18	0,24	*n.s.*
Anticipazione reazioni emotive[c]	5,33	0,14	5,34	0,15	4,21*	0,25	3,57	0,22	0,09

Nota. Con l'eccezione del numero di applicazioni al giorno in spiaggia, tutte le misure si riferiscono a una scala da 1 a 7. Le medie sono aggiustate per il rischio oggettivo e soggettivo di cancro della pelle. Valori più alti indicano una maggiore forza delle intenzioni, delle reazioni emotive e delle percezioni di efficacia.
[a] Test per l'effetto semplice dell'incorniciamento (all'interno del gruppo senza intenzioni iniziali di usare creme protettive): * $p < 0,06$, ** $p < 0,05$.
[b] Test incorniciamento × intenzione iniziale (1 e 210 gradi di libertà).
[c] Anticipazione reazioni emotive = previsioni dei partecipanti di una reazione emotiva positiva a seguito dell'applicazione di una crema protettiva e di una reazione emotiva negativa a seguito della mancata applicazione di una crema protettiva.

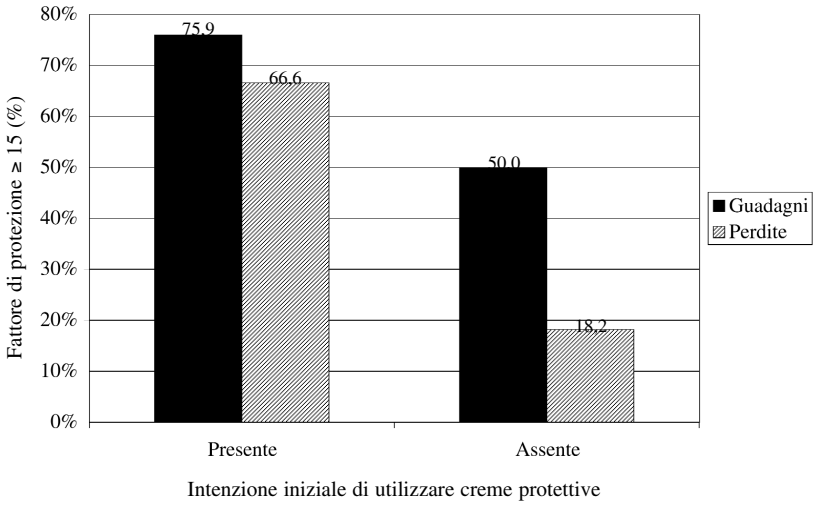
Abbiamo analizzato le intenzioni finali di utilizzare un fattore di protezione uguale o superiore a 15, secondo le raccomandazioni della American Cancer Society che erano state evidenziate negli opuscoli. I risultati della regressione logistica riguardo a questa variabile sono presentati in tab. 2.

Tab. 2 - Regressione logistica simultanea per la previsione delle intenzioni finali dei partecipanti di impiegare un fattore di protezione solare uguale o superiore a 15

Variabile predittiva	*b*	*SE*	*Rapporto incrociato* (odds ratio)	*p*
Intenzione iniziale (protezione ≥ 15)	3,65	0,53	38,39	< 0,001
Rischio oggettivo	0,55	0,28	1,74	< 0,05
Rischio soggettivo	0,19	0,20	1,20	*n.s.*
Incorniciamento	1,41	0,64	4,11	0,03
Intenzione iniziale (uso di creme protettive)	0,93	0,60	2,55	0,12
Incorniciamento × intenzione iniziale (uso di creme protettive)	1,49	0,82	4,42	0,07

Tenendo conto del rischio di cancro della pelle, si è rilevato un significativo vantaggio della presentazione in termini di guadagni nel promuovere l'intenzione di utilizzare creme solari con un adeguato fattore di protezione. Come osservato in relazione alle altre variabili associate alle intenzioni finali dei partecipanti, la natura di questo effetto si precisa alla luce dell'interazione fra il tipo di incorniciamento e le intenzioni iniziali dei partecipanti di usare creme protettive per quel giorno (si veda fig. 2). I messaggi incorniciati in termini di guadagni sono stati infatti particolarmente efficaci fra i bagnanti che per quel giorno non avevano previsto di usare creme protettive. In una regressione logistica separata, l'effetto dell'incorniciamento in questo gruppo di parte-

Fig. 2 - Intenzioni finali di far uso di creme solari con fattore di protezione uguale o superiore a 15



cipanti si è rivelato affidabile: Wald, χ^2 (1, n = 59) = 4.86, p < 0,03 (b = 1.37, SE = 0,62, rapporto incrociato = 3,93).

In conclusione, si è osservato un chiaro vantaggio della presentazione in termini di guadagni per tre delle nostre cinque variabili dipendenti: la richiesta di un campione gratuito di crema solare protettiva, l'intenzione di applicare ripetutamente durante la giornata creme solari protettive e l'intenzione di utilizzare un adeguato fattore di protezione. Per quanto riguarda le intenzioni finali riportate dai soggetti, l'effetto riscontrato è qualificato da una stabile interazione fra il tipo di incorniciamento e le intenzioni iniziali di usare creme protettive, di modo che il vantaggio della presentazione in termini di guadagni è risultato più forte fra i bagnanti che per quel giorno non avevano previsto di utilizzare creme solari. Per una quarta variabile dipendente, cioè l'intenzione di usare creme solari in spiaggia, questa stessa tendenza si è rivelata marginalmente significativa.

3.3. *Potenziali mediatori dell'effetto di incorniciamento*

Si sa poco sui possibili fattori che mediano gli effetti differenti delle presentazioni in termini di guadagni e in termini di perdite (Wilson, Purdon, Wallston, 1988; Wegener, Petty, Klein, 1994). Abbiamo incluso tre variabili nel tentativo di investigare questo aspetto, ma i risultati sono stati modesti (si veda tab. 1). Sebbene le reazioni emotive immediate dei partecipanti alla lettura degli opuscoli, i loro giudizi riguardo all'efficacia di comportamenti protettivi e le loro predizioni del modo in cui si sarebbero sentiti applicando o, in alternativa, non potendo applicare creme protettive fossero generalmente correlati con le intenzioni finali di utilizzare creme protettive (medie degli indici di correlazione: r = +0,17, +0,20 e +0,45, rispettivamente), il tipo di incorniciamento ha influito solo sull'anticipazione delle reazioni emotive e, comunque, in modo marginale. È emersa anche in questo caso la caratteristica interazione fra il tipo di incorniciamento e le intenzioni iniziali di impiegare creme protettive. Fra i partecipanti che non erano inizialmente intenzionati a usare creme protettive, coloro che avevano letto una presentazione in termini di guadagni predicevano che si sarebbero sentiti meglio utilizzando creme protettive e che si sarebbero sentiti peggio non potendole utilizzare in confronto a coloro che avevano letto presentazioni in termini di perdite. Questa interazione sostiene l'ipotesi di un vantaggio della presentazione in termini di guadagni fra i partecipanti che quel giorno non avevano previsto di usare creme protettive. Tuttavia, l'effetto osservato si è rivelato molto più debole di quello sulle intenzioni dei partecipanti a seguito della lettura degli opuscoli. Abbiamo quindi ritenuto di non approfondire ulteriormente lo studio dei possibili mediatori dell'effetto di incorniciamento.

4. Discussione

Dal momento che gli interventi preventivi sono percepiti come più sicuri, nelle loro conseguenze, rispetto agli interventi che mirano all'identificazione tempestiva di possibili problemi clinici, abbiamo ipotizzato che le presentazioni in termini di guadagni siano più efficaci nel persuadere le persone a metterli in atto (Rothman, Salovey, 1997). Nell'esperimento qui presentato, abbiamo previsto che i bagnanti sarebbero stati maggiormente interessati a richiedere un campione gratuito dopo aver letto un opuscolo informativo che presentava i vantaggi associati all'uso di creme solari protettive.

Uno dei nostri obiettivi era replicare i risultati di Rothman *et al.* (1993), che avevano rilevato come in un contesto specifico (una classe universitaria) la presentazione in termini di guadagni determinasse un numero maggiore di richieste di campioni di creme protettive rispetto alla presentazione in termini di perdite. I risultati del presente esperimento suggeriscono che il vantaggio di presentazioni in termini di guadagni è robusto e si manifesta sia fra gli uomini sia fra le donne che si recano in spiaggia. Abbiamo infatti osservato un vantaggio dei due tipi di presentazioni in termini di guadagni in confronto ai due tipi di presentazione in termini di perdite nell'indurre fra i bagnanti la richiesta di un campione gratuito di crema protettiva, nel sollecitare l'intenzione espressa di utilizzare ripetutamente creme protettive durante il giorno e di far uso di un fattore di protezione uguale o superiore a 15. Questi effetti si sono inoltre rivelati più forti fra i partecipanti che non si erano recati in spiaggia con l'intenzione iniziale di utilizzare creme protettive.

Il secondo scopo dello studio era determinare se diverse operazionalizzazioni delle nozioni di guadagno e di perdita fossero funzionalmente equivalenti: non abbiamo riscontrato differenze significative fra diversi tipi di presentazioni in termini di guadagni e fra diverse presentazioni in termini di perdite rispetto a nessuna delle variabili dipendenti considerate.

Un risultato interessante di questa indagine è l'assenza di un effetto di incorniciamento in relazione alle intenzioni espresse di usare creme protettive in attività quotidiane al di fuori della spiaggia. Questo risultato può essere in parte dovuto alle limitate informazioni offerte ai partecipanti attraverso gli opuscoli. Il messaggio degli opuscoli riguardava in modo specifico i comportamenti adottati in spiaggia e le intenzioni immediate (per es., «In questo preciso momento ti trovi esposto al sole – ti stai proteggendo, evitando così danni alla tua pelle?»). Sebbene le informazioni presentate trattassero il tema del rischio di cancro alla pelle e l'importanza dell'uso di prodotti protettivi in generale, i partecipanti potrebbero non aver considerato a sufficienza l'importanza di estendere questo comportamento ad attività quotidiane più comuni (si vedano, in proposito, Ajzen, 1982, e Ajzen, Fishbein, 1977).

I nostri risultati permettono di dissipare il sospetto che gli effetti di incorniciamento documentati in ricerche precedenti non dipendessero esclusivamente da una presentazione delle informazioni in termini di guadagni o di perdite, ma piuttosto da due dimensioni (tipo di possibili esiti dell'azione e caratteristiche di tali esiti) che sono sottesi alla distinzione fra guadagni e perdite. Sebbene sia difficile trarre conseguenze dai risultati negativi di un singolo studio, è il caso di osservare che, rispetto a tutte le variabili che abbiamo considerato, l'entità degli effetti dovuti a ciascuna di queste due specifiche dimensioni è stata estremamente modesta. Per il momento, quindi, ci sono scarse ragioni per ritenere che le distinzioni connesse a queste due dimensioni siano di per sé importanti nell'incorniciamento di informazioni e di messaggi relativi alla salute. È invece lecito assumere che, di fronte a una presentazione incentrata sui guadagni, le persone siano in grado di riflettere sul comportamento raccomandato sia nei termini dei benefici che se ne potrebbero ottenere sia nei termini dei danni che si potrebbero evitare. Analogamente, di fronte a una presentazione incentrata sulle perdite, le persone possono soffermarsi sia sui danni in cui potrebbero incorrere per la mancata attuazione del comportamento raccomandato sia sui benefici che potrebbero andare in tal modo perduti. A quanto pare, il primo punto di vista (qualunque sia il tipo di presentazione dei guadagni) produce una maggiore motivazione del secondo (qualunque sia il tipo di presentazione delle perdite) a mettere in atto azioni preventive, come l'uso di creme protettive.

Uno dei punti di forza di questo studio rispetto a indagini precedenti è che ci siamo rivolti a individui per i quali il tipo di comportamento studiato era estremamente saliente. Il vantaggio della presentazione in termini di guadagni nel favorire la richiesta di campioni gratuiti di una crema protettiva, l'intenzione di applicarla ripetutamente e l'intenzione di impiegare un fattore di protezione adeguato conferma i precedenti risultati di Rothman *et al.* (1993), che avevano documentato un vantaggio simile soprattutto fra i partecipanti maggiormente coinvolti (cioè fra le donne). Il contesto della spiaggia potrebbe avere immediatamente caratterizzato il nostro intero campione come «altamente coinvolto». Si ritiene che il coinvolgimento o l'interesse verso un problema incoraggi l'elaborazione sistematica delle informazioni e l'effetto di incorniciamento potrebbe manifestarsi precisamente in quegli individui che tendono a elaborare le informazioni in modo sistematico (Wegener, Petty, Klein, 1994).

Con l'eccezione della richiesta di campioni gratuiti di crema solare protettiva, i nostri risultati sono stati principalmente determinati dalle reazioni di coloro che non avevano inizialmente previsto di usare creme protettive durante la loro permanenza in spiaggia. Riteniamo che questo risultato sia intuitivamente comprensibile. È infatti probabile che i partecipanti che si erano recati in spiaggia con l'intenzione di usare creme protettive abbiano fornito fin

da subito risposte corrispondenti a valori molto alti delle nostre misure, e che tali misure non fossero sufficientemente sofisticate per registrare successive modifiche nei loro atteggiamenti o nelle loro intenzioni. Questi partecipanti potrebbero inoltre non aver reagito in modo significativamente differente alle presentazioni in termini di guadagni e di perdite, nel senso che entrambi i tipi di presentazione possono essere apparsi ugualmente persuasivi alla luce della loro iniziale inclinazione a usare creme protettive. Per contro, i partecipanti che non avevano inizialmente previsto di utilizzare creme protettive durante la loro permanenza in spiaggia hanno avuto maggiori margini di cambiamento del loro atteggiamento, oppure sono stati più sensibili alla differenza fra presentazioni in termini di guadagni e in termini di perdite. I messaggi presentati in termini di guadagni hanno chiaramente avuto un forte impatto su questi partecipanti.

Questo studio, nella misura in cui rappresentava un intervento educativo sul campo, ha un certo numero di limiti, tre dei quali meritano di essere commentati. Innanzi tutto, l'«intervento» è stato estremamente breve ed è consistito nella presentazione di un numero limitato di informazioni contenute in un testo di pochi paragrafi. Abbiamo cercato di tenere conto della forza e della qualità del messaggio presentato negli opuscoli, ma abbiamo dovuto anche considerare i limiti di spazio e la quantità di tempo che i bagnanti sarebbero stati disposti a investire come partecipanti allo studio.

Il secondo limite riguarda la nostra principale misura comportamentale: la richiesta di un campione di crema solare con un fattore di protezione 15. Tale variabile dipendente rappresenta certamente un indicatore imperfetto dei comportamenti relativi alla protezione dall'esposizione solare. Non c'è alcuna garanzia che i partecipanti che hanno richiesto il campione fossero intenzionati a usarlo, né, del resto, che quelli che non l'hanno richiesto non fossero intenzionati a usare un proprio prodotto. Ciononostante, riteniamo che l'impiego di questa variabile sia difendibile per tre ragioni. Innanzi tutto, l'incorniciamento ha esercitato un effetto sistematico sulla richieste del campione gratuito di crema protettiva. In secondo luogo, questo risultato è sostenuto dalle intenzioni, espresse dopo la lettura degli opuscoli, di utilizzare un fattore di protezione uguale o superiore a 15 e di applicare ripetutamente creme solari protettive, così come da una tendenza nelle intenzioni di utilizzare creme solari durante la permanenza in spiaggia. Infine, i presenti risultati replicano ed estendono un insieme di osservazioni simili (Rothman *et al.*, 1993). Sarebbe comunque importante raccogliere ulteriori dati relativi all'uso effettivo di creme e ad altri comportamenti volti a ridurre l'esposizione solare (come l'utilizzo di ombrelloni o di indumenti protettivi).

Il terzo limite di questo studio riguarda il fatto che non sono stati raccolti dati relativi al comportamento sul lungo periodo. Le presentazioni in termini di guadagni hanno avuto un impatto significativo sulle intenzioni e i compor-

tamenti individuali nella giornata in cui lo studio si è svolto; non sappiamo, però, quanto tali intenzioni e comportamenti possano essere durevoli. Una prospettiva per future indagini consiste quindi nello sviluppo di un intervento più articolato che permetta di seguire il comportamento dei bagnanti su intervalli di tempo più estesi. Sarebbe inoltre importante estendere l'intervento in modo da includere ulteriori comportamenti protettivi e sottolineare l'importanza dell'impiego di creme solari nelle attività quotidiane anche al di fuori della spiaggia. In conclusione, consideriamo il presente studio un importante passo in avanti rispetto alla sperimentazione in laboratorio, ma ancora un piccolo passo verso più articolate ricerche sul campo.

La ricerca riportata in questo articolo è stata finanziata dalla American Cancer Society (Grant RPG-93-028-05-PBP). Desideriamo anche segnalare il sostegno ottenuto dal National Cancer Institute (Grant RO1-CA68427) e dal National Institute of Mental Health (Grant PO1-MH/DA56826). Infine, questo lavoro è stato parzialmente finanziato da una borsa di studio post-laurea della National Science Foundation.

Ringraziamo Duane Wegener e Mark Whisman per i loro commenti su precedenti stesure di questo articolo. Ringraziamo inoltre Kristi Lemm, Janet Zullo, Brian Lizotte e i membri del Laboratorio 1995-1997 su *Salute, emozioni e comportamento* (*Health, Emotions, and Behavior*) per l'assistenza nel concepimento e nello sviluppo del materiale impiegato nello studio e per i commenti durante la presentazione dei risultati. Ringraziamo infine Amos Tversky per avere ispirato questo lavoro.

*16. Un approccio cognitivo alle decisioni sul cancro: implicazioni per la prevenzione e il trattamento**

di *Baruch Fischhoff*

1. Introduzione

Per prevenire, trattare e sopravvivere al cancro è necessario far fronte a numerose decisioni. Tali decisioni riguardano, per esempio, quale dieta adottare, se fumare, come cercare di smettere, a quali controlli sottoporsi, quali sintomi segnalare al medico, se informare familiari (e colleghi) dei risultati di test genetici, quando richiedere un secondo parere medico (e come comportarsi di fronte a pareri discordi), come valutare informazioni mediche reperite sul web, quanto dovrebbe essere aggressivo il trattamento, come richiedere antiemetici (nel caso non vengano offerti), come sostenere familiari malati, se sostenere ricerche mediche controverse (per esempio, sulle cellule staminali o sui primati) e, più in generale, quanto preoccuparsi dei rischi associati al cancro.

Il nostro benessere individuale e collettivo dipende in parte dalla qualità di decisioni come queste. Per decidere efficacemente, un individuo deve comprendere i fatti e le proprie personali valutazioni in misura sufficiente da poter identificare l'opzione più conforme ai suoi interessi, e metterla in atto. Nella teoria della decisione, le scelte vengono descritte in termini di *opzioni* (le azioni che è possibile intraprendere, compresa l'inazione), *esiti* (le possibili conseguenze di quelle azioni), *valori* (l'importanza relativa dei diversi esiti), *gradi di incertezza* (riguardo agli esiti che potrebbero verificarsi), e *regole decisionali* (che integrano i diversi elementi determinando la scelta). Questo approccio, noto come analisi *normativa*, mira a cogliere la natura di una decisione per come essa apparirebbe a un individuo pienamente informato.

Due esempi mostrano come le scelte di prevenzione e di trattamento del cancro possano essere ricostruite in questi termini.

* Traduzione di *Decision research strategies*, «Health psychology», 24 (2005), pp. S9-S16, copyright © 2005 American Psychological Association.

Un problema decisionale comune, che ha potenziali implicazioni per la prevenzione del cancro, riguarda l'abbronzatura. Le opzioni coinvolte comprendono diversi modi per regolare l'esposizione a raggi ultravioletti o per ottenere l'abbronzatura con strumenti artificiali. Gli esiti rilevanti riguardano per esempio il proprio aspetto fisico, il modo in cui gli altri reagiscono a esso, conseguenze di tipo emotivo e costi economici. Le questioni di valore hanno a che fare con l'importanza che si attribuisce all'approvazione sociale e agli effetti a lungo termine sulla salute. L'incertezza riguarda, per esempio, gli effetti sociali dell'abbronzatura, l'efficacia della protezione solare dopo che si è nuotato, la qualità della manutenzione delle attrezzature nei centri di abbronzatura, la suscettibilità individuale al cancro della pelle, e quanta preoccupazione causerà una brutta scottatura. Assumendo che sia possibile confrontare fra loro i diversi esiti, ci si può servire di una regola decisionale basata sull'*utilità attesa*, che indirizza la scelta sull'opzione i cui possibili esiti risultano i più attraenti (in termini di *utilità*) una volta combinati con la probabilità che ciascuno di essi si verifichi. Una regola di decisione alternativa consiste nell'escludere tutte le opzioni con possibili esiti considerati inaccettabili (per esempio, un qualsiasi rischio di cancro, o la necessità di continue preoccupazioni). Un'altra ancora consiste nel selezionare un'opzione che favorisca specifici esiti considerati importanti (per esempio, un'abbronzatura imbattibile).

