

## Dos segles de conflicte-col·laboració entre la medicina i les matemàtiques

Alvan R. Feinstein

Traducció de Joan. M. V. Pons Ràfols i Gaietà Permanyer Miralda de l'article: Feinstein AR. Two centuries of conflict-collaboration between medicine and mathematics. *Journal of Clinical Epidemiology*. 1996;49:1339-1343.

En aquest exemplar del *Journal of Clinical Epidemiology*, Alfred Morabia descriu alguna de les primeres aventures de Pierre Louis en el camp de l'estadística mèdica i Jan Vandenbroucke discuteix algunes de les velles i presents controvèrsies<sup>1</sup>. Aquests temes són també el subjecte d'aquest assaig, el qual representa quasi una revisió i comentari addicional d'un nou llibre esplèndid, *Quantification and the quest for medical certainty*, de J. Rosser Matthews<sup>2</sup>.

Les controvèrsies estadístiques s'han fet tan habituals en la pràctica i en la recerca mèdica que potser hem oblidat quant de temps ha passat i potser no reconeixem que les disputes d'avui en dia sovint repeteixen les batalles del passat o s'hi assemblen. En el seu nou llibre, Matthews discuteix conflictes "matemèdics" que encara existeixen actualment i que s'han anat produint durant aproximadament uns 200 anys, des que el prominent matemàtic francès, Pierre-Simon Laplace, afirmés, el 1795, que "el principal mitjà per determinar la veritat [...] està basat en probabilitats".

Començant amb els orígens i desenvolupaments francesos d'aquesta creença, augmentada subsegüentment amb contribucions d'Alemanya i el Regne Unit, Matthews mostra que molts dels "temes mèdics actualment a l'avantguarda del debat públic [...] són remarcablement semblants als abordats inicialment (en un conjunt de disputes públiques) a la parisenca Acadèmia de Medicina el 1837". El llibre descriu la manera en què els metges, després de lluitar vigorosament contra la intrusió dels matemàtics en la medicina, finalment es convertiren en col·laboradors i proponents, a través dels assaigs clínics aleatoritzats. Matthews conclou el llibre citant l'antic aforisme francès, *Plus ça change, plus c'est la même chose* (com més coses canvien, més es mantenen iguals).

En els paràgrafs que segueixen comentaré cinc dels temes de Matthews per als quals els mateixos vells principis es discuteixen encara actualment, tot i que amb diferents detalls moderns. Són: 1) el contingut de la ciència mèdica; 2) el focus clínic en individus o grups; 3) el concepte de

mesura; 4) l'avaluació de l'estabilitat dels resultats numèrics, i 5) la qualitat de les dades bàsiques.

### 1. El contingut de la ciència mèdica

Malgrat que pocs metges actualment veurien les matemàtiques com un ingredient fonamental de la ciència mèdica, la idea de quantificar —en l'era anterior a les proves de laboratori i als experiments— va ser introduïda a principis de 1800 amb l'objectiu de donar a la medicina un fonament "científic". En lloc de confiar en anècdotes, judicis no especificats i expressions vagues de freqüència (com "molts" i "sovint"), es va instar els metges a utilitzar nombres per als resultats clínics.

El "model" científic d'aleshores era l'astronomia, la qual havia donat lloc a dos menes d'èxits matemàtics en el segle precedent. Un d'ells era la construcció d'equacions per identificar i predir l'òrbita dels planetes i altres entitats celestes. L'altre èxit era la solució aparent al problema de la "variabilitat de l'observador". Quan diferents observadors, utilitzant la llavors nova tecnologia astronòmica, produïen diferents mesures per a la mateixa entitat, s'havia de decidir quin era el valor *correcte*. Després de diverses formes de raonament, inclòs el desenvolupament del "mètode dels mínims quadrats", la mitjana aritmètica de les observacions va ser designada com a "veritat" i les desviacions de la mitjana van ser anomenades "errors".

El segle XVIII també havia portat altres tres assoliments matemàtics. Els "ajuts" rebuts com a suport de la reialesa interessada en les apostes havien inspirat mètodes per a formular probabilitats; s'havien desenvolupat mètodes com el càlcul per a diferenciar i integrar equacions matemàtiques; l'enumeració de naixements i defuncions, preparades com a part de la "política aritmètica", s'havia convertit en proporcions i s'expressaven com a taxes de mortalitat. Les taxes foren subsegüentment ordenades en taules de vida (supervivència), les quals foren aleshores utilitzades pragmàticament per formar expectacions actuarials per les companyies que venien assegurances de vida.

