

Frau, distorsió, autoengany i consens: els problemes de l'engany humà i de l'engany natural en la ciència epidemiològica

Alvan R. Feinstein

Traducció de Gaietà Permanyer Miralda i Joan M. V. Pons de l'article: Feinstein AR. Fraud, distortion, delusion, and consensus: The problems of human and natural deception in epidemiologic science. *Am J Med.* 1988;84:475-8.

En les darreres dues dècades, la cerca de la veritat científica en el món de la medicina ha sofert problemes seriosos quan s'ha vist que l'evidència era falsificada o quan, sense falsificació, plantejava dubtes importants sobre la seva validesa o interpretació. Si la cerca de la veritat és tant un objectiu professional com moral en la pràctica de la investigació científica, qualsevol tipus d'engany que l'emascarar és particularment repugnant. Però el grau de repugnància que produeixin els diferents tipus d'engany dependrà de la manera com es duguin a terme.

Alguns dels enganys són frau planificats de manera deliberada, però molts altres tenen lloc de manera inadvertida. Apareixen com a distorsions o autoenganys, produïts per defectes en els mètodes convencionals emprats en la pràctica "estàndard" de la recerca científica. Així com tothom se sent molt trasbalsat pels aspectes immorals del frau, als altres tipus d'engany no s'hi associa un rebuig ètic. Habitualment són acceptats com una part dels riscos inevitables associats als beneficis de la recerca científica. Vist així, el procés de la recerca és d'alguna manera semblant a un tractament eficaç que ocasionalment pot tenir efectes col·laterals adversos. Ens complauen les bones coses que esdevenen amb la recerca i esperem que les inevitables coses dolentes seran infreqüents o, almenys, no massa dolentes.

D'acord amb això, s'emprenen vigoroses mesures correctores quan es descobreix que els enganys han estat deliberats, però no quan han estat accidentals. Si un investigador es refereix a ratolins ficticis, falsifica dades o crea malalts inexistents, la "immoralitat" científica resulta extremadament escandalosa i es prenen mesures correctores ràpidament. Algunes persones perden la feina o la reputació; els comitès existents són reprovats i se'n creen de nous; i es desenvolupen millors procediments de supervisió, educació i monitoratge de les activitats de recerca. No obstant això, quan el furor va de baixa, la "indústria" de la recerca acostuma a retornar al seu equilibri previ. Ens sentim alleujats amb les millores dutes a terme per evitar enganys deliberats, però no ens preocupem gaire, o gens,

pels problemes molt més grans i més freqüents dels enganys inadvertits que s'esdevenen quan els estàndards de la pràctica "establerts" porten a distorsions o autoenganys.

En els enganys inadvertits no s'han falsificat dades ni observacions. Els investigadors han estat honorats, responsables i sincers, han treballat dur i les dades poden haver estat recollides i analitzades amb cura meticulosa; però ha anat malament alguna cosa. En la *distorsió*, quelcom ha afectat l'evidència que es descriu com a "nova": les dades poden ser inexactes o les comparacions esbiaixades. En l'*autoengany*, quelcom ha afectat el "procés interpretatiu" que té lloc quan les dades observades condueixen a formular hipòtesis o conclusions. La distorsió es produeix habitualment perquè no es reconeixen distincions importants en la complexitat de la natura. Habitualment, l'autoengany ve d'un zel excessiu en les expectatives, creences o conducta dels investigadors.

Per exemple, es produeix una distorsió quan es presenten dades convencionals per demostrar que els malalts sotmesos a tractament quirúrgic del càncer tenen millors taxes de supervivència que els no operats. Aquesta distorsió té lloc perquè la comparació és esbiaixada. Els criteris d'"operabilitat" ja donen un millor pronòstic pretractament al grup quirúrgic, encara que la cirurgia no es dugui a terme, que al dels malalts "inoperables", que són rebutjats per a l'operació. També es pot produir distorsió si la millora subjectiva després d'un tractament és avaluada per malalts i metges que coneixen aquest tractament. Segons com es percebin els esdeveniments que tenen lloc després d'un tractament que se suposa actiu, de manera diferent als que no segueixen cap tractament o segueixen un placebo manifest, els fenòmens resultants rebran una avaluació esbiaixada.