Un secondo esempio riguarda i pazienti con un cancro in fase avanzata, che spesso si trovano a dover scegliere fra trattamenti che comprendono cure convenzionali, alternative e palliative. Gli esiti rilevanti possono riguardare la longevità, il dolore, la funzionalità dei trattamenti, i costi e le relazioni personali. I valori coinvolti riflettono il peso relativo associato a tali esiti. Gli elementi di incertezza possono riferirsi all'efficacia dei trattamenti, alle reazioni dei familiari e alla capacità del paziente di sopportare il dolore e di rimanere sereno. Anche in questo caso, è possibile scegliere in base alla massimizzazione dell'utilità attesa. In alternativa, il problema di scelta può essere semplificato isolando le opzioni che si ritengono coerenti con un obiettivo globale, come «morire con dignità» o «lottare fino all'ultimo». La considerazione dell'utilità attesa potrebbe quindi intervenire nella scelta all'interno di questo più ristretto insieme di opzioni (per esempio, nel decidere come servirsi delle cure palliative disponibili).

Caratterizzare i problemi di scelta nei termini della teoria della decisione aiuta a descriverli in modo chiaro e articolato: permette, per esempio, di evidenziare quando la scelta ottimale differisce da persona a persona, perché gli esiti coinvolti sono diversi o perché non vengono valutati nello stesso modo. L'analisi normativa può rivelare strutture comuni in scelte fra loro diverse: essa per esempio prevede che vengano considerati tanto i costi *diretti* della scelta di un'opzione quanto i costi *di opportunità* che conseguono dalla man-

cata scelta di un'opzione. L'analisi normativa ignora invece i «costi affondati» (*sunk costs*), vale a dire impegni e investimenti passati (come l'iscrizione a un centro di abbronzatura o la promessa di fare a meno di potenti antidolorifici), a meno che essi non risultino rilevanti nella considerazione di esiti futuri (per esempio, riuscire a salvare la faccia o poter sostenere di essere infallibili). L'analisi normativa è d'aiuto nel chiarire la struttura di scelte complesse: per esempio, essa opera una distinzione fra l'incertezza sul modo in cui si reagirà al verificarsi di un certo esito e l'incertezza relativa a ciò che realmente si desidera; così come distingue il caso in cui si dà poco valore a esiti lontani nel tempo dal caso in cui si considera improbabile sopravvivere abbastanza a lungo da farne esperienza. Infine, l'analisi normativa è necessaria per definire gli *effetti di contesto*, vale a dire i casi in cui problemi formalmente equivalenti inducono scelte differenti in base al modo in cui vengono presentati.

Rappresentare i problemi di scelta nei termini dell'analisi normativa non significa però assumere che gli individui prendano le loro decisioni seguendone i principi. Di fatto, gli psicologi si sono impegnati a identificare e spiegare significative discrepanze fra l'ideale normativo e la realtà descrittiva della decisione umana, mettendo a punto possibili interventi *prescrittivi* in grado di aiutare le persone a decidere meglio (cioè in modo più vicino ai principi normativi). Tali interventi possono migliorare le decisioni modificando gli individui (per esempio, attraverso la formazione) o modificando il loro ambiente (per esempio, attraverso una comunicazione più efficace). In casi in cui sembra improbabile che gli individui possano decidere in modo appropriato, gli interventi prescrittivi possono richiedere l'indebolimento, e persino la limitazione, della libertà di scelta (per esempio nel caso in cui venga bandito un prodotto le cui caratteristiche e il cui corretto utilizzo verranno probabilmente fraintesi).

La rappresentazione formale delle decisioni degli individui evidenzia anche i problemi strategici dei ricercatori impegnati a studiare (e ad aiutare) tali decisioni. Il prossimo paragrafo descrive questi problemi strategici ed è seguito da una breve esposizione delle scelte strategiche caratteristiche dello studio cognitivo delle decisioni. L'approccio così definito sarà quindi illustrato attraverso cinque casi-studio che sollevano questioni analoghe a quelle coinvolte in numerose decisioni che riguardano il cancro. L'articolo si conclude con la definizione di una strategia di ricerca per lo studio e il miglioramento delle scelte di prevenzione e di trattamento del cancro.

2. Quattro scelte strategiche

In ogni progetto di ricerca per lo studio delle decisioni si presentano diverse scelte strategiche.

(1) *Il progetto di ricerca deve partire da un'analisi normativa del problema decisionale che si intende studiare?* È cioè opportuno caratterizzare esplicitamente le opzioni, gli esiti, i valori e i gradi di incertezza coinvolti, e quindi identificare le scelte ottimali che ne conseguono? I ricercatori che non svolgono un'analisi di questo tipo devono assumere di conoscere intuitivamente in che modo le scelte degli individui influiranno sulle loro vite e in che modo gli individui reagiranno agli effetti di tali scelte. Nella prevenzione e nel trattamento del cancro, alcuni problemi decisionali presentano scelte familiari che non richiedono un'estesa analisi formale; altre scelte sono invece nuove, complesse e incerte e mettono alla prova le intuizioni degli stessi ricercatori. In questi casi, senza un'analisi dettagliata si corre il rischio di fraintendere la natura del problema decisionale studiato.

(2) *Gli interventi che si intende praticare hanno un carattere persuasivo?* Hanno, cioè, l'obiettivo di promuovere particolari comportamenti o lasciano la scelta interamente agli individui coinvolti? Un approccio persuasivo presuppone che i ricercatori padroneggino i fatti rilevanti, che siano a conoscenza dei valori degli individui e che siano legittimati a manipolarli. Se queste assunzioni non si rivelano valide, allora l'intervento praticato può fornire consigli inappopriati, generare false sicurezze o violare l'autonomia degli individui coinvolti. Un punto di vista non persuasivo, per contro, presuppone che le persone siano in grado di decidere meglio se le opzioni che hanno a disposizione risultano più chiare e comprensibili. Questo secondo approccio può quindi mostrarsi inefficace se gli individui non riescono ad afferrare i termini del problema, non sanno ciò che vogliono o rifiutano la responsabilità della scelta. Nella prevenzione e nel trattamento del cancro gli esiti delle diverse opzioni sono spesso complessi, incerti, poco familiari e difficili da confrontare fra loro. Quando chi è chiamato a decidere non riesce a comprendere e a valutare gli esiti che ha di fronte, è possibile che voglia farsi guidare da qualcuno in cui ripone la sua fiducia, anche se solitamente preferisce decidere in modo autonomo. Se le cose stiano effettivamente così nelle diverse situazioni è una questione empirica. A seconda della risposta, si può valutare la praticabilità di un approccio non persuasivo.

(3) *Gli interventi che si intende praticare mirano a migliorare scelte specifiche o a promuovere la padronanza generale di un ambito decisionale?* Vale a dire, tali interventi collocano uno specifico problema di scelta in un contesto più ampio? Qualsiasi scelta (a meno che non abbia esiti fatali) non è altro che un episodio in una sequenza: può essere rivista («dovrei continuare a seguire questo corso di azione?»), può dar luogo a nuovi problemi decisionali

(per esempio, «dovrei rivelare i risultati del mio test?») e può modificare la persona coinvolta (per esempio, generando rimpianti o sofferenze che risulteranno rilevanti per scelte successive). Ampliare il contesto può rendere un problema decisionale più comprensibile (mettendo in luce come ciascuna opzione influenzi l'intero processo di creazione e selezione degli esiti) e maggiormente significativo (chiarendo il suo rapporto con le vite degli individui). D'altra parte, la considerazione del contesto può anche aumentare l'impegno cognitivo richiesto a chi decide. La prevenzione e il trattamento del cancro presentano molti problemi decisionali in cui gli elementi del contesto hanno un'importanza critica, ma le capacità degli individui di tenerne conto sono spesso limitate.

(4) *Quali differenze individuali dovrebbero essere considerate?* In altri termini: con quali strumenti i ricercatori valutano come una stessa decisione varia fra i diversi individui che l'affrontano? I singoli individui possono trovarsi di fronte a insiemi di opzioni differenti (per esempio, in funzione del loro stato di salute, del sostegno che ricevono da altri, delle risorse finanziarie che hanno a disposizione); possono avere valori differenti (per esempio, in termini di orizzonte temporale, di tolleranza dell'ambiguità, di preoccupazione per le conseguenze sugli altri); possono favorire strategie decisionali differenti (in funzione della loro attitudine al ragionamento, o del loro equilibrio emotivo); possono risentire di contesti sociali differenti (per esempio, in termini di aspettative culturali o di partner decisionali). Pertanto un dato problema decisionale (per esempio, relativo all'abbronzatura o a un trattamento contro il cancro) può avere caratteristiche differenti per i diversi individui, così come differenti possono rivelarsi gli effetti di un qualsiasi intervento messo in atto dai ricercatori. Nella misura in cui la malattia sa essere una grande livellatrice, il cancro riguarda persone molto diverse fra loro in periodi molto diversi delle loro vite. In particolare, numerose decisioni relative alla prevenzione del cancro riguardano la maggior parte delle persone in molti momenti della loro vita. I ricercatori sono quindi chiamati a decidere quali considerare fra le fonti di variabilità individuale potenzialmente rilevanti.

3. Una strategia per la ricerca cognitiva

Lo studio cognitivo delle decisioni ha rappresentato la risposta della psicologia all'analisi normativa delle decisioni. La teoria alla base dell'analisi normativa mostra come si possano individuare le scelte *razionali* a partire da alcune regole semplici e altamente plausibili (per esempio, la coerenza). Gli economisti hanno spesso implicitamente assunto che la ricostruzione normativa delle decisioni fosse valida anche dal punto di vista descrittivo, mentre gli psicologi hanno messo in luce quanto possa essere difficile conformarsi ai

principi di razionalità, specialmente quando gli individui affrontano situazioni complesse, incerte e poco familiari con risorse cognitive limitate. Le ricerche descrittive degli psicologi si sono quindi rivolte allo studio del comportamento a partire dallo scarto dall'ideale normativo, ponendo così le basi per interventi prescrittivi (von Winterfeldt, Edwards, 1986; Hastie, Dawes, 2002; Yates, 1990).

In pratica, quello che può apparire come un passaggio fra livelli di indagine separati spesso richiede l'interazione fra i diversi punti di vista. La ricerca descrittiva può mettere in luce esiti o circostanze che non avevano trovato posto nell'analisi normativa del problema decisionale (per esempio, il modo in cui gli individui effettivamente assumono i farmaci o percepiscono i loro effetti collaterali). L'elaborazione di interventi prescrittivi spesso promuove ulteriori indagini descrittive, sul modo in cui le persone li recepiscono, o ulteriori analisi normative, per l'individuazione degli aspetti che richiedono una particolare enfasi. Può risultare necessario rivedere le ricerche svolte alla luce di nuove indagini (per esempio, in base ai risultati di nuovi studi clinici) o di eventi esterni (per esempio, la diffusione del timore per una data malattia o la guarigione di una celebrità). Infine, verificare quanto è possibile ottenere con interventi prescrittivi permette di identificare i limiti che caratterizzano le procedure di scelta informata.

Nei termini delle scelte strategiche considerate nel paragrafo precedente:

(1) *La ricerca cognitiva inizia con un'analisi normativa*. In tal modo essa riconosce che i ricercatori non possono affidarsi all'intuizione per identificare la scelta ottimale per chi deve decidere. Piuttosto occorre elicitare gli obiettivi degli interessati, consultare esperti della materia per determinare gli esiti che possono seguire a ciascuna opzione e fare ricorso a una regola decisionale appropriata per selezionare l'alternativa più conforme agli interessi di chi affronta il problema di scelta studiato.

(2) *L'approccio di fondo della ricerca cognitiva è non persuasivo*. L'obiettivo è aiutare le persone a decidere per se stesse. Da un punto di vista filosofico, la speranza è quella di rafforzare l'autonomia decisionale – un proposito che la ricerca cognitiva condivide con l'economia neoclassica. Da un punto di vista pragmatico, i ricercatori spesso lavorano con persone che non vogliono sentirsi dire che cosa fare (per esempio, funzionari, dirigenti, e molti pazienti). L'indipendenza delle scelte (con o senza il beneficio di interventi prescrittivi) comporta dei costi in termini di efficiacia, come mostra lo scarto fra le scelte ottimali (identificate da un'appropriata analisi normativa) e le scelte effettive (identificate dalla ricerca descrittiva). Se il sacrificio di scelte ottimali per scelte autonome sia giustificato è una questione etica; la ricerca empirica può definirne i termini, ma non può fornire una risposta.

(3) *La ricerca cognitiva tipicamente si rivolge a problemi decisionali specifici*. Molte scelte specifiche e ben definite si rivelano già piuttosto difficili

per le persone. Concentrandosi su problemi di questo tipo, i ricercatori impegnati nello studio cognitivo delle decisioni mettono a frutto le loro principali abilità: l'elicitazione di credenze e di valori, e la loro combinazione attraverso regole decisionali. Occorre comunque osservare che concentrarsi su problemi decisionali specifici può comportare dei rischi: è possibile che l'impatto complessivo delle scelte venga oscurato (per esempio, le conseguenze di scelte ripetute) o che vengano trascurati elementi del contesto necessari per apprezzare il reale significato del problema decisionale (per esempio, le ragioni per cui i rischi sono così alti, o così bassi, come gli esperti affermano; o quali sono le effettive conseguenze di un certo cambiamento delle condizioni di salute). Per includere questi elementi del contesto, è necessaria un'analisi estesa dei processi di creazione e selezione dei possibili esiti (Fischhoff, Downs, Bruine de Bruin, 1998).

(4) *La ricerca cognitiva deve tenere conto delle condizioni diverse in cui gli individui si trovano a decidere, e dei loro valori diversi*. Come si è detto, in un dato problema decisionale le scelte ottimali possono risultare diverse per persone le cui condizioni o i cui obiettivi differiscono. La ricerca cognitiva non può quindi analizzare normativamente le decisioni senza tenere conto delle possibili differenze individuali in termini di opzioni disponibili, valori e conseguenze attese. La considerazione di ulteriori fattori individuali dipende dall'interesse dei risultati che ci si attende. Per esempio, tenere conto di differenze individuali nell'avversione al rischio non si è rivelato particolarmente produttivo (Yates, 1990). Come accade con molti altri aspetti della personalità, queste caratteristiche mostrano una forte dipendenza dalle situazioni (per esempio, una persona che rischia molto nella carriera lavorativa potrebbe essere un investirore o un viaggiatore molto prudente). Considerare differenze individuali in termini di competenze decisionali (Parker, Fischhoff, 2004; Stanovich, West, 1998) e di stati affettivi (Loewenstein, Lerner, 2003) potrebbe rivelarsi una linea di indagine fruttuosa. In questo contesto emergono possibili divergenze fra interessi teorici e pratici: l'influenza di un particolare fattore potrebbe risultare robusta e rivelatrice, ma di entità troppo ridotta per alterare il comportamento in molti problemi di scelta.

Il prossimo paragrafo riporta cinque applicazioni dell'approccio strategico qui delineato, soffermandosi sulle possibili implicazioni per problemi decisionali di prevenzione e trattamento del cancro. Nei casi in cui si rivelano riuscite, tali applicazioni non solo producono decisioni migliori, ma promuovono il progresso scientifico attraverso l'interazione fra *ricerca di base applicata* e *ricerca applicata di base* (Baddeley, 1979): da una parte, infatti, esse indagano la validità ecologica della ricerca di base, cioè la sua capacità di guidare la pratica; dall'altra parte, attraverso l'applicazione a problemi pratici, identificano nuovi temi per la ricerca di base.

4. Applicazioni della ricerca cognitiva sulle decisioni

4.1. *Consenso informato*

Spesso le procedure mediche implicano un beneficio incerto in cambio dell'assunzione di alcuni rischi. Informare le persone di tutti i possibili effetti collaterali di una procedura può produrre un sovraccarico di dettagli in una situazione già difficile. Ciò di cui le persone hanno bisogno è la comunicazione efficace di pochi fatti che risultano critici per le loro decisioni. In termini formali, questa esigenza riflette la necessità di una analisi normativa del *valore dell'informazione*. Merz e collaboratori hanno svolto un'analisi di questo tipo in relazione all'endarterectomia carotidea, i cui possibili rischi e benefici sono rappresentativi di molti interventi chirurgici (Merz *et al.*, 1992). Liberare la carotide da parziali ostruzioni fino al cervello può ridurre il rischio di ictus. Tuttavia, tale operazione può anche causare numerosi problemi di diversa gravità, fra cui morte, lo stesso ictus, emicranie, rottura dei denti. L'analisi di Merz e collaboratori mostra che soltanto tre di questi rischi sono sufficientemente seri da dissuadere un numero significativo di potenziali pazienti: morte, ictus e paresi facciale. Sebbene nulla debba essere nascosto, per poter esprimere un consenso informato è necessario che i pazienti comprendano questi tre possibili effetti collaterali. Ciò significa *sia* comprendere quanto è probabile che ciascuno di essi si verifichi *sia* che cosa significhi incorrervi.

Merz e collaboratori non hanno svolto ricerche descrittive sulle credenze dei pazienti potenzialmente interessati all'endarterectomia carotidea, osservando che pochi di loro potrebbero essere a conoscenza della probabilità di eventi così inusuali, a meno che non ne vengano informati. Per quanto riguarda il significato di tali eventi, si può suppore che il rischio di morte sia familiare a tutti i pazienti e quello di ictus alla maggior parte di loro (tutti infatti lo temono e molti ne hanno già fatto esperienza). Il rischio di paresi facciale potrebbe invece richiedere specifiche spiegazioni. Il compito prescrittivo di comunicare efficacemente la natura di questi rischi può avvalersi di tutte le indicazioni rilevanti provenienti dalla ricerca di base e da quella applicata. Per esempio, molti studi suggeriscono che è preferibile servirsi di probabilità numeriche piuttosto che di quantificatori puramente verbali (come «probabile», o «raro»), la cui interpretazione può variare notevolmente fra individui e contesti differenti. Se le probabilità in questione fossero molto piccole, renderle comprensibili costituirebbe una sfida considerevole; per fortuna (ma solo dal punto di vista della comunicazione) tutti e tre gli eventi sono relativamente frequenti. Se la procedura chirurgica fosse un evento che si ripete, allora potrebbe essere difficile far capire in che modo i rischi si accumulano attraverso l'esposizione ripetuta; per fortuna (sia dal punto di vista della comunicazione, sia per i pazienti) si tratta di un evento singolo (Fischhoff, Bostrom, Quadrel, 2002).

Nel caso dell'endarterectomia carotidea sembrerebbe che i fatti rilevanti siano sufficientemente semplici da rendere pienamente praticabile un approccio non persuasivo, cioè volto ad aiutare i pazienti a identificare autonomamente la migliore opzione di intervento. Da un punto di vista empirico tale considerazione resta però da verificare. A questo proposito occorre sottolineare che la valutazione di un intervento richiede l'identificazione di una misura di efficacia: si tratta «solo» della conoscenza dei fatti che risultano critici per il valore informativo che contengono? Oppure è anche necessario controllare se i pazienti sono capaci di trarre delle conclusioni da quei fatti, sanno valutare il proprio livello di conoscenza e si sentono in grado di agire sulla base delle proprie credenze? Le decisioni dei pazienti dovrebbero coincidere con quelle identificate come razionali dall'analisi normativa? È necessario controllare se i pazienti sono soddisfatti del processo decisionale, anche se esso li mette di fronte a una scelta estremamente importante e difficile?

La maggior parte dei trattamenti per il cancro hanno lunghe liste di possibili effetti collaterali. Anche in questo ambito può quindi essere opportuno lo svolgimento di analisi normative per l'identificazione delle questioni più rilevanti per i pazienti, seguite dalle ricerche descrittive e prescrittive necessarie per promuovere il consenso informato e per controllare in che misura esso sia effettivamente tale. Sebbene questo approccio richieda ulteriori analisi e indagini, esso è necessario per tradurre in termini empiricamente controllabili l'importante requisito del consenso informato e potrebbe permettere ai medici di mostrare con chiarezza che i pazienti hanno compreso i fatti determinanti per le loro scelte.