A començament del segle XIX, els metges eren apresats a convertir-se en científics utilitzant el "càlcul de probabilitats", que rarament es definia específicament però que presumiblement incloïa totes les tècniques matemà-

tiques referides anteriorment. La resposta clínica inicial a aquesta proposta fou d'indignació i resistència. La major part dels metges creia que l'art mèdic no necessitava ni volia convertir-se en ciència i que el procés, sens dubte, no havia d'utilitzar mètodes matemàtics.

Avui en dia, òbviament, en la mesura en què la medicina clínica s'ha impregnat d'estadística (o ha estat dominada per ella), la principal disputa sobre ciència és l'elecció de l'orientació adequada. Amb el desenvolupament, en el segle XIX, de la necròpsia, l'examen microscòpic, les proves de laboratori i els experiments de laboratori, l'objectiu científic principal va passar a ser l'explicació del mecanisme de la malaltia. Aquest objectiu havia estat incitat pels avenços en la radiologia a principis del segle XX i en altres mètodes utilitzats per establir etiquetes diagnòstiques i explicacions fisiopatològiques. En les darreres dècades, les explicacions espectaculars dels avenços en la biologia molecular han portat a la seva designació com a "ciència biomèdica bàsica", a partir de la qual presumiblement s'aplica tot el treball clínic. Tot i l'admiració pels èxits i la utilització entusiàstica de la nova tecnologia, els clínics han argumentat que una "ciència bàsica" diferent, o si més no complementària, és necessària en la pràctica clínica, ja que els magnífics èxits moleculars en l'explicació de la malaltia no ofereixen mètodes o material apropiat per escollir i avaluar les intervencions aplicades en l'atenció als pacients.

Un problema addicional, recentment emfatitzat per dos metges que han rebut el premi Nobel pel seu treball en els mecanismes moleculars de la hipercolesterolèmia, és que un focus excessiu o exclusiu en la ciència de la biologia molecular pot donar lloc a clínics que no comprenen la fisiopatològica dels òrgans i sistemes, i conduir a una recerca que està massa distant en la seva relació amb els reptes en salut i la malaltia.

## 2. Individus *versus* grups

Un altre fort argument inicial, ofert contra la intrusió de les matemàtiques en la medicina, era que el "mètode numèric", el qual havia estat batejat i advocat per Pierre Louis, requeria que les dades fossin agregades en grups i que les decisions fossin preses segons els resultats en els grups. Aquesta aproximació era un anatema per als clínics que creien que cada pacient necessitava una atenció i anàlisi individuals.

Aquesta creença estava probablement també suportada pel vell concepte humoral de la malaltia, que no havia estat completament reemplaçat per noves idees durant el segle XIX sobre la patologia dels òrgans i per la taxonomia nosològica derivada dels avenços contemporanis en la necròpsia. D'acord amb aquell vell concepte humoral, cada pacient tenia una malaltia única basada en els desequilibris individuals dels quatre humors. Els clínics havien de discernir aquests desequilibris fisiològics mentre examinaven cada pacient i havien de corregir-los mitjançant

l'aplicació d'una teràpia "racional" com la sagnia, les venoses, les purgues i els emètics.

L'atac a les dades grupals, més que a les anàlisis individuals, va esdevenir particularment vehement després que Louis, el 1828 i el 1835, referís els resultats del seu famós estudi amb el mètode numèric sobre l'eficàcia de les sagnies, les quals havien estat establertes com una modalitat de tractament durant segles. El treball de Louis ha estat considerat erròniament com un "assaig clínic", que comparava la sagnia amb un grup control. De fet, les seves dades provenien de la revisió de notes de casos mèdics en un examen prolectiu (sovint dit prospectiu) d'una cohort de pacients hospitalitzats amb el diagnòstic de pneumònia, erisipela facial i amigdalitis. No hi havia grup control sense tractament, donat que tots els pacients havien rebut sagnies. En lloc d'això, Louis comparava les taxes de mortalitat i la durada mitjana dels símptomes en els pacients que rebien una sagnia primerenca (abans de quatre dies) o tardana. Quan no va trobar pràcticament diferència entre els dos grups, va concloure que "la utilitat de la sagnia ha estat molt limitada en els casos fins ara analitzats".