L'autoengany pot tenir lloc en la interpretació de dades aparentment acurades i lliures de biaix que mostren que taxes més altes de desenvolupament de càncer es correlacionen amb increments de la industrialització. Si la conclusió és que la industrialització ha produït el càncer, aquesta interpretació pot ser un autoengany. La taxa de desenvolupament del càncer pot haver augmentat no perquè el propi càncer augmenti, sinó perquè càncers fins aleshores desapercebut es detecten més sovint amb els

millors mètodes diagnòstics que s'han desenvolupat de manera concomitant amb la industrialització.

Les distorsions tenen lloc de manera regular quan les relacions causa-efecte s'interpreten a partir de l'observació d'esdeveniments humans naturals i no a partir d'actuacions plantejades experimentalment. Els grups que es formen a efectes de comparació en circumstàncies observacionals sovint estan esbiaixats pels molts fenòmens naturals que afecten l'elecció dels elements considerats "causes"; sovint les dades estaran també esbiaixades en la detecció dels resultats subsegüents que es consideren "efectes". Els judicis individuals i altres decisions selectives que es duen a terme en circumstàncies naturals (és a dir, no experimentals) poden produir diferents tipus ben definits de distorsions o biaixos "innats".

Independentment del nom que es doni a aquests biaixos, el reconeixement de les distorsions potencials és la raó principal per la qual actualment s'exigeix que les relacions causa-efecte siguin demostrades, sempre que sigui possible, amb evidència d'alta qualitat, no obtinguda a partir d'estudis observacionals sinó amb experiments realitzats en forma d'assajos clínics aleatoritzats amb cegament doble.

Apareixen també autoenganyos de manera regular quan dades que poden ser acurades i no esbiaixades s'interpreten amb el prejudici de creences preconcebudes. Per evitar cites ofensives d'esdeveniments recents, aquest problema es pot il·lustrar amb dos exemples epidemiològics clàssics. El 1852, William Farr (el prestigiós fundador de les modernes "estadístiques vitals") va aplegar un grup correcte d'observacions i dades que mostraven que el còlera era més freqüent a nivell del mar, a pressió atmosfèrica elevada, que en zones de muntanya, on la pressió atmosfèrica és més baixa. Però Farr interpretà les dades segons la teoria dels miasmes, que aleshores era popular per explicar la patogènia de les malalties. Concloué que el còlera era produït per miasmes que s'observen a pressions atmosfèriques elevades i rebutjà una hipòtesi alternativa nova (proposada per John Snow) segons la qual el còlera era transmès per l'aigua, més pura en regions de muntanya que a nivell del mar.

Un altre autoengany clàssic, en la primera meitat del segle xx, fa referència a la patogènia de la pel·lagra. En diferents revisions epidemiològiques a gran escala s'obtingueren acurats reculls de dades que mostraven que la pel·lagra era especialment freqüent en famílies i veïns. Utilitzant els conceptes clàssics de patogènia, els investigadors epidemiològics varen interpretar els resultats suposant que indicaven que la malaltia era infecciosa i probablement també familiar. Finalment, Joseph Goldberger, a la primeria dels anys 1920, va dur a terme l'heroica recerca que demostrà l'error d'aquestes dues hipòtesis i que revelà el paper etiològic de la carència dietètica.

Les febleses humanes que produeixen distorsions i autoenganyos són difícils d'evitar. Provenen de tres reaccions humanes naturals: la pressió per creure allò que volem creure, el refús d'acceptar allò que és contrari a les nostres creences i una profunda repugnància per qualsevol cosa que suggereixi que ens hem equivocat o que hem fet mal. Els metges actuals s'horroritzen quan coneixen, a partir de la història de la medicina, els errors que els nostres avantpassats professionals varen cometre en els seus raonaments etiològics i terapèutics. Els seus conceptes humerals sobre la patogènia de les malalties eren un imaginari prestigiós que no s'acompanyava d'evidència documental, i les seves sagnies, ventoses, purgues i emètics feien, amb seguretat, més mal terapèutic que no pas bé. Però molts errors anàlegs han tingut lloc també en el segle XX i molts d'altres, sens dubte, encara existeixen avui, malgrat els nostres avenços científics en assajos aleatoritzats i els sofisticats mètodes estadístics.