4.2. *Consigli per la sicurezza personale*

Spesso gli individui che affrontano procedure mediche impegnative possono avvalersi di consulenze personali in sostegno alle loro decisioni. Nel caso di molti altri rischi per la vita ciò non accade: le persone ricevono piuttosto indicazioni generali rivolte a un uditorio esteso. Dal punto di vista normativo, tali indicazioni generali presuppongono che tutti i destinatari affrontino situazioni simili e perseguano obiettivi simili. Per esempio, nel tentativo di ridurre il rischio di violenza sessuale, le donne spesso si trovano di fronte a indicazioni assertive, universali e fra loro contraddittorie sull'opportunità di opporre o meno una resistenza fisica (se lottare oppure no con un eventuale aggressore). Anche consulenti guidati da buone intenzioni possono così aumentare la confusione in situazioni già difficili. Inoltre, in queste condizioni, risulta più facile incolpare le vittime: qualunque cosa una donna abbia fatto per proteggersi, avrà infatti violato le indicazioni di qualche «esperto». Per molti versi, i problemi di scelta che le donne affrontano in questo contesto somi-

gliano a quelli di chi cerchi di orientarsi fra i numerosi suggerimenti che riceve per la riduzione del rischio di cancro (proteggiti dal sole, perdi peso, fai uso di certe erbe, ecc.).

Gli elementi fondamentali di qualunque analisi normativa sono le opzioni, i possibili esiti, e le probabilità che mettono in relazione le prime con i secondi. In un progetto di ricerca (esposto sinteticamente in Fischhoff, 1992) abbiamo identificato le opzioni e gli esiti rilevanti nella prevenzione delle violenze sessuali in base a una rassegna dei materiali pubblicati e a interviste con donne ed esperti del campo. Dalle interviste è emerso un gran numero di alternative di azione (più di 1.200) sufficientemente diverse fra loro da influenzare in modo differente gli esiti che i partecipanti avevano indicato come importanti. Il numero delle opzioni e degli esiti coinvolti mostra la complessità del compito decisionale che le donne fronteggiano. In questo studio, donne appartenenti a gruppi diversi individuavano opzioni diverse e davano importanza a esiti differenti, e lo stesso valeva per gli esperti consultati. I membri di questi diversi gruppi si trovano quindi di fronte a scelte fra loro differenti, con il risultato che indicazioni utili per certe donne potranno risultare inappropriate per altre.

Una meta-analisi degli studi sull'efficacia di diverse strategie difensive ha rivelato che le informazioni scientifiche disponibili sono di limitata utilità, mostrando che qualsiasi suggerimento espresso in termini assertivi è fuori luogo. Le ricerche forniscono però un'indicazione, per quanto debole: in confronto ad altre forme di reazione, la resistenza fisica tende a diminuire la probabilità che un'aggressione venga portata a compimento. Nella nostra indagine, l'analisi delle credenze dei partecipanti ha rivelato che tanto le persone comuni quanto gli esperti sopravvalutano l'efficacia di strategie con limitati margini di applicazione (per esempio, controllare sempre il sedile posteriore dell'auto).

La comunicazione di consigli appropriati e qualificati deve rispettare la complessità, l'incertezza e l'eterogeneità delle decisioni che le donne si trovano di fronte nel tentativo di prevenire le violenze sessuali. Sfortunatamente, nel momento in cui eravamo pronti per avviare la parte prescrittiva del progetto, l'ente finanziatore, l'Istituto Nazionale di Salute Mentale (National Institute of Mental Health), ha smantellato il suo Centro per la prevenzione e il trattamento delle violenze sessuali. Non solo il nostro progetto si è fermato, ma lo stesso è accaduto alle ricerche sull'efficacia di varie misure protettive. In assenza dei risultati di tali ricerche, è necessaria molta cautela nell'offrire suggerimenti e nel criticare chi li segue (o li respinge).

Le indicazioni relative a scelte di prevenzione e trattamento del cancro sono talvolta sostenute da una più solida base scientifica di quella disponibile nel campo della prevenzione delle violenze sessuali. Ciononostante, non è infrequente che tali indicazioni presentino un quadro confuso. Naturalmente, si

dovrebbero evitare il fumo, le scottature solari e i luoghi di lavoro fortemente inquinati. Ma la situazione è molto meno chiara quando si tratta di evitare luoghi e pratiche i cui effetti cancerogeni restano presunti (come la fluorizzazione o la lavatura a secco) o di impiegare cibi o integratori dietetici che si suppone possano proteggere dal cancro. In questi casi è possibile imbattersi in suggerimenti assertivi e fra loro contraddittori, specie se non vengono specificati l'entità degli effetti e la qualità delle ricerche, così come le credenziali e le motivazioni di chi fornisce le informazioni (Woloshin *et al.*, 2000). C'è solo da augurarsi che gli esperti che analizzano più accuratamente gli effetti di possibili strategie preventive e li comunicano in modo responsabile e comprensibile guadagnino le più ampie quote di mercato.

4.3. *Infezioni a trasmissione sessuale fra gli adolescenti*

I giovani ricevono molte informazioni su come ridurre il rischio di infezioni a trasmissione sessuale, soprattutto in relazione a HIV e AIDS. Nonostante ciò, la presenza di infezioni a trasmissione sessuale (e di gravidanze inattese) è ancora diffusa. Di fronte a questo problema, alcuni adulti hanno perso fiducia nell'efficacia di un approccio basato sull'informazione, considerando con favore interventi volti a influenzare il comportamento in modo più diretto. Nella maggior parte dei casi, però, i messaggi rivelatisi «inefficaci» non erano stati elaborati sulla base di un'analisi normativa delle scelte degli adolescenti e dei loro bisogni informativi. Ciò che in prevalenza è accaduto è che gli educatori decidevano ciò che gli adolescenti avevano bisogno di sapere; quindi cercavano una presentazione coinvolgente. Ma campagne del tipo «Just say "No"»[1] sono spesso state criticate proprio per il fatto che semplificano in modo eccessivo l'analisi normativa delle scelte, ignorando molte delle preoccupazioni degli adolescenti.

Alcuni programmi di intervento di più ampio respiro cercano di informare gli adolescenti sui «fatti della vita» rilevanti per compiere scelte appropriate. In un'analisi normativa del problema, abbiamo ricostruito il complesso di processi cognitivi, affettivi, sociali e psicologici che determinano il rischio di infezioni a trasmissione sessuale (Fischhoff, Downs, Bruine de Bruin, 1998). Abbiamo quindi osservato che i messaggi più diffusi si concentrano su una piccola parte di questi numerosi fattori. L'attenzione su alcuni aspetti specifici ha dato i suoi frutti: gli adolescenti sono a conoscenza di questi aspetti. Ma non sempre ne hanno una vera padronanza. Inoltre, altri fattori di importanza

1. «Di' di no, e basta!»: è uno slogan coniato negli Stati Uniti e utilizzato per campagne contro il consumo di droghe illegali, soprattutto fra i giovani [N.d.C.].

cruciale non trovano posto nei modelli mentali che guidano il loro comportamento. Alcune di queste omissioni sono deliberate. Per esempio, sebbene non sia difficile spiegare quali sono i rischi associati al sesso orale o anale, poche scuole negli Stati Uniti acconsentono a spiegazioni così «esplicite». Anche la confusione di molti adolescenti riguardo all'uso corretto del preservativo o al significato di alcuni termini (come «sesso sicuro») va ricondotta a influenze di carattere sociale: molti insegnanti e genitori non sono a loro agio nell'affrontare questi argomenti. Altre omissioni riflettono questioni estranee all'ambito dell'educazione sanitaria. Per esempio, gli adolescenti (proprio come gli adulti) tendono a concentrarsi sugli effetti di azioni singole piuttosto che sulla ripetizione di quelle azioni (per esempio, i rapporti sessuali o la guida ad alta velocità). Quando si chiede loro di ragionare sul lungo periodo, tendono a sottovalutare con quanta velocità un'esposizione ripetuta faccia crescere rischi di piccola entità (Shaklee, Fischhoff, 1990; Morgan *et al.*, 2002).

In una serie di indagini abbiamo elaborato un intervento prescrittivo volto a colmare queste lacune nei modelli mentali degli adolescenti. Uno degli obiettivi era dare agli adolescenti un'idea più realistica di quanto essi effettivamente sapessero: quando i messaggi diffusi ripetono con forza solo una parte dei fatti rilevanti, il rischio è quello di creare la sensazione infondata di «sapere già tutto». Inoltre, anche le migliori informazioni significano ben poco se non sono percepite dagli adolescenti come un arricchimento delle proprie possibilità di azione. In una parte del progetto, il nostro approccio prescrittivo è consistito nella creazione di un DVD interattivo destinato alle giovani donne. Il DVD comunicava alcuni fatti critici (riguardo alle infezioni a trasmissione sessuale, all'uso del preservativo, ecc.) e mostrava come si potessero identificare e gestire alcune particolari scelte nell'ambito delle relazioni sessuali. In uno studio randomizzato, l'utilizzo del DVD ha prodotto un aumento significativo della conoscenza delle infezioni a trasmissione sessuale e dell'uso (riportato dai partecipanti) del preservativo, e una diminuzione significativa degli incidenti nell'uso del preservativo e del tasso di infezioni da *chlamydia* (Downs *et al.*, 2004).

Il presupposto di un intervento prescrittivo come questo è l'identificazione, attraverso l'indagine descrittiva, di specifici limiti e lacune nelle credenze degli individui. L'obiettivo è aiutare gli individui a elaborare modelli mentali che comprendano i fatti indicati come rilevanti dall'analisi normativa. Ricerche di questo tipo sono complementari alle indagini che si rivolgono direttamente alla stima dell'entità dei rischi e dei benefici. Anche nell'ambito della prevenzione e del trattamento del cancro, difficilmente i dati statistici possono «parlare da sé» senza la mediazione di modelli mentali dei processi coinvolti. Per esempio, tenere sotto controllo l'esposizione a raggi ultravioletti significa sapere quando il problema si presenta (per esempio, in quale stagione

o in quali ore del giorno), saper gestire i processi sociali coinvolti e, all'occorrenza, improvvisare delle soluzioni (per esempio, quando le creme protettive finiscono). In breve, significa saperne abbastanza per evitare la «fallacia dell'eccezione» (*special pleading*), in cui un individuo si convince, senza valide ragioni, che egli, diversamente dagli altri, non è in realtà esposto ad alcun rischio.

4.4. *Vaccinazioni*

Talvolta è possibile che le persone compiano in modo consapevole scelte potenzialmente dannose per la salute perché danno più importanza ad altri obiettivi. Per esempio, gli adolescenti possono decidere di avere rapporti sessuali non protetti per ragioni edonistiche o per preservare un rapporto. Una particolare forma di comportamento apparentemente subottimale per la salute è il rifiuto della raccomandazione di una vaccinazione per i propri figli (per esempio, contro morbillo, parotite, rosolia, difterite, pertosse, tetano) o per se stessi (per esempio, la vaccinazione contro l'antrace per i militari, o quella contro il vaiolo per il personale chiamato a intervenire immediatamente in casi di emergenza).

Una rappresentazione restrittiva di questi problemi di scelta suggerisce che solo chi è particolarmente avverso agli effetti collaterali della vaccinazione dovrebbe respingerla. In questa prospettiva, l'analisi normativa confronta i dati statistici sulla diffusione e le conseguenze della malattia in questione e quelli sugli effetti collaterali della vaccinazione, combinandoli con i corrispondenti valori di disutilità. Sebbene gli effetti collaterali più seri siano rari, alcune indagini descrittive suggeriscono che i genitori sono molto restii ad accettare il rischio di causare potenziali danni ai propri figli attraverso la vaccinazione (Ritov, Baron, 1995). Adottando un approccio persuasivo, gli interventi prescrittivi potrebbero mostrare ai genitori i dati statistici rilevanti e spingerli a considerarsi ugualmente responsabili delle conseguenze della scelta di effettuare la vaccinazione come della scelta opposta.

Un'analisi normativa di più ampio respiro considera fattori estranei all'ambito medico-sanitario, facendosi guidare dagli altri aspetti a cui le persone danno importanza, messi in evidenza dalle indagini descrittive. Per esempio, alcune persone potrebbero valutare positivamente un contributo al livello generale di «protezione immunitaria» – quale appunto la vaccinazione – proprio in relazione a individui che non dovrebbero essere vaccinati (per esempio, perché immunodepressi). Alcune persone potrebbero valutare positivamente il fatto di prepararsi a una missione che comporta la minaccia di una grave malattia. Alcune persone potrebbero voler evitare le sanzioni imposte a chi non è vaccinato (per esempio, l'impossibilità dell'iscrizione a certe scuole).

Alcune persone potrebbero volersi opporre al carattere coercitivo di alcuni programmi di vaccinazione, spingendosi magari fino al punto di non utilizzare scuole pubbliche. Trascurando di segnalare i benefici di un vaccino, i programmi di informazione perderebbero certamente un'opportunità per risultare persuasivi. Ma ignorando le ragioni per opporsi alla vaccinazione (per quanto difficili siano da accettare o condividere), essi potrebbero aggravare elementi di tensione già presenti (nel corso della discussione di un'analisi normativa degli effetti della vaccinazione contro l'antrace, un ufficiale medico ha suggerito a questo proposito l'obiettivo di un «assenso informato»: i militari avrebbero il dovere di vaccinarsi, ma anche il diritto di sapere che cosa rende necessario il vaccino e quali sono i suoi rischi).

Un'approccio normativo ancora più ampio tiene conto delle fonti d'informazione che diffondono le stime dei rischi e dei benefici. Se non vengono considerati credibili, anche messaggi ben confezionati possono mancare il bersaglio o addirittura suscitare reazioni di rifiuto. Nel caso delle vaccinazioni, alcuni scettici mettono in dubbio i benefici perché considerano trascurabile il rischio di malattia o perché non accettano le dichiarazioni sull'efficacia dei vaccini o sulla loro sicurezza. È possibile che non ritengano affidabili le fonti ufficiali, forse perché convinti che siano state ingannevoli in passato (per esempio, in relazione alla «sindrome della Guerra del Golfo» o all'episodio dell'influenza suina[2]), e che contestino il sistema (per esempio perché il metodo di rilevazione degli effetti collaterali si basa sull'iniziativa autonoma degli interessati). Un'ombra di dubbio è spesso difficile da cancellare e l'indifferenza per le osservazioni dei critici può peggiorare la situazione. Un'analisi normativa di ampia portata potrebbe chiarire i fattori su cui si fonda la fiducia degli esperti nei vaccini e convincere gli scettici.

Le scelte di prevenzione e di trattamento del cancro non sono estranee a controversie (per esempio, in relazione a trattamenti in competizione, forme di medicina alternativa e nuovi test di controllo). Anche in quest'ambito potrebbe quindi essere d'aiuto l'approccio qui delineato, emerso in relazione a controversie su rischi di carattere ambientale e tecnologico in cui la mancanza di un confronto costruttivo con i critici ha spesso aggravato i sospetti e impedito l'identificazione di un terreno comune di discussione (Fischhoff, 1995).

2. All'inizio del 1976 si osservarono negli Stati Uniti alcuni casi di questa forma di influenza particolarmente aggressiva. Il governo reagì con un massiccio (e discusso) programma di vaccinazione a livello nazionale [N.d.C.].

4.5. *Rischi relativi a malattie infettive*

Quando le misure preventive falliscono, ci si può trovare a dover reagire a un'emergenza. Un'analisi normativa del problema potrebbe in questo caso consigliare l'attuazione (seppur tardiva) di quelle stesse misure preventive o l'adozione di interventi del tutto differenti. Per esempio, la campagna del 2003 per la vaccinazione contro il vaiolo negli Stati Uniti ha raggiunto solo una piccola parte dei suoi obiettivi. Se dovesse verificarsi un attacco terroristico, la vaccinazione sarebbe ancora un'opzione possibile (anche a seguito dell'esposizione), così come la quarantena. La definzione di una politica di intervento richiede l'elaborazione di modelli che fanno assunzioni realistiche sul comportamento degli individui (per esempio, le persone recepiranno un messaggio di allarme? ne comprenderanno il contenuto? confideranno nelle istruzioni che contiene? le metteranno in atto?). L'utilizzo dei mezzi di comunicazione di massa non garantisce la riuscita: per esempio, alla fine del 2002, a dispetto di un'estesa copertura mediatica del problema, i cittadini americani non avevano compreso che l'antrace non è contagioso, che il virus *West Nile* è raramente fatale o che il vaccino contro il vaiolo può funzionare anche a seguito dell'esposizione (Fischhoff *et al.*, 2003).

Il *Cryptosporidium* è un parassita trasmesso con l'acqua che le autorità sanitarie pubbliche conoscono bene e riguardo al quale spesso pianificano programmi di intervento. Secreto attraverso le feci dei mammiferi, può penetrare nelle riserve d'acqua a causa di forti precipitazioni piovose (per esempio, trasportato da rivoli provenienti da recinti per l'allevamento intensivo di bestiame). Nel 1993, a Milwaukee, un'epidemia ha colpito 400.000 persone uccidendone 100. Le persone affette da AIDS sono particolarmente a rischio. In caso di emergenza, l'impatto della comunicazione dipende dal suo raggio d'azione e dalla sua persuasività, che a loro volta risentono di fattori quali l'entità del rischio di penetrazione del *Cryptosporidium*, le procedure di sorveglianza, l'efficacia dei controlli, le tecnologie disponibili per il soccorso, le reazioni dei media, il livello di sensibilizzazione della popolazione e il coordinamento istituzionale (per esempio, fra i servizi di fornitura dell'acqua, le autorità pubbliche, le farmacie, ecc.). In uno studio incentrato sulla costruzione di modelli dei processi coinvolti in caso di emergenza, è stato possibile concludere che un messaggio di allarme tipicamente non avrebbe *alcun effetto*, anche se raggiungesse tutti gli interessati e venisse seguito scrupolosamente (Casman *et al.*, 2000). Il problema è che gli strumenti di controllo disponibili non consentono di rilevare la presenza del *Cryptosporidium* in modo sufficientemente tempestivo da permettere alle persone di reagire efficacemente. Lo stesso approccio potrebbe invece proteggere la popolazione da agenti patogeni rilevabili con test veloci ed economici (per esempio, l'*Escherichia Coli*), a condizione che gli altri elementi del sistema funzionino come

previsto (per esempio, che gli ingegneri comunichino efficacemente con la autorità sanitarie). In ogni caso, sembra chiaro che l'elaborazione esplicita di modelli dei diversi scenari possibili è una necessaria integrazione delle intuizioni di chi deve pianificare gli interventi.

In indagini descrittive basate su interviste approfondite in due comunità che erano state colpite dal *Cryptosporidium*, abbiamo rilevato un buon livello di conoscenza dei relativi rischi fra individui HIV-positivi e nessuna conoscenza fra gli altri residenti. Nessuno dei due gruppi di persone poteva comunque permettersi il consumo regolare di acqua in bottiglia. Se la penetrazione del *Cryptosporidium* non può essere efficacemente prevenuta, allora fornire acqua in bottiglia ai cittadini più vulnerabili è un investimento migliore che mettere a punto interventi di emergenza: nei casi in cui gli individui non hanno a disposizione valide possibilità di azione, l'elaborazione di una «soluzione» prescrittiva non dovrebbe assumere irrealisticamente che essi siano in condizioni di scegliere e di influenzare il corso degli eventi.

Nella prevenzione e nel trattamento del cancro, una classe di decisioni con proprietà simili a queste riguarda l'impiego dei farmaci ufficialmente approvati. Il fatto che i benefici offerti dai farmaci corrispondano a quanto è documentato dagli studi clinici dipende dal modo in cui i pazienti ne fanno uso: seguono l'intero corso delle cure? Evitano l'uso concomitante di cibi o di altri farmaci che, in base alle conoscenze disponibili, possono dar luogo a interazioni? Notano i sintomi che potrebbero suggerire controindicazioni? Notano e comprendono le avvertenze d'uso ricavate da indagini successive all'approvazione ufficiale del farmaco? L'analisi normativa, basata su assunzioni descrittive realistiche, è uno strumento imprescindibile per stimare l'efficacia delle prescrizioni e quindi aiutare i pazienti (e gli stessi medici) a prendere decisioni migliori in relazione ai farmaci contro il cancro[3].