Louis, amb precaució, no va instar a l'abandonament de la sagnia i, fins i tot, digué d'ella "que no havia de ser descuidada en inflamacions que eren greus o assentaven en un òrgan important". També va remarcar "que s'havia de preferir l'ús de la llanceta al de les sangoneres". Malgrat aquestes conclusions conservadores, el seu treball va ser percebut com un atac a les sagnies; havia ofert també una proposició més "perillosa": instava a l'aplicació del mètode numèric en l'avaluació de totes les altres formes de teràpia mitjançant la utilització de "l'anàlisi exacta de fets numèrics, ben constatats, classificats amb atenció i comptats".

En el prefaci a la traducció anglesa del 1835, James Jackson, que era aleshores el cap de medicina a l'Hospital General de Massachusetts, escriví "penso que no és pas agosarat predir que en cinquanta anys l'art de curar se sustentará en moltes regles exactes [...] les quals no aniran endavant com a derivacions de grans principis fisiològics o patològics, sinó deduïdes a partir de l'agregació d'observacions acurades i fidels de fets individuals".

Pel fet que Louis havia introduït també la idea d'utilitzar mitjanes, així com proporcions (o taxes) per expressar els resultats postterapèutics, la idea de mitjana més que no pas la del malalt individual va passar a ser objecte del contraatac. Matthews descriu de manera excel·lent l'angoixa dels clínics per les amenaces a la santedat de la sagnia i a les afirmacions terapèutiques sense proves i per la perspectiva que allò característic de cada pacient pogués ser reduït a la imprecisió biològica d'una mitjana.

Louis va respondre que havia analitzat en el seu estudi les distincions individuals en molts atributs com l'edat i la gravetat de la malaltia i que, a més, qualsevol diferència entre els grups tractats s'equilibraria quan es comparés un

gran nombre de casos. Les subsegüents estadístiques terapèutiques, particularment en la primera meitat del segle XX, mostraren, tanmateix, que Louis s'equivocava en l'esperança que els grans nombres per ells mateixos crearien un equilibri apropiat. Les eleccions de tractament basades en el judici sovint estaven associades amb diferències pronòstiques en els pacients. Conseqüentment, quan una forma de tractament seleccionada segons judici es comparava amb una altra, els grans nombres simplement amplificaven el biaix de susceptibilitat més que no pas l'eliminaven. No ha estat fins el passat mig segle que el biaix s'ha evitat mitjançant l'estratègia estadística d'assignar aleatòriament el tractament als grups que es comparen.

Malgrat els èxits i assoliments dels assajos aleatoritzats avui en dia, aquests han plantejat qüestions una altra vegada sobre si les decisions clíniques s'han de prendre sobre individus o sobre grups. Els assajos aleatoritzats van ser dissenyats i pensats per demostrar l'eficàcia terapèutica mitjançant la demostració que el tractament A és, de mitjana, millor que el tractament B o el placebo. Els resultats són esplèndids per a les decisions i per a la política de les companyies farmacèutiques i les agències reguladores. Per als clínics i els pacients, tanmateix, el resultat mitjà de l'assaig aleatoritzat, o la metaanàlisi que construeix mitjanes de mitjanes, pot no ser adequat per a les importants distincions entre subgrups clínics pertinents que s'han de considerar quan els tractaments s'escullen per a pacients individuals.

La disputa moderna, que va més enllà de l'abast d'aquest assaig, no es focalitza en la singularitat d'un individu *versus* un col·lectiu amorf. En canvi, els clínics estan incòmodes amb el fet que les importants distincions en subgrups clínics apropiats estan absents de les dades utilitzades en la “medicina basada en l'evidència”, que deriva dels assajos aleatoritzats i les metaanàlisis.