Els assajos clínics han estat formidables per demostrar l'acció curativa dels agents terapèutics, però sovint han creat més calor que no pas llum en investigar els efectes profilàctics a llarg termini en la prevenció primària de les malalties o en la prevenció secundària de les complicacions de la malaltia establerta. Els mètodes estadístics han estat un mitjà esplèndid per avaluar l'evidència quantitativa, però les anàlisis elaborades a vegades han enterbolit més que no pas aclarit els problemes científics, particularment pel que fa als estudis observacionals. Per exemple, si els càlculs estadístics de William Farr es repetissin avui, probablement no s'expressarien com a correlacions simples sinó en forma d'anàlisis multivariades fent ús de la regressió logística o de les funcions discriminants. Qualsevol persona que dubtés intuïtivament de la conclusió de Farr sobre els "miasmes" atmosfèrics que produïen el còlera potser quedaria tan confusa o atemorida per la complicada anàlisi matemàtica que ni tan sols gosaria parlar.

Els problemes moderns es veuen avui dia magnificats per alguns dels dobles estàndards que actualment s'apliquen a l'avaluació científica de les relacions causa-efecte. Encara que cal suposar que el raonament sobre aquestes relacions es basa en els mateixos principis, tant si l'agent causal té efectes etiològics com terapèutics, habitualment l'etiologia i la terapèutica s'avaluen amb estàndards científics diferents. Quan els clínics estudien els beneficis de la terapèutica, els seus resultats seran acceptats tan sols si provenen d'un model experimental precís i de mètodes científics rigorosos per prevenir les distorsions. Però quan els epidemiòlegs estudien agents etiològics sospitats o efectes adversos de teràpies no es plantegen exigències científiques equivalents. Les dades procedeixen d'estudis observacionals i no d'experiments. El "model" de mostreig estadístic emprat en la recerca observacional no conté obligatòriament estratègies científiques específiques

per identificar i eliminar els biaixos inherents. Com a mínim dues altres activitats que serien rebutjades en la ciència experimental han estat acceptades sense obstacles en la recerca observacional: els investigadors poden fer canvis arbitraris i retroactius en els grups control, i hipòtesis que han estat generades pel “dragatge computat” de dades es poden proclamar i difondre sense l'ajust estadístic o científic oportú i sense que s'exigeixi que les hipòtesis generades es verifiquin amb dades noves.

En absència d'estàndards científics rigorosos, moltes distorsions i autoenganys (així com les polèmiques que se'n deriven) seran inevitables en qualsevol tipus de recerca. No obstant això, amb el progrés científic ordinari, els problemes finalment s'identificaran quan apareguin resultats contradictoris i podran desenvolupar-se solucions quan es proposin mètodes nous o millors per corregir els defectes.

Però aquest tipus de progrés pot quedar obstaculitzat per un altre concepte nou en la ciència mèdica moderna: la “síndrome del consens”. En la forma més lleu d'aquesta síndrome, les disputes no es resolen mitjançant el procés científic d'identificació i reparació dels defectes metodològics, sinó amb una mena de procés polític: el vot de la majoria en un grup d'experts seleccionats. En la forma més greu de la síndrome, el procés passa de la política a la religió. El consens es converteix en un dogma instituit que exigeix una conformitat fidel i la crítica dels dissidents es considera una activitat herètica. Per tal com sovint els heretges són silenciats amb el refús de beques sol·licitades —l'equivalent acadèmic de ser cremats a la foguera—, el manteniment del consens pot ser recomanat vigorosament i, fins i tot, forçat activament.

Encara que habitualment es mostra a porta tancada durant l'avaluació per consemblants de les propostes de recerca, actualment la síndrome del consens s'ha posat més a la vista. Des de fa pocs anys, s'han discutit temes que han plantejat incertesa o polèmica en una sèrie de conferències de consens als Estats Units i, més recentment, al Regne Unit. Les conferències arriben al “consens” en forma d'acord o vot de la majoria d'un selecte grup d'experts reunits. No es fa cap intent d'abordar, identificar o millorar les febleses científiques que han donat lloc a la incertesa i a la controvèrsia.

Aquest tipus de mecanisme de consens pot ser satisfactori per oferir guies “prudents” sobre l'estat actual d'un coneixement, però l'acord entre experts ha estat una font tradicional de tots els errors que s'han establert al llarg de la història de la medicina. Així com el bon govern es basa en una llei formal i no en opinions individuals, la bona ciència depèn de mètodes i evidència adequats, no d'un consens entre autoritats.