5. Ostacoli per l'approccio strategico della ricerca cognitiva

Sebbene si possa sperare di portare a compimento la componente normativa, descrittiva e prescrittiva di ogni progetto di ricerca, anche l'esplorazione di uno solo di questi aspetti può influenzare gli altri in modo significativo: l'analisi normativa guida lo studio scientifico dei processi che determinano rischi e benefici in un problema decisionale; l'indagine descrittiva mette in evidenza le questioni alle quali le persone sono maggiormente interessate e

3. Durante la ricerca sul *Cryptosporidium*, mi capitò di leggere messaggi informativi rivolti a pazienti in chemioterapia. Questi messaggi dicevano ben poco su agenti patogeni trasmessi attraverso l'acqua. Un'analisi normativa potrebbe indicare se rischi di questo tipo sono più o meno importanti dei temi che di solito vengono trattati esplicitamente.

individua ciò che già sanno (e che non hanno bisogno di sentirsi ripetere); lo studio di un intervento prescrittivo mette alla prova la robustezza delle teorie descrittive e identifica le lacune più significative della ricerca in vista delle applicazioni pratiche.

I programmi di ricerca integrati affrontano però inevitabilmente una serie di ostacoli. Alcuni riguardano le comunità professionali coinvolte, altri le istituzioni finanziatrici (e il parziale controllo che esse esercitano). Nel corso del tempo, l'approccio normativo, quello descrittivo e quello prescrittivo si sono evoluti come aree di specializzazione separate nello studio delle decisioni. In ciascuno di questi ambiti è già sufficientemente impegnativo acquisire una formazione professionale, restare aggiornati e conservare la propria posizione. In linea di principio, la comunicazione dovrebbe essere facilitata dal comune riferimento alla teoria dell'utilità. In pratica, però, la routine dell'attività professionale riduce le occasioni e gli incentivi per la comunicazione e l'interazione. In effetti, è anche possibile che siano presenti alcuni disincentivi: al fine di rendere trattabile il proprio approccio a un problema, a ciascuno dei tre livelli si fanno alcune assunzioni semplificanti in relazione agli altri. Per esempio, gli studiosi impegnati nella ricerca descrittiva devono assumere che i compiti che studiano catturino alcune caratteristiche essenziali dei processi decisionali del mondo reale. Lo svolgimento di un'analisi normativa su questo tema non solo complica il loro lavoro, ma comporta anche una critica implicita di qualunque altro studio in cui il problema decisionale sia stato definito su basi intuitive.

Ostacoli simili emergono dalle altre scelte strategiche discusse in precedenza. Nel considerare una particolare dimensione di variabilità individuale, per esempio, i ricercatori mettono implicitamente in questione gli studi che l'hanno ignorata. Sebbene la scienza proceda estendendo l'analisi a nuovi fattori («che cosa succede se facciamo variare…?»), la sua retorica spesso tratta le omissioni come peccati («come hai potuto trascurare…?»). Su questo tema, un approccio integrato allo studio delle decisioni può contribuire a una visione equilibrata. In particolare, l'analisi normativa può chiarire l'effettiva importanza pratica dei risultati della ricerca descrittiva – come accade quando un fenomeno statisticamente e teoricamente significativo non ha effetti sulla qualità di molte scelte. Una fonte sistematica di errore può infatti conservarsi nel tempo proprio perché le sue conseguenze sono solitamente troppo modeste perché l'errore sia notato e corretto attraverso l'apprendimento. Col risultato che quella tendenza all'errore risulterà inalterata in quelle particolari occasioni in cui si rivelerà determinante.

Anche la scelta di un atteggiamento persuasivo o non persuasivo può creare ostacoli e barriere. Questa scelta talvolta riflette le personali preferenze del ricercatore: alcune persone amano orientare la vita altrui, mentre altre odiano le manipolazioni. In altri casi, la scelta riflette priorità di finanziamento (al-

cune istituzioni si occupano di persuadere, altre possono soltanto consigliare) o esigenze legate ai metodi di valutazione (è più facile rilevare i cambiamenti nella diffusione di un particolare comportamento che misurare il livello di qualità delle scelte degli individui). Infine, la scelta può riflettere la percezione di un bisogno, cioè la convizione che l'uno o l'altro dei due approcci non possa funzionare in relazione al problema trattato. È da notare che una definizione restrittiva del problema decisionale incoraggia la comunicazione persuasiva, volta a modificare uno specifico comportamento; anche perché la mancata considerazione di alcuni dei possibili obiettivi delle persone tende a farle apparire meno razionali (e quindi potenzialmente bisognose di un intervento persuasivo). Tutto ciò comporta il rischio che una certa istituzione che si interessa a un unico e specifico comportamento (per esempio, una dieta più adeguata), trascurando altre dimensioni rilevanti per chi decide (come gusti, costi, convenienza), incorra nella frustrazione dei propri sforzi e in una critica immotivata del pubblico («noi l'abbiamo detto, ma loro non hanno voluto ascoltare»). Infine, quale che sia la ragione della scelta, l'adozione di uno dei due approcci (persuasivo/non persuasivo) può complicare la comunicazione e il confronto dei risultati con ricercatori che adottano l'approccio alternativo.

Un ultimo ostacolo è rappresentato dal fatto che, in molti settori importanti, la nicchia della ricerca sulle decisioni è occupata da paradigmi alternativi. La «competizione» più forte è forse con l'utilizzo di modelli di tipo computazionale (come lo *health belief model*). Negli anni Settanta è emerso che questi modelli, rimasti sostanzialmente estranei alla ricerca cognitiva sulle decisioni, possono avere un significativo potere predittivo in relazione al comportamento umano anche se non riproducono accuratamente i processi psicologici che lo guidano, purché comprendano alcune variabili correlate con quei processi. La distinzione fra modelli computazionali fra loro alternativi è peraltro difficile a causa di alcuni problemi di misurazione (Dawes, 1979). Di conseguenza, è spesso possibile creare diversi modelli dotati di una certa efficacia predittiva, a condizione di tenere conto di alcuni aspetti fondamentali di una certa classe di decisioni. Le capacità predittive ed esplicative di tali modelli hanno però dei limiti. Proprio il riconoscimento di questi limiti ha indotto gli studiosi impegnati nella ricerca cognitiva a concentrarsi sui processi che intervengono nelle scelte (per esempio, la stima delle probabilità o l'elaborazione dei propri valori) e ad appoggiarsi alla comprensione di quei processi per la messa a punto di interventi prescrittivi. Lo scontro fra differenti paradigmi di ricerca è difficile da risolvere, o anche solo da affrontare, quando i ricercatori partecipano a convegni separati, pubblicano su riviste differenti e interagiscono limitatamente alla valutazione delle proposte di ricerca (con i redattori delle proposte che non riescono a spiegare, o a difendere, il loro paradigma di fronte a valutatori che lavorano nel paradigma alternativo).

6. Un approccio cognitivo alle decisioni sul cancro

Dal punto di vista strategico, lo studio cognitivo delle scelte di prevenzione e di trattamento del cancro richiede lo svolgimento di analisi normative. Ciò significa che è necessario identificare i fatti rilevanti per la decisione, determinare la loro importanza per coloro che devono decidere e individuare le scelte più opportune nel loro interesse (attraverso la teoria dell'utilità o una euristica efficiente; si veda Simon, 1957). Sebbene i dettagli differiscano molto, gli stessi principi normativi si applicano a decisioni che riguardano le medicine alternative, i test genetici, l'abbronzatura e i pericoli per pazienti in chiemioterapia (per esempio, parassiti trasmessi con l'acqua, o deficienze dietetiche). L'analisi normativa può essere finalizzata a compiti molto diversi, quali la definizione delle priorità comunicative per il consenso informato, l'identificazione di ragioni potenzialmente legittime per il rifiuto di un intervento medico o la stima dell'efficacia di istruzioni per maneggiare agenti chimici cancerogeni (Riley *et al.*, 2002).

Le altre scelte strategiche (ampiezza della definizione del problema decisionale, considerazione di differenze individuali, adozione di un approccio persuasivo) possono riflettere le personali preferenze dei ricercatori, così come le loro preoccupazioni pratiche. Un messaggio di portata limitata può essere opportuno se le possibilità di comunicazione sono ristrette; può essere invece fuori luogo se troppo conciso perché i destinatari ne comprendano le ragioni. Un intervento ad ampio raggio (per esempio, un messaggio radiofonico, o la distribuzione di un opuscolo in uno studio medico) può essere appropriato se le differenze individuali sono ridotte e rendono un unico strumento adatto a tutti; può invece essere inopportuno se risulta fuorviante per alcuni destinatari che si trovano in circostanze atipiche o se consuma risorse che potrebbero essere impiegate in interventi più mirati. Le differenze individuali rilevanti possono riguardare le competenze decisionali dei destinatari (il compito è troppo difficile per loro? o è stato eccessivamente semplificato?), i loro valori (sono stati trascurati aspetti che essi considerano importanti?) o le loro reazioni emotive (sono state evocate emozioni che essi desiderano evitare?).

Nei casi in cui i valori delle persone e le circostanze in cui si trovano sono molto variabili, è possibile che nessun singolo messaggio risulti giustificato. Se non c'è la possibilità di adeguare il messaggio ai diversi casi, si dovrebbe intervenire in modo non persuasivo, lasciando che gli interessati si incarichino di applicare le informazioni alla loro specifica situazione. È preferibile un approccio non persuasivo anche quando l'incertezza è tale che le scelte dipendono dalla disponibilità degli individui ad accettare una fra diverse possibili scommesse. Come si è detto, un approccio persuasivo presuppone che l'esperto conosca la situazione meglio di quanto le persone comuni possano mai sperare di conoscerla o che l'esperto conosca le persone interessate me-

glio di quanto loro conoscano se stesse. Non è impossibile che ciò avvenga, come nel caso di un paziente che domandi «dottore, che cosa voglio fare?». Occorre però sottolineare che, quando un individuo condivide (o delega) il controllo, deve essere presente un legame di fiducia: è necessario che egli consideri l'esperto competente e onesto, in possesso di abilità che egli non ha (o che non può esercitare) e intenzionato ad agire nel suo interesse. Gli interventi fondati su un approccio persuasivo possono quindi rivelarsi inappropriati se allontanano gli individui, usurpandone l'autonomia o negando loro la legittima possibilità di giungere alla comprensione dei problemi; mentre gli interventi fondati su un approccio non persuasivo possono rivelarsi inappropriati se negano agli individui protezioni di cui essi hanno bisogno, lasciandoli di fronte a «scelte impossibili» date le risorse personali che hanno a disposizione per comprenderle.

La definizione di un programma di ricerca produttivo sarebbe notevolmente semplificata se fosse possibile assumere un livello generale di competenza decisionale. Su questo tema è comune imbattersi in frettolose generalizzazioni che normalmente riflettono punti di vista interessati, se non vere e proprie prese di posizione strategiche. La tesi secondo cui il pubblico sarebbe incompetente a decidere giustifica istituzioni paternalistiche (per esempio, l'adozione di regolamenti restrittivi o il controllo tecnocratico). Per contro, la tesi per cui il pubblico sarebbe perfettamente competente a decidere giustifica istituzioni libertarie (per esempio, soluzioni basate sul mercato o su processi partecipativi). La strategia di fondo della ricerca cognitiva sulle decisioni prevede che l'adeguatezza dei processi decisionali (tanto degli esperti quanto delle persone comuni) dipenda dal problema di scelta, dagli individui e dal contesto. Nel loro complesso, le indicazioni della ricerca di base e applicata suggeriscono la seguente ipotesi di lavoro: le persone spesso scelgono in modo ragionevole, a condizione che (a) ricevano una descrizione chiara e credibile dei fatti, (b) la qualità delle loro scelte sia valutata sulla base dei loro reali obiettivi, (c) abbiano un qualche controllo su se stesse e sul loro ambiente, e (d) abbiano un livello minimo di competenza decisionale.

Per perseguire programmi di ricerca in grado di migliorare le scelte di prevenzione e di trattamento del cancro, è necessario innanzi tutto un impegno pratico a fornire le risorse che l'analisi formale e la ricerca empirica richiedono. È inoltre necessario un impegno filosofico volto all'espansione dell'autonomia decisionale degli individui, fin dove essa è possibile, pur nel riconoscimento dei limiti di tale autonomia (e della potenziale utilità di interventi di tipo persuasivo). È infine necessario un impegno politico per l'elaborazione e la diffusione di valutazioni imparziali degli effetti delle opzioni di prevenzione e trattamento, in modo che le persone abbiano pronto accesso alle informazioni di cui hanno bisogno per scegliere.

Riferimenti bibliografici

Afifi, A.A. (1990), *Computer-aided multivariate analysis*, Van Nostrand-Reinhold, New York.

Ajzen, I. (1982), «On behaving in accordance with one's attitudes», in Zanna, M.P., Higgins, E.T., Herman, C.P. (a cura di), *Consistency in social behavior: the Ontario Symposium* (vol. 2), Erlbaum, Hillsdale (NJ), pp. 3-15.

Ajzen, I., Fishbein, M. (1977), *Attitude-behavior relationship: a theoretical analysis and review of empirical research*, «Psychological bulletin», 84, pp. 888-918.

Allen, J.G. (1965), «Breast surgery, principles and practice», in Moyer, C.A. (a cura di), *Surgery: principles and practice*, Lippincott, Philadelphia.

Allman, R.M., Steinberg, E.P., Keruly, J.C., Dans, P.E. (1985), *Physician tolerance for uncertainty. Use of liver-spleen scans to detect metastases*, «Journal of the American Medical Association», 254, pp. 246-248.

American Child Health Association (1934), *Physical defects: the pathway to correction*, American Child Health Association, New York.

American Heart Association (1976), *Coronary risk handbook: estimating the risk factors for coronary heart disease in daily practice*, American Heart Association, New York.

Apple, F.S., Henderson A.R. (1999), «Cardiac function», in Burtis, C.A., Ashwood, E.R. (a cura di), *Tietz textbook of clinical chemistry*, Saunders, Philadelphia.

Arkes, H.R., Blumer, C. (1985), *The psychology of sunk cost*, «Organizational behavior and human decision processes», 35, pp. 125-140.

Arkes, H.R., Christensen, C., Lai, C., Blumer, C. (1987), *Two methods of reducing overconfidence*, «Organizational behavior and human decision processes», 39, pp. 133-144.

Arkes, H.R., Harkness, A.R. (1980), *The effect of making a diagnosis on subsequent recognition of symptoms*, «Journal of experimental psychology: human learning and memory», 6, pp. 568-575.

Aronson, E. (1978), «The theory of cognitive dissonance: a current perspective», in Berkovitz, L. (a cura di), *Cognitive theories in social psychology*, Academic press, New York.

Aronson, M.D., Bor, D.H. (1987), *Blood cultures*, «Annals of internal medicine», 106, pp. 246-453.

Arrow, K.J. (1982), *Risk perception in psychology and economics*, «Economic inquiry», 20, pp. 303-317.

Asch, D.A., Baron, J., Hershey, J.C., Kunreuther, H., Meszaros, J., Ritov, I., Spranca, M. (1994), *Omission bias and pertussis vaccination*, «Medical decision making», 14, pp. 118-123.
Asch, D.A., Hershey, J.C. (1995), *Why some health policies don't make sense at the bedside*, «Annals of internal medicine», 122, pp. 846-850.
Asch, D.A., Patton, J.P., Hershey, J.C. (1990), *Knowing for the sake of knowing: the value of prognostic information*, «Medical decision making», 10, pp. 47-57.
Axtell, L.M., Cutler, S.J., Myers, M.H. (1972) (a cura di), *End results in cancer (report n. 4)*, National Cancer Institute, Bethesda (MD).
Ayanian, J.Z., Berwick, D.M. (1991), *Do physicians have a bias toward action?*, «Medical decision making», 11, pp. 154-158.
Baddeley, A.D. (1979), «Applied cognitive and cognitive applied research», in Nilsson, L.G. (a cura di), *Perspectives on memory research*, Erlbaum, Hillsdale (NJ).
Baines, C.J., To, T., Wall, C. (1990), *Women's attitudes to screening after participation in the National Breast Screening Study. A questionnaire survey*, «Cancer», 65, pp. 1663-1669.
Bakwin, H. (1945), *Pseudodoxa pediatrica*, «New England journal of medicine», 232, pp. 691-697.
Banks, S.M., Salovey, P., Greener, S., Rothman, A.J., Moyer, A., Beauvais, J., Eppel, E. (1995), *The effects of message framing on mammography utilization*, «Health psychology», 14, pp. 178-184.
Bar-Hillel, M. (1983), «The base-rate fallacy controversy», in Scholtz, R.W. (a cura di), *Decision making under uncertainty*, Elsevier, North Holland.
Baron, J. (1998), *Judgment misguided: intuition and error in public decision making*, Oxfor University Press, New York.
Baron, J., Beattie, J., Hershey, J.C. (1988), *Heuristics and biases in diagnostic reasoning. II. Congruence, information, and certainty*, «Organizational behavior and human decision processes», 42, pp. 191-194.
Baron, J., Hershey, J.C. (1988a), *Outcome bias in decision evaluation*, «Journal of personality and social psychology», 54, pp. 569-579.
Baron, J., Hershey, J.C. (1988b), *Heuristics and biases in diagnostic reasoning. I. Priors, error costs, and test accuracy*, «Organizational behavior and human decision processes», 41, pp. 259-279.
Barron, F.J. (1953), *An ego-strength scale which predicts response to psycholterapy*, «Journal of consulting and clinical pathology», 17, pp. 327-333.
Bartlett, E.E. (1998), *Physicians' cognitive errors and their liability consequences*, «Journl of healthcare risk management», 18, pp. 62-69.
Bastardi, A., Shafir, E. (1998), *On the pursuit and misuse of useless information*, «Journal of personality and social psychology», 75, pp. 19-32.
Bastardi, A., Shafir, E. (2000), *Nonconsequential reasoning and its consequences*, «Current directions in psychological science», 9, pp. 216-219.
Bates, B. (1995), *A guide to physical examination and history taking*, Lippincott, Philadelphia.
Bates, D.W., Cook, E.F., Goldman, L., Lee, T.H. (1990), *Predicting bacteremia in hospitalized patients. A prospectively validated model*, «Annals of internal medicine», 113, pp. 495-500.
Beam, C.A., Layde, P.M., Sullivan, D.C. (1996), *Variability in the interpretation of screening mammograms by US radiologists. Findings from a national sample*, «Archives of internal medicine», 156, pp. 209-213.

Bedell, S.E., Delbanco, T.L., Cook, E.F., Epstein, F.H. (1983), *Survival after cardiopulmonary resuscitation in the hospital*, «New England journal of medicine», 309, pp. 569-576.

Bem, D.J. (1978), «Self-deception theory», in Berkovitz, L. (a cura di), *Cognitive theories in social psychology*, Academic press, New York.

Berwick, D.M., Fineberg, H.V., Weinstein, M.C. (1981), *When doctors meet numbers*, «American journal of medicine», 71, pp. 991-998.

Berwick, D.M., Thibodeau, L.A. (1983), *Receiver operating characteristic analysis of diagnostic skill*, «Medical care», 21, pp. 876-885.

Beyth-Marom, R., Fischhhoff, B. (1983), *Diagnosticity and pseudodiagnosticity*, «Journal of personality and social psychology», 45, pp. 1185-1195.

Blendon, R.J., DesRoches, C.M., Brodie, M., Benson, J.M., Rosen, A.B., Schneider, E., Altman, D.E., Zapert, K., Herrmann, M.J., Steffenson, A.E. (2002), *Views of practicing physicians and the public on medical errors*, «New England journal of medicine», 347, pp. 1933-1940.

Block, L.G., Keller, P.A. (1995), *When to accentuate the negative: the effects of perceived efficacy and message framing on intentions to perform a health-related behavior*, «Journal of marketing research», 32, pp. 192-203.

Block, M.A., Reynolds, W. (1974), *How vital is mammography in the diagnosis and management of breast carcinoma?*, «Archives of surgery», 108, pp. 588-591.

Bobbio, M., Demichelis, B., Giustetto, G. (1994), *Completeness of reporting trial results: effect on physicians' willingness to prescribe*, «Lancet», 343, pp. 1209-1211.