### 3. Conceptes de mesura

Termes com “mètode numèric” i “quantificació” impliquen la idea de mesura que s'expressa com a mesurament i enumeració. Mesurament és l'acte d'obtenir dades en brut quan una entitat observada s'assigna a una categoria pertinent escollida d'entre una escala disponible de categories. El mesurament es dona quan diem que algú té 42 anys (en una escala dimensional d'anys), està *embarassada* (en una escala binària embarassada/no embarassada), *moderadament malalt* (en una escala ordinal de gravetat) i que té un *adenocarcinoma* (en una escala nominal histològica). L'enumeració es dona quan comptem i obtenim totals numèrics per a persones individuals o per a categories dins un grup.

La matemàtica i l'estadística només es poden produir després que hagi tingut lloc el mesurament i l'enumeració. Per a categories sense dimensió, el recompte enumerat es pot expressar en forma de proporcions (o taxes) que po-

den ser subjectes de comparació i del “càlcul de probabilitats”. Per a dades dimensionals, l'addició de valors totals produeix una *quantitat*, no una *enumeració*. La suma se sol convertir en una mitjana, la qual pot ser la mediana dels valors ordenats o la mitjana aritmètica obtinguda quan la quantitat total es divideix pels membres enumerats del grup.

Poques vegades s'ha distingit clarament la diferència entre aquests dos tipus de mesura (no comentats per Matthews) en les disputes antigues o modernes sobre matemàtiques i medicina. La distinció clau en la mesura és la taxonomia: la identificació i construcció de categories apropiades que no necessiten ser expressades en dimensions. Els contraris a Louis potser es queixaven de l'ús de valors mitjans per a una persona individual, però la principal preocupació, reiterada avui en dia per metges que es preocupen pels defectes de la “medicina basada en l'evidència”, és si les dades eren (i són) taxonòmicament inadequades. Es descuiden massa distincions importants i categòricament heterogènies quan els resultats estadístics es comprimeixen per a una fal·laç agregació “homogènia” de pacients amb diverses diferències clíniques, els quals tenen la mateixa “malaltia”, o que compleixen el mateix “criteri d'admissió” per a un assaig aleatoritzat.

Les distincions entre formes de mesura dimensionals *versus* no dimensionals es passen també per alt en els debats actuals sobre la recerca qualitativa i quantitativa. Els proponents de la recerca “qualitativa” argumenten, correctament, que molts atributs humans crucials —dolor, angoixa, gravetat de la malaltia, temor, amor— no es poden esmentar com a dades dimensionals. Però, aleshores pot passar que els proponents no estableixin sistemes adequats de taxonomia i escales ordinals de mesura que permetrien la mesura d'aquests atributs per a la seva enumeració i anàlisi subsegüent. Els problemes s'han perpetuat per un enfocament estadístic a llarg termini restringit a dades dimensionals com a principal o única forma de mesura. La utilització d'escales ordinals per a categories ordenades i el desenvolupament de mètodes “no paramètrics” per analitzar dades ordenades no es va popularitzar fins després de la II Guerra Mundial; molts llibres de text d'estadística ensenyen encara avui en dia als clínics o bé a ometre la utilització d'escales categòriques ordinals o bé les ofereixen com una idea *a posteriori* cap al final del text. Els llibres també donen poca o nul·la atenció als avantatges i desavantatges dels mètodes psicomètrics moderns utilitzats per construir escales de categorització per a fenòmens complexos com la capacitat funcional, l'estat de salut, la satisfacció amb l'atenció, les preferències personals i la qualitat de vida.

No obstant, quan les dades dimensionals (edat, alçada, pes, circumferència toràtica, llargada dels dits, temperatura i, finalment, el torrent de mesures de laboratori) foren disponibles en el segle XIX, la seva distribució estadística

va ser estudiada per Adolphe Quetelet, el qual es va convèncer que la mitjana no era sols una mitjana o el valor correcte en una sèrie de mesuraments, sinó el valor *ideal* del grup. Proposà uns rangs de normalitat construïts estadísticament (encara utilitzats avui en dia) en els quals valors propers a la mitjana eren bons i desviacions extremes de la mitjana eren dolentes.

La persona mitjana ideal de Quetelet, tanmateix, va ser també rebutjada pels clínics. D'acord amb François Double, aquesta aproximació reduiria el metge a ser "un sabater que després de mesurar els peus de mil persones persistia en l'encaix de tothom a partir d'un model imaginari". Per als clínics d'aquella època, com per a molts avui en dia, els agregats estadístics i les mitjanes eren inacceptables perquè els pacients i les malalties humanes no podien manejar-se amb una estratègia de "talla única".