La síndrome del consens és especialment perjudicial per al progrés científic en epidemiologia, perquè poques vegades els investigadors poden dur a terme assajos expe-

rimentals i han de refiar-se d'estudis observacionals. Per millorar la qualitat científica dels mètodes no experimentals cal un nou tipus de conferència orientat cap a un focus situat més en la construcció que el consens. Aquestes conferències podrien començar identificant els conflictes, determinant les seves bases metodològiques, identificant els defectes en els seus mètodes i desenvolupant millores adequades. Poques vegades es pot assolir aquest tipus de raonament perceptiu en grups grans de persones, motiu pel qual pot caldre que les conferències siguin petits tallers i no reunions massives en les quals representants adequats dels punts de vista oposats treballin amb moderadors apropiats per guiar el debat cap a solucions constructives.

Entre els molts temes epidemiològics que es podrien debatre en aquestes noves conferències es poden citar tres qüestions destacades com a matèries que, des de fa temps, esperen un escrutini científic curós:

1. Fa més de 40 anys, Joseph Berkson va suggerir que els criteris habituals per indicar l'hospitalització podien invalidar la forma habitual en què la informació dels hospitals és utilitzada en els estudis epidemiològics de casos i controls. La proposta de Berkson i l'evidència empírica en què es basa, que no va ser aplegada fins més de 30 anys més tard, tenen implicacions molt importants que mai no han estat subjectes a un debat obert i il·lustrat.
2. Les tabulacions nacionals i internacionals relatives a les “taxes d'incidència de malaltia” depenen de canviar sovint les guies jeràrquiques per escollir un únic diagnòstic dels molts que es registren (o s'ometen) arbitràriament en els certificats de defunció. Aquesta informació estadística sobre la “causa de la mort” és tan poc digna de confiança que molt pocs científics clínics seriosos hi presten atenció. Entre els defectes més ostensibles trobem que en diferents regions o àrees els diagnòstics es veuen dràsticament afectats per les diferències tecnològiques, que els diagnòstics concrets sovint són erronis i que les taxes d'incidència, que sovint només es compten a partir d'una sola de les entitats esmentades en el certificat de defunció, són escandalosament baixes. Segurament, els futurs historiadors de la medicina es meravellaran de la complaença que va permetre que un procediment tan poc científic hagi estat tolerat tant temps i, pitjor encara, que els seus resultats erronis hagin estat acceptats i utilitzats per a decisions importants en la política sanitària nacional. I en canvi, no s'ha fet gairebé res per intentar millorar la situació, ni tan sols en regions petites on es podrien obtenir dades fiables si es fessin els esforços corresponents.
3. Fa uns 14 anys es va proposar que la reserpina era una possible causa de càncer de mama quan en tres

treballs diferents, publicats al mateix número de la mateixa prestigiosa revista, el fàrmac va ser implicat de manera concordant. Actualment, s'accepta de manera general que aquesta implicació és espúria, però no s'han dut a terme conferències per determinar què és el que va fallar, perquè tots tres grups de recerca independents van arribar a la mateixa conclusió errònia i com evitar problemes similars en el futur.

Un altre tema recent que també mereixeria ser debatut és la publicació, en el mateix número d'una altra revista prominent, de dos estudis epidemiològics a gran escala amb conclusions contradictòries. En un dels treballs es considerava que el tractament postmenopàusic amb estrògens causava malaltia coronària, mentre que l'altre els trobava protectors. En el comentari editorial que acompanyava aquests dos estudis contradictoris no hi figurava cap suggeriment sobre què es podia fer per resoldre el conflicte.

En la majoria de camps de la ciència, identificar, reconèixer obertament i examinar de manera reflexiva les contradiccions ha estat un estímul fonamental del progrés. Donat que les dades observacionals de l'epidemiologia són de primera importància per al futur de l'atenció mèdica, i també per a la planificació econòmica de la indústria i la salut, actualment la prevenció de les distorsions i els autoenganys representa un repte fonamental. Les conferències de consens poden ser necessàries per satisfer incentius polítics o per resumir l'estat del coneixement, però el progrés científic ve més del "diví descontentament" que de la conformitat forçada. L'objectiu de noves conferències hauria de ser la millora dels mètodes i estàndards amb què es pot assolir el consens, no mitjançant l'artefacte d'un acord entre autoritats seleccionades, sinó amb metodologia científica, evidència i interpretacions inequívocament vàlides.