Boom, R., Chavez-Oest, J., Gonzalez, C., Cantu, M.A., Rivero, F., Reyes, A., Aguilar, E., Santamaria, J. (1988), *Physicians' diagnoses compared with algorithmic differentiation of causes of jaundice*, «Medical decision making», 8, pp. 177-181.

Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J. (1993), *Classification and regression trees*, Wadsworth, Belmont (CA).

Brendl, C.M., Higgins, E.T., Lemm, K. (1995), *Sensitivity to varying gains and losses: The role of self-discrepancies and event framing*, «Journal of personality and social psychology», 69, pp. 1028-1051.

Bucher, H.C., Weinbacher, M., Gyr, K. (1994), *Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration*, «British medical journal», 309, pp. 761-764.

Bundeszentrale für gesundheitliche Aufklärung (1988-1993) (a cura di), *Wissenswertes über den HIV-Test*, vol. 1-10, Bundeszentrale für gesundheitliche Aufklärung, Köln.

Burke, D.S., Brundage, J.F., Redfield, R.R., Damato, J.J., Schnabel, C.A., Putman, P., Visitine, R., Kim, H.J. (1988), *Measurement of the false positive rate in a screening program for human immunodeficiency virus infection*, «New England journal of medicine», 319, pp. 961-964.

Burt, R.W. (2000), *Colon cancer screening*, «Gastroenterology», 119, pp. 837-853.

Burton, E.C., Troxclair, D.A., Newman, W.P. (1998), *Autopsy diagnoses of malignant neoplasms: how often are clinical diagnoses incorrect?*, «Journal of the American Medical Association», 280, pp. 1245-1248.

Busch, M.P. (1994), «HIV testing in blood banks», in Schochetman, G., George, J.R. (a cura di), *AIDS testing: a comprehensive guide to technical, medical, social, legal, and management issues*, Springer, New York, pp. 224-236.

Byrne, R.R. (1974), *Correlation of thermography, xeromammography and biopsy in a community hospital*, «Wisconsin medical journal», 73, pp. 35-37.
Cacioppo, J.T., Petty, R.E. (1982), *The need for cognition*, «Journal of personality and social psychology», 42, pp. 116-131.
Camerer, C., Babcock, L., Loewenstein, G., Thaler, R. (1997), *Labor supply of New York City cab drivers: one day at a time*, «Quartely journal of economics», 112, pp. 407-441.
Caplan, R.A., Posner, K.L., Cheney, F.W. (1991), *Effect of outcome on physicians' judgments of appropriateness of care*, «Journal of the American Medical Association», 265, pp. 1957-1960.
Capute, A.J., Neidermeyer, E.F.L., Richardson, F. (1968), *Electroencephalogram in children with minimal cerebral dysfunction*, «Pediatrics», 41, pp. 1104-1114.
Carroll, J.S., Wiener, R.L., Coates, D., Galegher, J., Alibrio, J.J. (1982), *Evaluation, diagnosis and prediction in parole decision making*, «Law and society review», 17, pp. 199-228.
Casman, E., Fischhoff, B., Palmgren, C., Small, M., Wu, F. (2000), *Integrated risk model of a drinking waterborne* Cryptosporidiosis *outbreak*, «Risk analysis», 20, pp. 493-509.
Casscells, W., Schoenberger, A., Graboys, T. (1978), *Interpretation by physicians of clinical laboratory results*, «New England journal of medicine», 299, pp. 999-1000.
Catalan, J., Pugh, K. (1995), *Suicidal behaviour and HIV infection - is there a link?*, «AIDS care», 7, pp. 117-121.
Centor, R.M., Schwartz, J.S. (1983), *Use of an electronic spreadsheet (Visicalc) to estimate the area under the ROC curve*, «Medical decision making», 3, pp. 380.
Chalmers T.C., Smith, H. Jr., Blackburn, B., Silverman, B., Schroeder, B., Reitman, D., Ambroz, A. (1981), *A method for assessing the quality of a randomized control trial*, «Controlled clinical trials», 2, pp. 31-49.
Chapman, C.R., Casey, K.L., Dubner, R., Foley, K.M., Gracely, R.H., Reading, A.E. (1985), *Pain measurement: an overview*, «Pain», 22, pp. 1-31.
Chapman, L.J., Chapman, J.P. (1967), *Genesis of popular but erroneous psychodiagnostic observations*, «Journal of abnormal psychology», 72, pp. 193-204.
Chapman, L.J., Chapman, J.P. (1969), *Illusory correlation as an obstacle to the use of valid psychodiagnostic signs*, «Journal of abnormal psychology», 74, pp. 271-280.
Christensen-Szalanski, J.J.J. (1986), «Improving the practical utility of judgment research», in Brehmen, B., Jungermann, H., Lourens, P., Sevon, G. (a cura di), *New directions in research on decision making*, Elsevier, North Holland.
Christensen-Szalanski, J.J.J., Beach, L.R. (1982), *Experience and the base-rate fallacy*, «Organizational behavior and human decision processes», 29, pp. 270-279.
Christensen-Szalanski, J.J.J., Bushyhead, J.B. (1983), *Physicians' misunderstanding of normal findings*, «Medical decision making», 3, pp. 169-175.
Christensen-Szalanski, J.J.J., Fobian, C.S. (1988), *On the significance of judgment studies*, «Medical decision making», 8, pp. 265-267.
Clark R.L., Copeland, M.M., Egan, R.L., Gallagher, H.S., Geller, H., Lindsay, J.P., Robbins, L.C., White, E.C. (1965), *Reproducibility of the technic of mammography (Egan) for cancer of the breast*, «American journal of surgery», 109, pp. 127-133.
Clark, R.L., Robbins, L.C. (1965), *Mammography (Egan) in cancer of the breast*, «American journal of surgery», pp. 125-126.

Cohn, H.E. (1972), *Mammography in its proper perspective*, «Surgery, gynecology and obstetrics», 134, pp. 97-98.
Comroe, J.H. Jr. (1978), *The road from research to new diagnosis and therapy*, «Science», 200, pp. 931-937.
Cook, D.J. (1997), *Health professional decision making in the ICU: a review of the evidence*, «New horizons», 5, pp. 15-19.
Cook, R.J., Sackett, D.L. (1995), *The number needed to treat: a clinically useful measure of treatment*, «British medical journal», 310, pp. 452-454.
Cosmides, L., Tooby, J. (1996), *Are humans good statisticians after all? Rethinking some conclusions from the literature on judgment and uncertainty*, «Cognition», 58, pp. 1-73.
Council on Scientific Affairs (1989), *Harmful effects of ultraviolet radiation*, «Journal of the American Medical Association», 262, pp. 380-384.
Daniel, W.G., Mugge, A., Martin, R.P., Lindert, O., Hausmann, D., Nonnast-Daniel, B., Laas, J., Lichtlen, P.R. (1991), *Improvement in the diagnosis of abscesses associated with endocarditis by transesophageal echocardiography*, «New England journal of medicine», 324, pp. 795-800.
Dannenberg, A.L., McNeil, J.G., Brundage, J.F., Brockmeyer, R. (1996), *Suicide and HIV-infection*, «Journal of the American Medical Association», 276, pp. 1743-1746.
Dawes, R.M. (1971), *A case-study of graduate admission: application of three principles of human decision making*, «American psychologist», 26, pp. 180-188.
Dawes, R.M. (1979), *The robust beauty of improper linear models in decision making*, «American psychologist», 34, pp. 571-582.
Dawes, R.M., Corrigan, B. (1974), *Linear models in decision making*, «Psychological bulletin», 81, pp. 95-106.
Dawson, N.V., Connors, A.F. Jr., Speroff, T., Kemka, A., Shaw, P., Arkes, H.R. (1993), *Hemodynamic assessment in managing the critically ill: is physician confidence warranted?*, «Medical decision making», 13, pp. 258-266.
Dawson, N.V., Arkes, H.R. (1987), *Systematic errors in medical decision making: judgment limitations*, «Journal of general internal medicine», 2, pp. 183-187.
Dawson, N.V., Arkes, H.R., Siciliano, C., Blinkhorn, R., Lakshmanan, M., Petrelli, M. (1988), *Hindsight bias: an impediment to accurate probability estimation in clinicopathologic conferences*, «Medical decision making», 8, pp. 259-264.
DeGowin, E., DeGowin, R. (1969), *Bedside diagnostic examination*, Macmillan, New York.
del Regato, J.A. (1970), «Diagnosis, treatment and prognosis», in Ackerman, L.V. (a cura di), *Cancer*, Mosby, St. Louis.
DeLuca, J.T. (1974), *A statistical comparison study of patients undergoing breast biopsy at a community hospital over a 16-year period*, «Radiology», 112, pp. 315-318.
DerSimonian, R., Charette, L.J., McPeek, B., Mosteller, F. (1982), *Reporting on methods in clinical trials*, «New England journal of medicine», 306, pp. 1332-1337.
Deskin, W.C., Hoye, R.E. (2004), *Another look at medical error*, «Journal of surgucal oncology», 88, pp. 122-129.
DeSmet, A.A., Fryback, D.G., Thornbury, J.R. (1979), *A second look at the utility of radiographic skull examination for trauma*, «American journal of Roentgenology», 132, pp. 95-99.

Deutscher Bundestag (1988) (a cura di), *AIDS: Fakten und Konsequenzen. Progress report of the enquete committee of the 11th German Bundestag, 3/88*, Bonner Universitäts Buchdruckerei, Bonn.
Diener, E., Wirtz, D., Oishi, S. (2001), *End effects of rated life quality*, «Psychological science», 12, pp. 124-128.
Dietz, K. Seydel, J., Schwartländer, B. (1994), *Back-projection of German AIDS data using information on dates of tests*, «Statistics in medicine», 13, pp. 1991-2008.
Doherty, M.E., Mynatt, C.R., Tweney, R.D., Schiavo, M.D. (1979), *Pseudodiagnosticity*, «Acta psychologica», 43, pp. 111-121.
Doherty, M.E., Schiavo, M., Tweney, R., Mynatt, C. (1981), *The influence of feedback and diagnostic data on pseudodiagnositicity*, «Bulletin of the psychonomic society», 18, pp. 191-194.
Doll, L.S., Kennedy, M.B. (1994), «HIV counseling and testing: what is it and how well does it work?», in Schochetman, G., George, J.R. (a cura di), *AIDS testing: a comprehensive guide to technical, medical, social, legal, and management issues*, Springer, New York, pp. 302-319.
Downs, J.S., Murray, P., Bruine de Bruin, Penrose, J., Palmgren, C., Fischhoff, B. (2004), *Interactive video behavioral intervention to reduce adolescent females' STD risk: a randomized control trial*, «Social science and medicine», 59, pp. 1561-1572.
Dunagan, W.C., Woodward, R.S., Medoff, G., Gray, J.L. 3rd, Casabar, E., Smith, M.D., Lawrenz, C.A., Spitznagel, E. (1989), *Antimicrobial misuse in patients with positive blood cultures*, «American journal of medicine», 87, pp. 253-259.
Dunham, H.W., Meltzer, B.M. (1946), *Predicting length of ospitalization of mental patients*, «American journal of sociology», 52, pp. 123-131.
Eagly, A.H., Chaiken, S. (1993), *The psychology of attitudes*, Harcourt Brace Jovanovich, New York.
Eberle, J.F., Deinhardt, K.O., Habermehl, M.A. (1988), *Die Zuverlässigkeit des HIV-Antikörpertests*, «Deutsches Ärzteblatt», 85, pp. 1512-1514.
Egan, R.L. (1969), *Fundamentals of mammographic diagnoses of benign and malignant diseases*, «Oncology», 23, pp. 126-148.
Egan, R.L. (1971), *Contributions of mammography in the detection of early breast cancer*, «Cancer», 28, pp. 1555-1557.
Egan, R.L. (1972), *Mammography*, Charles T. Thomas, Springfield (IL).
Einhorn, H.J. (1972), *Expert judgment and mechanical combination*, «Organizational behavior and human decision processes», 7, pp. 86-106.
Einhorn, H.J., Hogarth, R.M. (1975), *Unit weighting schemas for decision making*, «Organizational behavior and human decision processes», 13, pp. 171-192.
Einhorn, H.J., Hogarth, R.M. (1978), *Confidence in judgment: persistence of the illusion of validity*, «Psychological review», 85, pp. 395-416.
Eisenberg, J.M. (1986), *Doctors' decisions and the cost of medical care*, Health Administration Press, Ann Arbor (MICH).
Eisenberg, R.L., Heineken, P., Hedgcock, M.W., Federle, M., Goldberg, H.I. (1982), *Evaluation of plain abdominal radiographs in the diagnosis of abdominal pain*, «Annals of internal medicine», 97, pp. 257-261.
Elstein, A.S. (1988), «Cognitive processes in clinical inference and decision making», in Turk, D.C., Salovey, P. (a cura di), *Reasoning, inference and judgment in clinical psychology*, The Free Press, New York, pp. 17-50.
Elstein, A.S., Shulman, L.S., Sprafka, S.A. (1978), *Medical problem solving: an analysis of clinical reasoning*, Harvard University Press, Cambridge (MA).

Elwyn, G., Edwards, A., Eccles, M., Rovner, D. (2001), *Decision analysis in patient care*, «Lancet», 358, pp. 571-574.
Epstein, J.S. (1994), «Regulation of HIV-related tests and procedures», in Schochetman, G., George, J.R. (a cura di), *AIDS testing: a comprehensive guide to technical, medical, social, legal, and management issues*, Springer, New York, pp. 52-61.
Eraker, S.A., Sox, H.C. (1981), *Assessment of patients' preferences for therapeutic outcomes*, «Medical decision making», 1, pp. 29-39.
Erskine, A., Morley, S., Pearce, S. (1990), *Memory for pain: a review*, «Pain», 41, pp. 255-265.
Espinosa, J.A., Nolan, T.W. (2000), *Reducing errors made by emergency physicians in interpreting radiographs: longitudinal study*, «British medical journal», 320, pp. 737-740.
Farkouh, M.E., Sauve, J.S., Kassam, H., Lee, H.N., Sackett, D.L. (1993), *Varying formats in which trial results are reported can affect clinical decision making (abstract)*, «Clinical research», 41, pp. 517.
Faust, D. (1984), *The limits of scientific reasoning*, University of Minnesota Press, Minneapolis (MN).
Feinstein A.R. (1992), *Invidious comparisons and unmet clinical challenges*, «American journal of medicine», 92, pp. 117-120.
Feinstein, A.R. (1983), *An additional basic science for clinical medicine: III. The challenges of comparison and measurement*, «Annals of internal medicine», 99, pp. 705-712.
Feinstein, A.R. (1985), «Statistical indexes in contrast», in *Clinical epidemiology: the architecture of clinical research*, Saunders, Philadelphia.
Feinstein, A.R. (1988), *Fraud, distortion, delusion, and consensus: the problems of human and natural deception in epidemiologic science*, «American journal of medicine», 84, pp. 475-478.
Festinger, L. (1957), *A theory of cognitive dissonance*, Row & Peterson, Evanston (IL).
Fetherstonhaugh, D., Slovic, P., Johnson, S.M., Friedrich, J. (1997), *Insensitivity to the value of human life: a study of psychophysical numbing*, «Journal of risk and uncertainty», 14, pp. 282-300.
Filice, G.A., Weiler, M.D., Hughes, R.A., Gerding, D.N. (1989), *Nosocomial febrile illnesses in patients on an internal medicine service*, «Archives of internal medicine», 149, pp. 319-324.
Fillingim, R.B. (2000), *Sex, gender, and pain: women and men really are different*, «Current review of pain», 4, pp. 24-30.
Fischer, D. (1990), *AIDS-Modellprogramm der Bundesregierung – eine Zwischenbilanz*, «Öffentliches Gesundheitswesen», 52, pp. 425-431.
Fischer, J.E., Steiner, F., Zucol, F., Berger, C., Martignon, L., Bossart, W., Altwegg, M., Nadal, D. (2002), *Use of simple heuristics to target macrolide prescription in children with community-acquired pneumonia*, «Archives of pediatrics and adolescent medicine», 156, pp. 1005-1008.
Fischhoff, B. (1975), *Hindsight ≠ foresight: the effect of outcome knowledge on judgment under uncertainty*, «Journal of experimental psychology: human perception and performance», 1, pp. 288-299.
Fischhoff, B. (1980), «For those condemned to study the past», in Schweder, R.A., Fiske, D.W. (a cura di), *New Directions for Methodology of Social and Behavioral Science*, Jossey-Bass, San Francisco (CA), pp. 79-93.

Fischhoff, B. (1992), *Giving advice: decision theory perspectives on sexual assault*, «American psychologist», 47, pp. 577-588.
Fischhoff, B. (1995), *Risk perception and communication unplugged: twenty years of process*, «Risk analysis», 15, pp. 137-145.
Fischhoff, B., Beyth-Marom (1975), *«I knew it would happen» – remembered probabilities of once-future things*, «Organizational behavior and human decision processes», 13, pp. 1-16.
Fischhoff, B., Bostrom, A., Quadrel, M.J. (2002), «Risk perception and communication», in Detels, R., McEwen, J., Beaglehole, R., Tanaka, H. (a cura di), *Oxford textbook of public health*, Oxford University Press, London, pp. 1105-1123.
Fischhoff, B., Downs, J., Bruine de Bruin, W. (1998), *Adolescent vulnerability: a framework for behavioral interventions*, «Applied and preventive psychology», 7, pp. 77-94.
Fischhoff, B., Gonzalez, R., Small, D., Lerner, J. (2003), *Evaluating the success of terror risk communication*, «Biosecurity and bioterrorism», 1, pp. 255-258.
Fischhoff, B., Slovic, P., Lichtenstein, S. (1980), «Knowing what you want: measuring labile values», in Wallsten, T. (a cura di), *Cognitive processes in choice and decision behavior*, Erlbaum, Hillsdale (NJ), pp. 117-141.
Fitzgerald, R. (2001), *Error in radiology*, «Clinical radiology», 56, pp. 938-946.
Foucar, E., Foucar, M.K. (2000), «Error in anatomic pathology», in Foucar, M.K. (a cura di), *Bone marrow pathology*, ASCP Press, Chicago.
Frazier, A.L., Colditz, G.A., Fuchs, C.S., Kuntz, K.M. (2000), *Cost-effectiveness of screening for colorectal cancer in the general population*, «Journal of the American Medical Association», 284, pp. 1954-1961.
Frederickson, B.K., Kahneman, D. (1993), *Duration neglect in retrospective evaluations of affective episodes*, «Journal of personality and social psychology», 65, pp. 45-55.
Friedman, A.K., Ashovitz, S.I., Berger, S.M., Dodd, M.D., Fisher, M.S., Lapayowker, M.S., Moore, J.P., Parlee, D.E., Stein, G.N., Pendergrass, E.P. (1966), *A cooperative evaluation of mammography in seven teaching hospitals*, «Radiology», 86, pp. 886-891.
Fries, J.F., Bloch, D.A., Sharp, J.T., McShane, D.J., Spitz, P., Bluhm, G.B., Forrester, D., Genant, H., Gofton, P., Richman, S., *et al*. (1986), *Assessment of radiologic progression in rheumatoid arthritis. A randomized, controlled trial*, «Arthritis and rheumatism», 29, pp. 1-9.
George, J.R., Schochetman, G. (1994), «Detection of HIV infection using serologic techniques», in Schochetman, G., George, J.R. (a cura di), *Aids testing: a comprehensive guide to technical, medical, social, legal, and management issues*, Springer, New York, pp. 62-102.
Gershon-Cohen, J., Borden, A.G.G. (1964), *Biopsy and mammography*, «New York State journal of medicine», pp. 2751-2756.
Gigerenzer, G. (1996), *The psychology of good judgment: frequency formats and simple algorithms*, «Medical decision making», 16, pp. 273-280.
Gigerenzer, G., Hoffrage, U. (1995) *How to improve Bayesian reasoning without instruction: frequency formats*, «Psychological review», 102, pp. 684-704.
Gigerenzer, G., Hoffrage, U. (1997), *How to improve diagnostic inferences in physicians*, manoscritto.
Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., Krüger, L. (1989), *The empire of chance: how probability changed science and everyday life*, Cambridge University Press, Cambridge (MA).

Gigerenzer, G., Todd, P.M., ABC Research Group (1999), *Simple heuristics that make us smart*, Oxford University Press, New York.
Gilovich, T., Griffin, D., Kahneman, D. (2002) (a cura di), *Heuristics and biases: the psychology of intuitive judgment*, Cambridge University Press, New York.
Glück, D., Vornwald, A., Grossau, E., Kubanek, B. (1990), *HIV prevalence in blood donors in urban and rural areas of the Federal Republic of Germany*, «Blut», 60, pp. 304-307.
Goitein, M. (1990), *Waiting patiently*, «New England journal of medicine», 323, pp. 604-608.
Gold, R.H. (1969), *The accuracy of mammography in the diagnosis of benign and malignant disease*, «Oncology», 23, pp. 159-163.
Goldberg, L.R. (1965), *Diagnosticians* vs. *diagnostic signs: the diagnosis of psychosis* vs. *neurosis from the MMPI*, «Psychological monographs», 79.
Goldberg, L.R. (1968), *Simple models or simple processes? Some research on clinical judgments*, «American psychologist», pp. 483-496.
Goldberg, L.R. (1970), *Man* versus *model of man: a rationale, plus some evidence, for a method of improving on clinical inferences*, «Psychological bulletin», 73, pp. 422-432.
Goldberg, R.J., Gore, J.M., Alpert, J.S., Osganian, V., de Groot, J., Bade, J., Chen, Z., Frid, D., Dalen, J.E. (1991), *Cardiogenic shock after acute myocardial infarction. Incidence and mortality from a community-wide perspective, 1975 to 1988*, «New England journal of medicine», 325, pp. 1117-1122.
Goldman, B. (1989), *Science and finance face-off in battle over colon cancer screening*, «American college of physicians observer», 1989, 9, pp. 10-11.
Goldman, L., Cook, E.F., Brand, D.A., Lee, T.H., Rouan, G.W., Weisberg, M.C., Acampora, D., Stasiulewicz, C., Walshon, J., Terranova, G., *et al*. (1988), *A computer protocol to predict myocardial infarction in emergency department patients with chest pain*, «New England journal of medicine», 318, pp. 797-803.
Goldman, L., Weinberg, M., Weisberg, M., Olshen, R., Cook, E.F., Sargent, R.K., Lamas, G.A., Dennis, C., Wilson, C., Deckelbaum, L., Fineberg, H., Stiratelli, R. (1982), *A computer-derived protocol to aid in the diagnosis of emergency room patients with acute chest pain*, «New England journal of medicine», 307, pp. 588-596.
Gorter, S., van der Heijde, D.M.F.M., van der Linden, S., Houben, H., Rethans, J.-J., Scherpbier, A.J.J.A., van der Vleuten, C.P.M. (2002), *Psoriatic arthritis: performance of rheumatologists in daily practice*, «Annals of rheumatology disease», 61, pp. 219-224.
Gottman, J.M., Levenson, R.W. (1985), *A valid procedure for obtaining self-report of affect in marital interaction*, «Journal of consulting and clinical psychology», 53, pp. 151-160.
Graber, M. (2005), *Diagnostic errors in medicine: a case of neglect*, «Joint commission journal on quality and patient safety», 31, pp. 106-113.
Green, L., Mehr, D.R. (1997), *What alters physicians' decisions to admit to the coronary care unit?*, «Journal of family practice», 45, pp. 219-226.
Green, L.A., Becker, M.P. (1994), *Physician decision making and variation in hospital admission rates for suspected acute cardiac ischemia. A tale of two towns*, «Medical care», 32, pp. 1086-97.
Greenwald, A.G., Pratkanis, A.R., Leippe, M.R., Baumgardner, M.H. (1986), *Under what conditions does theory obstruct research progress?*, «Psychological review», 93, pp. 216-229.

Grice, H.P. (1975), «Logic and conversation», in Cole, P., Morgan, J.L. (a cura di), *Syntax and semantics. III. Speech acts*, Academic press, New York, pp. 41-58.
Griffith, J.R., Wilson, P.A., Wolfe, R.A., Bischak, D.P. (1985), *Clinical profiles of hospital discharge rates in local communities*, «Health services research», 20, pp. 131-151.
Grove, W.M., Meehl, P.E. (1996), *Comparative efficiency of informal (subjective, impressionistic) and formal (mechanical, algorithmic) prediction procedures: the clinical-statistical controversy*, «Psychology, public policy, and law», 2, pp. 1-31.
Grove, W.M., Zald, D.H., Lebow, B.S., Snitz, B.E., Nelson, C. (2000), *Clinical* vs. *mechanical prediction: a metaanalysis*, «Psychological assessment», 12, pp. 19-30.
Guilmette, T.J., Faust, D., Hart, K., Arkes, H.R. (1990), *A national survey of psychologists who offer neuropsychological services*, «Archives of clinical neuropsychology», 5, pp. 373-392.
Guyatt, G.H., Sackett, D.L., Cook, D.J. for the Evidence-Based Working Group (1994), *Users' guides to the medical literature. 11. How to use an article about therapy or prevention, B: what are the results and will they help me in caring for my patients?*, «Journal of the American Medical Association», 271, pp. 59-63.
Halbower, C.C. (1955), *A comparison of actuarial* vs. *clinical prediction to classes discriminated by the MMPI*, Dissertazione di dottorato, University of Minnesota.
Haley, R.W., Quade, D., Freeman, H., Bennett,, J.V. (1980) *CDC SENIC planning committee. Appendix E: algorithms for diagnosing infections*, «American journal of epidemiology», 111, pp. 635-643.
Ham, C., Hunter, D.J., Robinson, R. (1995), *Evidence based policymaking*, «British medical journal», 310, pp. 71-72.
Hammond, K.R., Summers, D.A. (1965), *Cognitive dependence on linear and nonlinear cues*, «Psychological review», 72, pp. 215-224.
Hanley, J.A., McNeil, B.J. (1982), *The meaning and use of the area under a ROC curve*, «Radiology», 143, pp. 29-36.
Harris, T., Cook, E.F., Kannel, W.B., Goldman, L. (1988), *Proportional hazards analysis of risk factors for coronary heart disease in individuals aged 65 or older. The Framingham Heart Study*, «Journal of the American geriatrics society», 36, pp. 1023-1028.
Hastie, R., Dawes, R.M. (2001), *Rational choice in an uncertain world*, Sage, Thousand Oaks (CA).
Haynes, R.B., Sackett, D.L., Tugwell, P. (1983), *Problems in the handling of clinical and research evidence by medical practitioners*, «Archives of internal medicine», 143, pp. 1971-1975.
Hayward, J. (1994), *Purchasing clinically effective care*, «British medical journal», 309, pp. 823-824.
Higbee, KL. (1969), *Fifteen years of fear arousal: research on threat appeals*, «Psychological bulletin», 72, pp. 426-444.
Higgins, E.T., Roney, C.J.R., Crowe, E., Hymes, C. (1994), *Ideal versus ought predilections for approach and avoidance: distinct self-regulatory systems*, «Journal of personality and social psychology», 66, pp. 276-286.
Higgins, E.T., Tykocinski, O. (1992), *Self-discrepancies and biographical memory: personality and cognition at the level of psychological situation*, «Personality and social psychology bulletin», 18, pp. 527-535.

Hilton, G. (1960), *The present position relating to cancer of the lung. Results with radiotherapy alone*, «Thorax», 15, pp. 17-18.
Hoffman-Valentin, F. (1991), *AIDS. Gefahren, Schutz, Vorsorge, Behandlungsmöglichkeiten*, Ecomed, Landsberg.
Hoffrage, U., Gigerenzer, G. (1998), *Using natural frequencies to improve diagnostic inferences*, «Academic medicine», 73, pp. 538-540.
Hogarth, R.M. (1975), *Cognitive processes and the assessment of subjective probability distributions*, «Journal of the American Statistical Association», 70, pp. 271-294.
Hourihane, J. (1994), *CT scans and the common cold?*, «New England journal of medicine», 330, pp. 1826.
Huber, J., Payne, J.W., Puto, C. (1982), *Adding asymmetrically dominated alternatives: violations of regularity and the similarity hypothesis*, «Journal of consumer research», 9, pp. 90-98.
Hulley, S.B., Cummings, S.R. (1988), *Designing clinical research*, Williams & Wilkins, Baltimore (MD).
Hux, J., Naylor, C.D. (1994), *Effect of coronary artery bypass surgery on survival (letter)*, «Lancet», 344, pp. 1223-1224.
Hux, J.E., Levington, C.M., Naylor, C.D. (1994), *Prescribing propensity: influence of life-expectancy gains and drug costs*, «Journal of general internal medicine», 9, pp. 195-201.
Hux, J.E., Naylor, C.D. (1995), *Communicating the benefits of chronic preventive therapy: does the format of efficacy data determine patients' acceptance of treatment?*, «Medical decision making», 15, pp. 152-157.
Hypertension detection and follow-up program cooperative group (1979), *Five-year findings of the hypertension detection and follow-up program. I. Reduction in mortality of persons with high blood pressure, including mild hypertension*, «Journal of the American Medical Association», 242, pp. 2562-2571.
Imperiale, T.F., Wagner, D.R., Lin, C.Y., Larkin, G.N., Rogge, J.D., Ransohoff, D.F. (2000), *Risk of advanced proximal neoplasms in asymptomatic adults according to the distal colorectal findings*, «New England journal of medicine», 343, pp. 169-174.
Ingelfinger, J.A., Mosteller, F., Thibodeau, L.A., Ware, J.H. (1983) (a cura di), *Biostatistics in clinical medicine*, Macmillan, New York.
Institute of Health Sciences (1994), *Critical appraisal skills for purchasers*, IHS, Oxford (Report to the NHS Executive).
Jayes, R.L. Jr., Beshansky, J.R., D'Agostino, R.B., Selker, H.P. (1992), *Do patients' coronary risk factor reports predict acute cardiac ischemia in the emergency department? A multicenter study*, «Journal of clinical epidemiology», 45, pp. 621-626.
Jones, R.A., Howard, P.H., Haley, J.V. (1984), *Probability distortions and outcome desirability: experimental verification of medical folklore*, «Journal of applied social psychology», 14, pp. 319-333.
Jonides, J., Jones, C.M. (1992), *Direct coding for frequency of occurrence*, «Journal of experimental psychology: learning, memory and cognition», 18, pp. 368-378.
Kachelmeier, S.J., Shehata, M. (1992), *Examining risk preferences under high monetary incentives: experimental evidence from People's Republic of China*, «American economic review», 82, 1120-1141.

Kahneman, D. (2000a), «New challenges to the rationality assumption», in Kahneman, D., Tversky, A. (a cura di), *Choices, values and frames*, Cambridge University Press, New York, pp. 758-774.
Kahneman, D. (2000b), «Experiences utility and objective happiness: a moment-based approach», in Kahneman, D., Tversky, A. (a cura di), *Choices, values and frames*, Cambridge University Press, New York, pp. 673-692.
Kahneman, D. (2000c), «Evaluation by moments: past and future», in Kahneman, D., Tversky, A. (a cura di), *Choices, values and frames*, Cambridge University Press, New York, pp. 693-708.
Kahneman, D. (2002), «Mappe di razionalità limitata», trad. it. in Motterlini, M., Piattelli Palmarini, M. (a cura di), *Critica della ragione economica*, Il Saggiatore, Milano, 2005.
Kahneman, D., Fredrickson, B.L., Schreiber, C.A., Redelmeier, D.A. (1993), *When more pain is preferred to less: adding a better end*, «Psychological science», 8, pp. 255-258.
Kahneman, D., Slovic, P., Tversky, A. (1982) (a cura di), *Judgment under uncertainty: heuristics and biases*, Cambridge University Press, New York.
Kahneman, D., Tversky, A. (1973), *On the psychology of prediction*, «Psychological review», 80, pp. 237-251.
Kahneman, D., Tversky, A. (1979), «*Prospect theory:* un'analisi delle decisioni in condizioni di rischio», trad. it. in Motterlini, M., Guala, F. (a cura di), *Introduzione all'economia cognitiva e sperimentale*, UBE, Milano, 2005, pp. 63-95.
Kahneman, D., Tversky, A. (1982), *The psychology of preferences*, «Scientific American», 247, pp. 160-173.
Kahneman, D., Tversky, A. (1984), *Choices, values, and frames*, «American psychologist», 39, pp. 341-350.
Kahneman, D., Tversky, A. (2000) (a cura di), *Choices, values and frames*, Cambridge University Press, New York.
Kalichman, S.C., Coley, B. (1995), *Context framing to enhance HIV-antibody-testing messages targeted to African American women*, «Health psychology», 14, pp. 247-254.
Kannel, W.B. (1976), *Some lessons in cardiovascular epidemiology from Framingham*, «American journal of cardiology», 37, pp. 269-282.
Kassirer, J.P. (1983), *Adding insult to injury: usurping patients' prerogatives*, «New England journal of medicine», 308, pp. 898-901.
Kassirer, J.P. (1989), *Our stubborn quest for diagnostic certainty: a cause of excessive testing*, «New England journal of medicine», 320, pp. 1489-1491.
Kassirer, J.P., Kopelman, R.I. (1989), *Cognitive errors in diagnosis: instantiation, classification, and consequences*, «American journal of medicine», 86, pp. 433-441.
Katsikopoulos, K., Machery, E., Pachur, T., Wallin, A. (2005), *The search for models of clinical judgment: fast, frugal, and friendly in Paul Meehl's spirit*, http://jeannicod.ccsd.cnrs.fr/documents/disk0/00/00/05/33/.
Katz, J., Melzack, R. (1999), *Measurement of pain*, «Surgical clinics of North America», 79, pp. 231-252.
Keeney, R.L., Raiffa, H. (1976), *Decision with multiple objectives: preferences and value trade-offs*, Wiley, New York.
Keesling, B., Friedman, H.S. (1987), *Psychosocial factors in sunbathing and sunscreen use*, «Health psychology», 6, pp. 477-493.

Keffer, J.H. (1996), *Myocardial markers of injury. Evolution and insights*, «American journal of clinical pathology», 1996,105, p. 305-320.
Kent, G. (1985), *Memory of dental pain*, «Pain», 21, pp. 187-194.
Kern, L., Doherty, M.E. (1982), *«Pseudodiagnosticity» in an idealized medical problem-solving environment*, «Journal of medical education», 57, pp. 100-104.
Knaus, W.A., Draper, E.A., Wagner, D.P., Zimmerman, J.E. (1985), *APACHE II: a severity of disease classification system*, «Critical care medicine», 13, pp. 818-829.
Koenker, R.M., DeLuca, S.A. (1989), *Vertebral osteomyelitis*, «American family physician», 39, pp.169-174.
Kreger, B.E., Craven, D.E., McCabe, W.R. (1980), *Gram-negative bacteremia. IV. Re-evaluation of clinical features and treatment in 612 patients*, «American journal of medicine», 68, pp. 344-355.
Kunin, C.M. (1988), *Cost-effective use of antibiotics*, «Archives of internal medicine», 148, pp. 1709-1710.
Kushner, R. (1976), *Breast cancer – «The night I found out»*, «San Francisco chronicle», 24 marzo 1976.
Laupacis, A., Naylor, C.D., Sackett, D.L. (1992), *How should the results of clinical trials be presented to clinicians? (abstract)*, «American college of physicians journal club», 12, p. 4.
Laupacis, A., Sackett, D.L., Roberts, R.S. (1988), *An assessment of clinically useful measures of the consequences of treatment*, «New England journal of medicine», 318, pp. 1728-1733.
Leape, L.L. (1994), *Error in medicine*, «Journal of the American Medical Association», 272, pp. 1851-1857.
LeBlond, R.F. (1989), *Improving structured abstracts*, «Annals of internal medicine», 111, p. 764.
Lee, K.L., Pryor, D.B., Harrell, F.E. Jr, Califf, R.M., Behar, V.S., Floyd, W.L., Morris, J.J., Waugh, R.A., Whalen, R.E., Rosati, R.A. (1986), *Predicting outcome in coronary disease. Statistical models* versus *expert clinicians*, «American journal of medicine», 80, pp. 553-560.
Lee, T.H., Ting, H.H., Shammash, J.B., Soukup, J.R., Goldman, L. (1992), *Long-term survival of emergency department patients with acute chest pain*, «American journal of cardiology», 69, pp. 145-151.
Leli, D.A., Filskov, S.B. (1981), *Clinical-actuarial detection and description of brain impairment with the W-B form I*, «Journal of clinical psychology», 37, pp. 623-629.
Leli, D.A., Filskov, S.B. (1984), *Clinical detection of intellectual deterioration associated with brain damage*, «Journal of clinical psychology», 40, pp. 1435-1441.
Lesnick, G.J. (1966), *Mammography: a word of caution*, «New York State journal of medicine», 2005-2008.
Leventhal, H. (1970), «Findings and theory in the study of fear communications», in Berkowitz, L. (a cura di), *Advances in experimental social psychology* (Vol. 5), Academic Press, New York, pp. 1208-1224.
Levi, K. (1985), *A signal detection framework for the evaluation of probabilistic forecast*, «Organizational behavior and human decision processes», 36, pp. 143-166.
Lichtenstein, S., Fischhoff, B., Phillips, L.D. (1980), «Calibration of probabilities: the state of the art to 1980», in Kahneman, D., Slovic, P., Tversky, A. (1982) (a cura di), *Judgment under uncertainty: heuristics and biases*, Cambridge University Press, New York.

Lieberman, D.A., Weiss, D.G., Bond, J.H., Ahnen, D.J., Garewal, H., Chejfec, G. (2000), *Use of colonoscopy to screen asymptomatic adults for colorectal cancer. Veterans Affairs Cooperative Study Group 380*, «New England journal of medicine», 343, pp. 162-168.

Likert, R. (1932), *A technique for the measurement of attitudes*, «Archives of Psychology», 22, pp. 1-55.

Lipid research clinics coronary primary prevention trial (1984), *The lipid research clinics coronary primary prevention trial results I. Reduction in incidence of coronary heart disease*, «Journal of the American Medical Association», 251, pp. 351-364.

Lipsky, B.A. (1988), *Fever in medical service patients*, «Journal of general internal medicine», 3, pp. 201-202.

Loewenstein, G., Lerner J.S. (2003), «The role of affect in decision making», in Davidson, R., Scherer, K., Goldsmith, H. (a cura di), *Handbook of Affective Sciences*, Oxford University Press, New York, pp. 619-642.

Lord, C.G., Lepper, M.R., Preston, E. (1984), *Consider the opposite: a corrective strategy of social judgment*, «Journal of personality and social psychology», 47, pp. 1231-1243.

Lusted, L.B. (1968), *Introduction to medical decision making*, Charles T. Thomas, Springfield (IL).

Lusted, L.B., Bell, R.S., Edwards, W., Roberts, H.V., Wallace, D.L. (1977), «Evaluating the efficacy of radiological procedures by Bayesian methods: a progress report», in Snapper, K.J. (a cura di), *Models in metrics for decision makers*, Information Resources Press, Washington (DC).

Lyon, D., Slovic, P. (1976), *Dominance of accuracy information and neglect of base rates in probability estimation*, «Acta psychologica», 40, pp. 287-298.

Lyons, A.S. (1975), *Aids to the surgeon by mammography*, «Mt. Sinai journal of medicine. New York», 42, pp. 223-231.

MacMahon, B., Pugh, T.F. (1970), *Epidemiology*, Little Brown and Company, Boston (MA).

Maheswaran, D., Meyers-Levy, J. (1990), *The influence of message framing and issue involvement*, «Journal of marketing research», 27, pp. 361-367.

Makadon, H.J., Bor, D., Friedland, G., Dasse, P., Komaroff, A.L., Aronson, M.D. (1987), *Febrile inpatients: house officers' use of blood cultures*, «Journal of general internal medicine», 2, pp. 293-297.

Malenka, D.J., Baron, J.A., Johansen, S., Wahrenberger, J.W., Ross, J.M. (1993), *The framing effect of relative and absolute risk*, «Journal of general internal medicine», 8, pp. 543-548.

Månsson, S.A. (1990), *Psycho-social aspects of HIV testing – the Swedish case*, «AIDS care», 2, pp. 5-16.

Martin, W.B., Apostolakos, P.C., Roazen, H. (1960), *Clinical* versus *actuarial prediction in the differential diagnosis of jaundice. A study of the relative accuracy of predictions made by physicians and by a statistically derived formula in differentiating parenchymal and obstructive jaundice*, «American journal of medical science», 240, pp. 571-578.

Marzuk, P.M., Perry, S.W. (1993), *Suicide and HIV: researchers and clinicians beware*, «AIDS care», 5, pp. 387-390.

Maurer, C., Kiehl, W., Altmann, D. (1993), «Zur Prävalenz und HIV-Inzidenz bei Blutspendern in Baden-Württemberg», in Kretschmer, V., Stangel, W., Eckstein,

R. (a cura di), *Transfusionsmedizin 1992/93* («Beiträge zur Infusionstherapie und Transfusionsmedizin», 31), Karger, Freiburg, pp. 5-9.

McCabe, W.R., Jackson, G.G. (1962), *Gram-negative bacteremia. II. Clinical, laboratory and therapeutic observations*, «Archives of internal medicine», 110, pp. 856-864.

McCarthy, B.D., Beshansky, J.R., D'Agostino, R.B., Selker, H.P. (1993), *Missed diagnoses of acute myocardial infarction in the emergency department: results from a multicenter study*, «Annals of emergency medicine», 22, pp. 579-582.

McClow, M.V., Williams, A.C. (1973), *Mammographic examinations (4030): ten-year clinical experience in a community medical center*, «Annals of surgery», 177, pp. 616-619.

McNeil, B.J., Weichselbaum, R., Pauker, S.G. (1978), *Fallacy of the five-year survival in lung cancer*, «New England journal of medicine», 299, pp. 1397-1401.

Meehl, P.E. (1954), *Clinical* versus *Statistical Prediction*, University of Minnesota Press, Minneapolis.

Meehl, P.E. (1957), *When shall we use our heads instead of the formula?*, «Journal of counseling psychology», 4, pp. 268-273.

Meehl, P.E. (1959), *A comparison of clinicians with five statistical methods of identifying psychotic MMPI profiles*, «Journal of counseling psychology», 6, pp. 102-109.

Meehl, P.E. (1973), *Psychodiagnosis: selected papers*, University of Minnesota Press, Minneapolis (MN).

Meehl, P.E. (1986), *Causes and effects of my disturbing little book*, «Journal of personality assessment», 50, pp. 370-375.

Mellors, J.W., Horwitz, R.I., Harvey, M.R., Horwitz, S.M. (1987), *A simple index to identify occult bacterial infection in adults with acute unexplained fever*, «Archives of internal medicine», 147, pp. 666-671.

Merz, J., Fischhoff, B., Mazur, D.J., Fischbeck, P.S. (1993), *Decision-analytic approach to developing standards of disclosure for medical informed consent*, «Journal of toxics and liability», 15, pp. 191-215.

Meyerowitz, B.E., Chaiken, S. (1987), *The effect of message framing on breast self-examination attitudes, intentions, and behavior*, «Journal of personality and social psychology», 52, pp. 500-510.

Meyers, B.R., Sherman, E., Mendelson, M.H., Velasquez, G., Srulevitch-Chin, E., Hubbard, M., Hirschman, S.Z. (1989), *Bloodstream infections in the elderly*, «American journal of medicine», 86, pp. 379-384.

Morgan, M.G., Fischhoff, B., Bostrom, A., Atman, C. (2001), *Risk communication: the mental models approach*, Cambridge University Press, New York.

Morlock, H. (1967), *The effect of outcome desirability on information required for decisions*, «Behavioral science», 12, pp. 296-300.

Moses, L.E., Emerson, J.D., Hosseini, H. (1986), «Analyzing data from ordered categories», in Bailar, J.C. III, Mosteller, F. (a cura di), *Medical uses of statistics*, New England journal of medicine books, Waltham (MA), pp. 235-258.

Motterlini, M., Piattelli Palmarini, M. (2005) (a cura di), *Critica della ragione economica. Tre saggi: McFadden, Kahneman, Smith*, Il Saggiatore, Milano.

Mountain, C.F., Carr, D.T., Anderson, W.A. (1974), *A system for the clinical staging of lung cancer*, «American journal of Roentgenology radium therapy and nuclear medicine», 120, pp. 130-138.

Mountain, C.T. (1976), «The relationship of prognosis to morphology and the anatomic extent of disease: study of a new clinical staging system», in Israel, L., Chahinian, A.P. (a cura di), *Lung cancer: natural history, prognosis and therapy*, Academic Press, New York.

Naylor, C.D., Chen, E., Strauss, B. (1992), *Measured enthusiasm: does the method of reporting trial results alter perceptions of therapeutic effectiveness?*, «Annals of internal medicine», 117, pp. 916-921.

NHS Management Executive (1993), *Improving clinical effectiveness*, NHS Management Executive, Leeds.

Nisbett, R.E., Ross, L. (1980), *Human inference: strategies and shortcomings of social judgment*, Prentice Hall, Englewood Cliffs (NJ).

O'Connor P.M., Dowey, K.E., Bell, P.M., Irwin, S.T., Dearden, C.H. (1995), *Unnecessary delays in accident and emergency departments: do medical and surgical senior house officers need to vet admissions?*, «Journal of accident and emergency medicine», 12, pp. 251-254.

Oldridge, N.B., Guyatt, G.H., Fischer, M.E., Rimm, A.A. (1988), *Cardiac rehabilitation after myocardial infarction: combined experience of randomized controlled trials*, «Journal of the American Medical Association», 260, pp. 945-950.

Ong, L.M., de Haes, J.C., Hoos, A.M., Lammes, F.B. (1995), *Doctor-patient communication: a review of the literature*, «Social science and medicine», 40, pp. 903-918.

Oskamp, S. (1962), *The relation of clinical experience and training methods to several criteria of clinical prediction*, «Psychological monographs», 76.

Parker, A., Fischhoff, B. (2004), *Decision-making competence: external validation through an individual-differences approach*, «Journal of behavioral decision making», 17, pp. 1-27.

Patel, V.L., Evans, D.A., Kaufman, D.R, (1989), «Cognitive framework for doctor-patient interaction», in Evans, D.A., Patel, V.L. (a cura di), *Cognitive science in medicine: biomedical modelling*, MIT Press, Cambridge (MA).

Patel, V.L., Groen, G.J. (1986), *Knowledge based solution strategies in medical reasoning*, «Cognitive science», 10, pp. 91-116.

Patel, V.L., HoPingKong, H., Mark, V.C. (1984), *Role of prior knowledge in comprehension of medical information by medical students and physicians*, «Proceedings of the annual conference on research in medical education», 23, pp. 127-132.

Pauker, S.G., Kassirer, J.P. (1980), *The threshold approach to clinical decision making*, «New England journal of medicine», 302, pp. 1109-1117.

Pauker, S.G., Kopelman, R.I. (1992), *How sure is sure enough?*, «New England journal of medicine», 326, pp. 688-691.

Peichl-Hoffman, G. (1991), *Spezifitätsprobleme bei der Testung auf Anti-HIV 1 bzw. Anti-HIV 2 in der Routineuntersuchung von Blutspendern*, «Klinisches Labor», 37, pp. 320-328. Phelps, Shanteau, 1978;

Petty, R.E., Wegener, D.T. (1991), «Thought systems, argument quality, and persuasion», in Wyer, R.S. Jr., Srull, T.K. (a cura di), *Advances in social cognition* (vol. 4), Erlbaum, Hillsdale (NJ), pp. 147-161.

Petty, R.E., Wegener, D.T. (1998) «Attitude change: multiple roles for persuasion variables», in Gilbert, D.T., Fiske, S.T., Lindzey, G. (a cura di), *The handbook of social psychology*, I, McGraw-Hill, New York, pp. 328-390.

Phelps, R.H., Shanteau, J. (1978), *Livestock judges: how much information can an expert use?*, «Organizational behavior and human performance», 21, pp. 209-219.
Physician's Desk Reference (1997), Medical economics, Oradell (NJ).
Pickering, T.G. (1983), *Treatment of mild hypertension and the reduction of cardiovascular mortality: the «of or by» dilemma*, «Journal of the American Medical Association», 249, pp. 399-400.
Pocock, S.J., Hughes, M.D., Lee, R.J. (1987), *Statistical problems in the reporting of clinical trials. A survey of three medical journals*, «New England journal of medicine», 317, pp. 426-432.
Podbregar, M., Voga, G., Krivec, B., Skale, R., Pareznik, R., Gabrscek, L. (2001), *Should we confirm our clinical diagnostic certainty by autopsies?*, «Intensive care medicine», 27, pp. 1750-1755.
Pope, J.H., Aufderheide, T.P., Ruthazer, R., Woolard, R.H., Feldman, J.A., Beshansky, J.R., Griffith, J.L., Selker, H.P. (2000), *Missed diagnoses of acute cardiac ischemia in the emergency department*, «New England journal of medicine», 342, pp. 1163-1170.
Poses, R.M., Cebul, R.D., Collins, M., Fager, S.S. (1985), *The accuracy of experienced physicians' probability estimates for patients with sore throats*, «Journal of the American Medical Association», 254, pp. 925-929.
Poses, R.M., Bekes, C., Copare, F.J., Scott, W.E. (1989), *The answer to «What are my chances, doctor?» depends on whom is asked: prognostic disagreement and inaccuracy for critically ill patients*, «Critical care medicine», 17, pp. 827-833.
Poses, R.M., Bekes, C., Winkler, R.L., Scott, W.E., Copare, F.J. (1990), *Are two (inexperienced) heads better than one (experienced) head? Averaging house officers' prognostic judgments for critically ill patients*, «Archives of internal medicine», 150, pp. 1874-1878.
Poses, R.M., Cebul, R.D., Centor, R.M. (1988), *Evaluating physicians' probabilistic judgments*, «Medical decision making», 8, pp. 233-240.
Poses, R.M., Cebul, R.D., Wigton, R.S., Centor, R.M., Collins, M., Fleischli, G. (1992), *Controlled trial using computerized feedback to improve physicians' diagnostic judgments*, «Academic medicine», 67, pp. 345-347.
Pozen, M.W., D'Agostino, R.B., Selker, H.P., Sytkowski, P.A., Hood, W.B. Jr. (1984), *A predictive instrument to improve coronary-care-unit admission practices in acute ischemic heart disease. A prospective multicenter clinical trial*, «New England journal of medicine», 310, pp. 1273-1278.
Pryse-Phillips, W.E., Dodick, D.W., Edmeads, J.G., Gawel, M.J., Nelson, R.F., Purdy, R.A., Robinson, G., Stirling, D., Worthington, I. (1997), *Guidelines for the diagnosis and management of migraine in clinical practice*, «Canadian Medical Association journal», 156, pp. 1273-87.
Pugh, K., O'Donnell, I., Catalan, J. (1993), *Suicide and HIV disease*, «AIDS care», 5, pp. 391-400.
Raiffa, H. (1968), *Decision analysis: introductory lectures on choices under uncertainty*, Addison-Wesley, Reading (MA).
Redelmeier, D.A. (1991), *Cognitive psychology and medical judgment: some downfalls of studying pitfalls*, «Medical decision making», 11, pp. 169-170.
Redelmeier, D.A. (1995), *Screening for colorectal cancer*, «New England journal of medicine», 333, p. 461.
Redelmeier, D.A. (2005), *The cognitive psychology of missed diagnoses*, «Annals of internal medicine», 142, pp. 115-120.

Redelmeier, D.A., Kahneman, D. (1996), *Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures*, «Pain», 116, pp. 3-8.
Redelmeier, D.A., Shafir, E., *Medical decision making in situations that offer multiple alternatives*, «Journal of the American Medical Association», 273, pp. 302-305.
Redelmeier, D.A., Tan, S.H., Booth, G.L. (1998), *The treatment of unrelated disorders in patients with chronic medical diseases*, «New England journal of medicine», 338, pp. 1516-1520.
Rhoads, J.E. (1969), *Textbook of surgery*, Lippincott, Philadelphia.
Riley, D.M., Fischhoff, B., Small, M., Fischbeck, P. (2001), *Evaluating the effectiveness of risk-reduction strategies for consumer chemical products*, «Risk analysis», 21, pp. 357-369.
Ritov, I., Baron, J. (1990), *Reluctance to vaccinate: omission bias and ambiguity*, «Journal of behavioral decision making», 3, pp. 263-277.
Ritov, I., Baron, J. (1995), *Outcome knowledge, regret and omission bias*, «Organizational behavior and human decision processes», 64, pp. 119-127.
Robberson, M.R., Rogers, R.W. (1988), *Beyond fear appeals: negative and positive persuasive appeals to health and self-esteem*, «Journal of applied social psychology», 18, pp. 277-287.
Rorer, L.G., Goldberg, L.R. (1965), *Acquiescence in the MMPI?*, «Educational and psychological measurement», 25, pp. 801-817.
Rosato, F., Thomas, J., Rosato, E. (1973), *Operative management of nonpalpable lesions detected by mammography*, «Surgery, gynecology and obstetrics», 137, pp. 491-493.
Rothman, A.J., Salovey, P. (1997), *Shaping perceptions to motivate healthy behavior: the role of message framing*, «Psychological bulletin», 121, pp. 3-19.
Rothman, A.J., Salovey, P., Antone, C., Keough, K., Martin, C. (1993), *The influence of message framing on health behavior*, «Journal of experimental social psychology», 29, pp. 408-433.
Rusnak, R.A., Stair, T.O., Hansen, K., Fastow, J.S. (1989), *Litigation against the emergency physician: common features in cases of missed myocardial infarction*, «Annals of emergency medcine», 18, pp. 1029-1034.
Sackett, D.L., Cook, R.J. (1994), *Understanding clinical trials*, «British medical journal», 309, pp. 755-756.
Sackett, D.L., Haynes, R.B., Tugwell, P. (1985), *Clinical epidemiology: a basic science for clinical medicine*, Little Brown, Boston.
Salovey, P., Rothman, A.J., Rodin, J. (1998), «Health behavior», in Gilbert, D.T., Fiske, S.T., Lindzey, G. (a cura di), *The handbook of social psychology*, II, McGraw-Hill, New York, pp. 633-683.
SAS/STAT (1988), *User's Guide Release*, ed. 6.03, SAS Institute Inc., Cary (North Carolina).
Sawyer, J. (1966), *Measurement and prediction, clinical and statistical*, «Psychological bulletin», 66, 178-200.
Schering, A.G. (1992) (a cura di), *AIDS information*, Berlin.
Schreiber, C.A., Kahneman, D. (2000), *Determinants of the remembered utility of aversive sounds*, «Journal of experimental psychology: general», 129, pp. 27-42.
Schwartz, S., Kinosian, B.P., Pieskalla, W.P., Lee, H. (1990), *Strategies for screening blood for human immunodeficiency virus antibody; use of a decision support system*, «Journal of the American Medical Association», 264, pp. 1704-1710.

Sedlmeier, P., Gigerenzer, G. (2001), *Teaching Bayesian reasoning in less than two hours*, «Journal of experimental psychology: general», 130, pp. 380-400.
Seidman, H., Mushinski, M.H., Gelb, S.K., Silverberg, E. (1985), *Probabilities of eventually developing or dying of cancer - United States, 1985*, «CA: a cancer journal for clinicians», 35, pp. 36-56.
Selker, H.P., Griffith, J.L., D'Agostino, R.B. (1991a), *A tool for judging coronary care unit admission appropriateness, valid for both real-time and retrospective use. A time-insensitive predictive instrument (TIPI) for acute cardiac ischemia: a multicenter study*, «Medical care», 29, pp. 610-627.
Selker, H.P., Griffith, J.L., D'Agostino, R.B. (1991b), *A time-insensitive predictive instrument for acute myocardial infarction mortality: a multicenter study*, «Medical care», 29, pp. 1196-1211.
Shafir, E., Simonson, I., Tversky, A. (1993), «Scelta e ragioni», trad. it. in Motterlini, M., Guala, F. (a cura di), *Introduzione all'economia cognitiva e sperimentale*, UBE, Milano, 2005, pp. 167-194.
Shafir, E., Tversky, A. (1992), *Thinking through uncertainty: non-consequential reasoning and choice*, «Cognitive psychology», 24, pp. 449-474.
Shafir, E., Tversky, A. (1995), «Decision making», in Smith, E.E., Osherson, D. (a cura di), *Thinking: an invitation to cognitive science*, MIT Press, Cambridge (MA).
Shaklee, H., Fischhoff, B. (1990), *The psychology of contraceptive surprises: judging the cumulative risk of contraceptive failure*, «Journal of applied psychology», 20, pp. 385-403.
Shapiro, S., Strax, P., Venet, L. (1967), *Periodic breast cancer screening*, «Journal of the American Medical Association», 215, pp. 1777-1785.
Shedler, J., Manis, M. (1986), *Can the availability heuristic explain vividness effects?*, «Journal of personality and social psychology», 51, pp. 26-36.
Sherman, S.J. (1980), *On the self-erasing nature of errors of prediction*, «Journal of personality and social psychology», 2, pp. 211-221.
Shojania K.G., Burton, E.C., McDonald, K.M., Goldman, L. (2004), *Changes in rates of autopsy-detected diagnostic errors over time: a systematic review*, «Journal of the American Medical Association», 289, pp. 2849-2856.
Silverberg, E., Boring, C.C., Squires, T.S. (1990), *Cancer statistics*, «CA: a cancer journal for clinicians», 40, pp. 9-26.
Simon, H.A. (1957), *Models of man: social and rational*, Wiley, New York.
Simon, H.A. (1997), *Scienza economica e comportamento razionale*, trad. it. Comunità, Torino, 2000.
Simonson, I. (1989), *Choice based on reasons: the case of attraction and compromise effects*, «Journal of consumer research», 16, pp. 158-174.
Skin Cancer Foundation (1995), *About skin cancer*, Author, New York.
Slovic, P. (1975), *Choice between equally valued alternatives*, «Journal of experimental psychology: human perception and performance», 104, pp. 280-287.
Slovic, P. (1995), *The construction of preferences*, «American psychologist», 50, pp. 364-371.
Slovic, P., Finucane, M., Peters, E., MacGregor, D.G. (2002), «The affect heuristic», in Gilovich, T., Griffin, D., Kahneman, D. (a cura di), *Heuristics and biases: the psychology of intuitive judgment*, Cambridge University Press, New York, pp. 397-420.
Slovic, P., Fischhoff, B. (1977), *On the psychology of experimental surprises*, «Journal of experimental psychology: human perception and performance», 3, pp. 544-551.

Slovic, P., Fischhoff, B., Lichtenstein, S. (1982), «Facts *versus* fears: understanding perceived risk», in Kahneman, D., Slovic, P., Tversky, A. (1982) (a cura di), *Judgment under uncertainty: heuristics and biases*, Cambridge University Press, New York, pp. 463-489.

Slovic, P., Lichtenstein, S. (1983), *Preference reversals: a broader perspective*, «American economic review», 73, pp. 596-605.

Snedecor, G.W., Cochran, W.G. (1967), *Statistical methods*, Iowa State University, Ames (Iowa).

Snell, J.J.S., Supran, E.M., Tamashiro, H. (1992), *WHO international quality assessment scheme for HIV antibody testing: results from the second distribution of sera*, «Bulletin of the World Health Organization», 70, pp. 605-613.

Snyder, R.E. (1966), *Mammography: contributions and limitations in the management of cancer of the breast*, «Clinical obstetrics, and gynecology», 9, pp. 207-220.

Solomon, C.G., Lee, T.H., Cook, E.F., Weisberg, M.C., Brand, D.A., Rouan, G.W., Goldman, L. (1989), *Comparison of clinical presentation of acute myocardial infarction in patients older than 65 years of age to younger patients: the Multicenter Chest Pain Study experience*, «American journal of cardiology», 63, pp. 772-776.

Sox, H.C., Blatt, M.A., Higgins, M.C., Marton, K.I. (1988), *Medical decision making*, Butterworths, Boston (MA).

Spiegelhalter, D.J. (1985), *Statistical methodology for evaluating gastrointestinal symptoms*, «Clinical gastroenterology», 14, pp. 489-515.

Spielberg, F., Kabeya, C.M., Ryder, R.W., Kifuani, N.K., Harris, J., Bender, T.R., Heyward, W.L., Quinn, T.C. (1989), *Field testing and comparative evaluation of rapid visually read screening essays for antibody to human immunodeficiency virus*, «Lancet», 1, pp. 580-584.

Sporer, S.L., Penrod, S., Read, D., Cutler, B. (1995), *Choosing, confidence, and accuracy; a meta-analysis of the confidence-accuracy relation in eyewitness identification studies*, «Psychological bulletin», 118, pp. 315-327.

Spreen, O. (1981), *The relationship between learning disability, neurological impairment and delinquency: results of a follow-up study*, «Journal of nervous and mental disease», 169, pp. 791-799.

SSPS (1993), *SSPS for windows. Base system user's guide*, ed. 6.0. SSPS, Chicago.

Stamler, J. (1973), *Epidemiology of coronary heart disease*, «Medicine clinics of North America», 57, pp. 5-46.

Stanovich, K.E., West, R.F. (1998), *Individual differences in rational thought*, «Journal of experimental psychology: general», 127, pp. 161-188.

Statistisches Bundesamt (1994) (a cura di), *Statistisches Jahrbuch 1994 für die Bundesrepublik Deutschland*, Statistisches Bundesamt, Wiesbaden.

Stern, R.S., Weinstein, M.C., Baker, S.G. (1986), *Risk reduction for nonmelanoma skin cancer with childhood sunscreen use*, «Archives of dermatology», 122, pp. 537-545.

Stine, G.J. (1996), *Acquired immune deficiency syndrome: biological, medical, social, and legal issues*, Prentice Hall, Englewood Cliffs (NJ).

Stone, A.A., Broderick, J.E., Kaell, A.T., DelesPaul, P.A., Porter, L.E. (2000), *Does the peak-end phenomenon observed in laboratory pain studies apply to real-world pain in rheumatoid arthritics?*, «Journal of pain», 1, pp. 212-217.

Tabar, L., Fagerberg, C.J., Gad, A., Baldetorp, L., Holmberg, L.H., Grontoft, O., Ljungquist, U., Lundstrom, B., Manson, J.C., Eklund, G., *et al.* (1985), *Reduction in mortality from breast cancer after mass screening with mammography*, «Lancet», 1, pp. 829-832.

Tape, T.G., Heckerling, P.S., Ornato, J.P., Wigton, R.S. (1991), *Use of clinical judgment analysis to explain regional variations in physicians' accuracies in diagnosing pneumonia*, «Medical decision making», 11, pp. 189-197.

Thaler, R.H. (1999), *Mental accounting matters*, «Journal of behavioral decision making», 12, pp. 183-206.

Thompson, S.C., Jolley, D., Marks, R. (1993), *Reduction of solar keratoses by regular sunscreen use*, «New England journal of medicine», 329, pp. 1147-1151.

Thornbury, J.R., Fryback, D.G., Edwards, W. (1975), *Likelihood ratios as a measure of the diagnostic usefulness of excretory urogram information*, «Radiology», 114, pp. 561-565.

Tierney, W.M., Fitzgerald, J., McHenry, R., Roth, B.J., Psaty, B., Stump, D.L., Anderson, F.K. (1986), *Physicians' estimates of the probability of myocardial infarction in emergency room patients with chest pain*, «Medical decision making», 6, pp. 12-17.

Tu, X.T., Litvak, E., Pagano, M. (1992), *Issues in human immunodeficiency virus (HIV) screening programs*, «American journal of epidemiology», 136, pp. 244-255.

Tversky, A., Kahneman, D. (1974), *Judgment under uncertainty: heuristics and biases*, «Science», 185, 1124-1131.

Tversky, A., Kahneman, D. (1981), *The framing of decision and the psychology of choice*, «Science», 211, pp. 453-458.

Tversky, A., Kahneman, D. (1983) *Extensional* vs. *intuitive reasoning: the conjunction fallacy in probability judgment*, «Psychogical review», 90, pp. 293-315.

Tversky, A., Kahneman, D. (1986), *Rational choice and framing of decisions*, «Journal of business», 59, pp. 5251-5278.

Tversky, A., Kahneman, D. (1992), *Advances in prospect theory: cumulative representation of uncertainty*, «Journal of risk and uncertainty», 5, pp. 297-323.

Tversky, A., Koehler, D.J. (1994), *Support theory: a non-extensional representation of subjective probability*, «Psychological review», 101, pp. 547-567.

Tversky, A., Shafir, E. (1992a), *The disjunction effect in choice under conflict*, «Psychological science», 3, pp. 305-309.

Tversky, A., Shafir, E. (1992b), *Choice under conflict: the dynamics of deferred decision*, «Psychological science», 3, pp. 358-361.

Tversky, A., Simonson, I. (1993), *Context-dependent preferences: the relative advantage model*, «Management science», 39, pp. 1179-1189.

Ubel, P.A. (1999), *How stable are people's preferences for giving priority to severely ill patients?*, «Social science and medicine», 49, pp. 895-903.

Ubel, P.A., Goold, S. (1997), *Recognizing bedside rationing: clear cases and tough calls*, «Annals of internal medicine», 126, pp. 74-80.

Varey, C., Kahneman, D. (1992), *Experiences extended across time: evaluation of moments and episodes*, «Journal of behavioral decision making», 5, pp. 169-185.

Viscusi, W.K., Magat, W.A., Huber, J. (1987), *An investigation of the rationality of consumer valuations of multiple health risks*, «RAND journal of economics», 18, pp. 465-479.

von Winterfeldt, D., Edwards, W. (1986), *Decision analysis and behavioral research*, Cambridge University Press, New York.

Voytovich, A.E., Rippey, R.M., Suffredini, A. (1985), *Premature conclusions in diagnostic reasoning*, «Journal of medical education», 60, pp. 302-307.
Wachter, R., Shojania, K. (2004), *Internal bleeding. The truth behind America's terrifying epidemic of medical mistakes*, Rugged Land, New York.
Wainer, H. (1978), *On the sensitivity of regression and regressors*, «Psychological bulletin», 85, pp. 267-273.
Wallsten, T.S. (1981), *Physician and medical student bias in evaluating diagnostic information*, «Medical decision making», 1, pp. 145-164.
Ward, J.W. (1994), «Testing for human retrovirus infections: medical indications and ethical considerations», in Schochetman, G., George, J.R. (a cura di), *AIDS testing: a comprehensive guide to technical, medical, social, legal, and management issues*, Springer, New York, pp. 1-14.
Watley, D.J., Vance, F.L. (1974), *US Office of education cooperative research project n. 2022*, University of Minnesota, Minneapolis (MN).
Wedding, D. (1983), *Comparison of statistical and actuarial models for predicting lateralization of brain damage*, «International journal of clinical neuropsychology», 5, pp. 15-20.
Weeks, W.B., Foster, T., Wallace, A.E., Stalhandske, E. (2001), *Tort claims analysis in the Veterans Health Administration for quality improvement*, «Journal of law and medical ethics», 29, pp. 335-345.
Wegener, D.T., Petty, R.E., Klein, D.J. (1994), *Effects of mood on high elaboration attitude change: the mediating role of likelihood judgments*, «European journal of social psychology», 24, pp. 25-43.
Weinstein, M.C., Fineberg, H. (1980), *Clinical decision analysis*, Saunders, Philadelphia.
Weinstein, M.P., Murphy, J.R., Reller, L.B., Lichtenstein, K.A. (1983a), *The clinical significance of positive blood cultures: a comprehensive analysis of 500 episodes of bacteremia and fungemia in adults. II. Clinical observations, with special reference to factors influencing prognosis*, «Review of infectious diseases», 5, pp. 54-70.
Weinstein, M.P., Reller, L.B., Murphy, J.R., Lichtenstein, K.A. (1983b), *The clinical significance of positive blood cultures: a comprehensive analysis of 500 episodes of bacteremia and fungemia in adults. I. Laboratory and epidemiologic observations*, «Review of infectious diseases», 5, pp. 35-53.
Weissler, A.M., Miller, B.I., Boudoulas, H. (1989), *The need for clarification of percent risk reduction data in clinical cardiovascular trial reports*, «Journal of the American college of cardiology», 13, pp. 764-766.
Werner, P.D., Rose, T.L., Yesavage, J.A. (1983), *Reliability, accuracy, and decision-making strategy in a clinical predictions of imminent dangerousness*, «Journal of Consulting and Clinical Psychology», 51, pp. 815-825.
Westdrop, E.J., Gratton, M.C., Watson, W.A. (1992), *Emergency department interpretation of electrocardiograms*, «Annals of emergency medicine», 21, pp. 541-544.
WHO Global AIDS Statistics (1996), «AIDS care», 8, p. 501.
Wichstrom, L. (1994), *Predictors of Norwegian adolescents' sunbathing and use of sunscreen*, «Health psychology», 13, pp. 412-420.
Wigton, R.S. (1988), *Use of linear models to analyze physicians' decisions*, «Medical decision making», 8, pp. 241-252.

Wilber, J.C. (1991), «New development in diagnosing infections», in Volbering, P., Jacobson, M.A. (a cura di), *AIDS clinical review*, Marcel Dekker Inc., New York, pp. 1-15.
Wilks, S.S. (1938), *Weighting systems for linear functions of correlated variables when there is no dependent variable*, «Psychometrika», 3, pp. 23-40.
Wilson, D.K., Purdon, S.E., Wallston, K.A. (1988), *Compliance to health recommendations: a theoretical overview of message framing*, «Health education research», 3, pp. 161-171.
Wilson, R., Harrison, B.T., Gibberd, R.W., Hamilton, J.D. (1999), *An analysis of the causes of adverse events from the Quality in Australian Health Care Study*, «Medical journal of Australia», 170, pp. 411-415.
Winawer, S.J., Fletcher, R.H., Miller, L., Godlee, F., Stolar, M.H., Mulrow, C.D., Woolf, S.H., Glick, S.N., Ganiats, T.G., Bond, J.H., Rosen, L., Zapka, J.G., Olsen, S.J., Giardiello, F.M., Sisk, J.E., Van Antwerp, R., Brown-Davis, C., Marciniak, D.A., Mayer, R.J. (1997), *Colorectal cancer screening: clinical guidelines and rationale*, «Gastroenterology», 112, pp. 594-642.
Windeler, J., Köbberling, J. (1986), *Empirische Untersuchung zur Einschätzung diagnostischer Verfahren am Beispiel des Haemoccult-Tests*, «Klinische Wochenschrift», 64, pp. 1106-1112.
Winkler, R.L. (1986), «On "good probability appraiser"», in Goel, P., Zellner, A. (a cura di), *Bayesian inference and decision techniques*, Elsevier, North Holland.
Wittkowski, K. (1989), *Wann ist ein HIV Test indiziert? Schlusswort*, «Deutsches Ärzteblatt», 86, pp. B138-B140.
Wittman, M.P. (1941), *Mental efficiency levels before and after shock therapy*, «Elgin papers», 4, pp. 70-81.
Wolf, F.M., Gruppen, L.D., Billi, J.E. (1985), *Differential diagnosis and the competing-hypotheses heuristic*, «Journal of the American Medical Association», 253, pp. 2858-2862.
Wolf, F.M., Gruppen, L.D., Billi, J.E. (1988), *Use of the competing-hypotheses heuristic to reduce pseudodiagnosticity*, «Journal of medical education», 63, pp. 548-554.
Wolfe, J.N. (1964), *Mammography: report on its use in women with breasts abnormal and normal on physical examination*, «Radiology», 83, pp. 244-254.
Wolfe, J.N. (1965), *Mammography: errors in diagnosis*, «Radiology», 87, pp. 214-219. Wolfe, 1966
Wolfe, J.N. (1967), *Mammography*, Charles T. Thomas, Springfield (IL).
Woloshin, S., Schwartz, L.M., Byram, S.J., Sox, H.C., Fischhoff, B., Welch, H.G. (2000), *Women's understanding of the mammography screening debate*, «Archives of internal medicine», 160, pp. 1434-1440.
Wood, G. (1978), *The knew-it-all-along effect*, «Journal of experimental psychology: human perception and performance», 4, pp. 345-353.
Wright, G., Ayton, P. (1987), «The psychology of forecasting», in Wright, G., Ayton, P. (a cura di), *Judgmental forecasting*, Wiley, London.
Wu, A.W., Folkman, S., McPhee, S.J., Lo, B. (1991), *Do house officers learn from their mistakes?*, «Journal of the American Medical Association», 265, pp. 2089-2094.
Yates, J.F. (1990), *Judgment and Decision Making*, Prentice-Hall, Englewood Cliffs (NJ).

Yusuf, S., Zucker, D., Peduzzi, P., Fisher, L.D., Takaro, T., Kennedy, J.W., Davis, K., Killip, T., Passamani, E., Norris, R., *et al*. (1994), *Effect of coronary artery bypass surgery on survival: overview of 10-year results from randomised trials by the coronary artery bypass graft surgery trialists collaboration*, «Lancet», 344, pp. 563-570.

Zakay, D. (1983), *The relationship between the probability assessor and the outcome of an event as a determiner of subjective probability*, «Acta psychologica», 53, pp. 271-280.

Zeckhauser, R.J. (1991), *Strategy and choice*, MIT Press, Cambridge (MA).

Gli autori

Michele Anthony lavora come internista presso la Mays Landing Medical Association (Mays Landing, New Jersey). Ha svolto attività di ricerca presso il Department of Internal Medicine del Medical College of Virginia.

Hal R. Arkes è Professor of Psychology presso il Department of Psychology e Interim Director del Center for Health Outcomes, Policy and Evaluation Studies della Ohio State University. Ha diretto il Program for Decision, Risk, and Management Science della US National Science Foundation e presieduto la Society for Judgment and Decision Making.

Robert M. Arnold è Associate Professor of Medicine and Psychiatry e Director of Clinical Training presso la Division of General Internal Medicine del Department of Medicine dell'Università di Pittsburgh.

Prince S. Aujla svolge attività di ricerca presso la Division of Nephrology del Department of Medicine, Università di Toronto.

Brian T. Bedell è Associate Professor of Psychology presso lo Psychology Department del Lewis and Clark College (Portland, Oregon).

John E. Billi è Associate Dean for Clinical Affairs e Professor of Internal Medicine and Medical Education presso la University of Michigan Medical School.

Gretchen B. Chapman è Associate Professor of Cognitive Psychology e direttore del Medical Decision Making Laboratory presso il Department of Psychology della Rutgers University.

Robyn M. Dawes è Professor of Psychology presso il Department of Social and Decision Sciences e membro del Center for Risk Perception and Communication della Carnegie Mellon University di Pittsburgh.

Jerusha Detweiler è Associate Professor of Psychology presso lo Psychology Department del Lewis and Clark College (Portland, Oregon).

Axel Ebert ha svolto attività di ricerca presso il Max Planck Institute for Human Development di Berlino.

David M. Eddy è Senior Advisor for Health Policy and Management della Kaiser Permanente e membro dello Institute of Medicine (US National Academy of Sciences). È stato Professor of Engineering and Medicine presso l'Università di Stanford e Professor of Health Policy and Management e direttore del Center for Health Policy Research and Education presso la Duke University.

Tom Fahey è direttore del Trayside Centre for General Practice presso la University of Dundee School of Medicine.

David Faust è Professor of Psychology presso l'Università del Rhode Island.

Baruch Fischhoff è Professor of Social and Decision Sciences presso il Department of Social and Decision Sciences e Professor of Engineering and Public Policy presso il Department of Engineering and Public Policy della Carnegie Mellon University di Pittsburgh. È membro del Center for Risk Perception and Communication della Carnegie Mellon University e dello Institute of Medicine (US National Academy of Sciences) ed è stato presidente della Society for Judgment and Decision Making.

Lachlan Forrow è Associate Professor of Medicine presso la Harvard Medical School e internista presso la Division of General Medicine and Primary Care del Beth Israel Deaconess Medical Center (Boston), dove dirige il Palliative Care Program e presiede l'Ethics Advisory Committee.

Gerd Gigerenzer è direttore del Center for Adaptive Behavior and Cognition presso il Max Planck Institute for Human Development di Berlino. È stato Professor of Psychology presso l'Università di Chicago ed è Honorary Professor of Psychology presso la Humboldt Universität e la Freie Universität di Berlino.

Lee A. Green è Associate Professor presso il Department of Family Medicine dell'Università del Michigan.

Sian Griffiths è presidente della Faculty of Public Health Medicine del Royal College of Physicians britannico e Professor of Public Health presso l'Università di Oxford e la Oxford Brookes University.

Larry D. Gruppen è direttore e Professor of Medical Education presso il Department of Medical Education della University of Michigan Medical School.

Allan Harkness è direttore e Professor of Psychology presso lo Psychology Department dell'Henry Kandall College of Arts and Sciences dell'Università di Tulsa (Oklahoma).

Ulrich Hoffrage è Professeur de Psychologie presso l'Ecole des Hautes Etudes Commerciales dell'Università di Losanna.

Daniel Kahneman è Professor of Psychology presso il Department of Psychology e Professor of Public Affairs presso la Woodrow Wilson School of Public and International Affairs dell'Università di Princeton. Membro della American Academy of Arts and Sciences e della US National Academy of Science, nel 2002 è stato insignito del Premio Nobel per l'economia.

Joel Katz è Professor of Medicine presso il Department of Anesthesia della Faculty of Medicine dell'Università di Toronto, membro dello University of Toronto Centre for the Study of Pain e Canada Research Chair in Health Psychology presso l'Università di York (Ontario, Canada).

Derek J. Koehler è Associate Professor of Cognitive Psychology presso l'Università di Waterloo (Ontario, Canada).

Varda Liberman dirige la Mathematical and Statistical Studies Unit ed è vicepreside della Arison School of Business presso l'Interdisciplinary Center di Herzliya (Israele).

Barbara J. McNeil è Professor of Health Care Policy e Professor of Radiology presso la Harvard Medical School e il Brigham and Women's Hospital (Boston). Ha diretto il Department of Health Care Policy della Harvard Medical School ed è membro dello Institute of Medicine (US National Academy of Sciences) e della American Academy of Arts and Sciences.

Paul E. Meehl (1920-2003) è stato Professor Emeritus of Psychology presso il Department of Psychology e Member Emeritus presso il Center for Philosophy of Science della University of Minnesota. È stato inoltre presidente della American Psychological Association.

Stephen G. Pauker è Professor of Medicine presso la Tufts University School of Medicine, Vice-Chairman for Clinical Affairs e fondatore della Division of Clinical

Decision Making presso il Department of Medicine del New England Medical Center.

Roy M. Poses è Director of Research presso la Division of General Internal Medicine del Memorial Hospital del Rhode Island e Associate Professor of Medicine presso la Brown University School of Medicine.

Emily Pronin è Associate Professor of Psychology presso lo Psychology Department dell'Università di Princeton.

Donald A. Redelmeier è Professor of Medicine presso l'Università di Toronto, Canada Research Chair in Medical Decision Sciences e direttore della Clinical Epidemiology Unit presso il Sunnybrook & Women's College Health Sciences Centre (Toronto).

Alexander J. Rothman è Associate Professor of Psychology presso il Department of Psychology della University of Minnesota.

Peter Salovey è Professor of Psychology, Professor of Management e Professor of Epidemiology and Public Health presso il Department of Psychology dell'Università di Yale. È inoltre preside dello Yale College e direttore dello Health, Emotion and Behavior Laboratory.

Paul D. Saville svolge attività di ricerca presso la Charleston Division e il Department of Medicine della West Virginia School of Medicine.

Janet A. Schwartz è Research Associate del Department of Psychology e del Center for Health and Wellbeing presso la Woodrow Wilson School of Public and International Affairs dell'Università di Princeton.

Eldar Shafir è Professor of Psychology and Public Affairs presso il Department of Psychology e la Woodrow Wilson School of Public and International Affairs dell'Università di Princeton.

Harold C. Sox è Professor of Medicine e direttore del Department of Medicine del Dartmouth-Hitchcock Medical Center (Lebanon, New Hampshire) e direttore della Foundation for Informed Medical Decision Making. Direttore della rivista *Annals of internal medicine*, è stato presidente dell'American College of Physicians ed è membro dello Institute of Medicine (US National Academy of Sciences) e della American Association for the Advancement of Science.

William C. Taylor è Associate Professor of Medicine presso la Harvard Medical School.

Amos Tversky (1937-1996) è stato Professor of Behavioral Sciences presso il Department of Psychology dell'Università di Stanford e membro della American Academy of Arts and Sciences e della US National Academy of Science.

Fredric M. Wolf è direttore e Professor of Medical Education and Biomedical Informatics presso il Department of Medical Education and Biomedical Informatics della University of Washington School of Medicine.

Robert L. Wortmann è direttore e Professor of Internal Medicine presso il Department of Internal Medicine dell'University of Oklahoma College of Medicine.

J. Frank Yates è Professor of Psychology presso il Department of Psychology dell'Università del Michigan.