Les idees de Quetelet sobre les desviacions de la mitjana, no obstant, foren reflectides en les corbes "normal" o "gaussiana" utilitzades per representar patrons de distribució de dades biològiques i per il·lustrar molts conceptes teòrics ensenyats en l'estadística elemental. Les corbes encara s'utilitzen de manera preferencial avui, malgrat l'abundant evidència contrària que la major part de dades en les proves modernes de laboratori no tenen una distribució gaussiana. La idea original de Quetelet sobre la mitjana ha tingut també una evolució irònica. La zona de valors mitjans "ideal" al voltant de la mitjana es descriu actualment sovint amb el desdenyós terme de "mediocre".

#### 4. L'estabilitat en els resultats numèrics

Tot i que les idees sobre probabilitat matemàtica encara no existien, els clínics implicats en la quantificació després de Louis ja pogueren desenvolupar altres aproximacions per avaluar l'estabilitat dels resultats numèrics en grups petits. Per exemple, la simple tàctica de retirar un membre del grup i veure què passa en els resultats demostraria ràpidament que la proporció 0,40 era inestable si provenia de 2/5, però estable si provenia de 200/500 (aquesta tàctica s'utilitza actualment en la moderna estadística computada en un mètode conegut com "*jackknife*"). La tàctica de transferir un membre d'un grup a l'altre s'ha proposat també com a test d'"unitat de fragilitat" en comparar dues proporcions.

Els clínics que estaven desitjosos d'acceptar les matemàtiques, no obstant, varen anar elaborant diverses estratègies per abordar la probabilitat. Desenvolupades al segle XVIII, aquestes estratègies donaren lloc al mètode, formalitzat el 1837 per Siméon-Denis Poisson, per avaluar la "incertesa" dels grups petits. La "llei dels grans nombres" de Poisson per determinar la fiabilitat d'una proporció va ser posteriorment avalada el 1840 per Jules Gavarret i feta extensiva per a comparar dues proporcions. Expressada amb la notació moderna, si les dues pro-

porcions són  $p_1 = r_1/n_1$  i  $p_2 = r_2/n_2$ , el "límit d'oscil·lació" de Poisson-Gavarret pot calcular-se segons

$$2,828 \sqrt{[p_1 q_1 / n_1] + [p_2 q_2 / n_2]}$$

on  $q_1 = 1 - p_1$  i  $q_2 = 1 - p_2$ . Els lectors moderns reconeixeran que el terme arrel quadrada expressa el que avui es coneix com error estàndard de la diferència de dues proporcions i que 2,828 correspon al valor de  $Z_\alpha$  seleccionat (en una distribució gaussiana) per calcular a un interval de confiança  $1 - \alpha$ . Per a  $Z_\alpha = 2,828$ ,  $\alpha = 0,0047$ ; per tant, la proposta de Poisson-Gavarret va establir un interval de confiança del 99,53%, que mostra que l'increment en dues proporcions comparades excedeix els límits de l'"oscil·lació" per l'atzar. Aquest mètode de l'interval de confiança per avaluar la "incertesa" matemàtica havia estat proposat molt abans que l'estadística matemàtica i la bioestadística desenvolupessin les idees sobre els valors  $p$ , raons de versemblança i "significació estadística".

Durant els debats a París sobre medicina *versus* matemàtiques, els principis de Poisson-Gavarret foren tan lloats com atacats. Els partidaris de la sagnia argumentaven (correctament) que els resultats de Louis (i de molts altres usuaris de l'estadística) depenien de grups massa petits que contenien "dades insuficients". Els defensors del mètode numèric argumentaven que les formulacions de Poisson-Gavarret estaven basades no en observacions sinó en les abstraccions de la "probabilitat matemàtica". Una línia de defensa diferent, reconeixent que es necessitarien mostres de grandària enorme per assolir l'aprovació si els nivells d' $\alpha$  tenien un exigent límit de 0,0047, argumentava que aquesta exigència de grups grans "arruïnaria" el mètode numèric.

Quan el lideratge en les ciències mèdiques es traslladà des de França a Alemanya durant el segle XIX, diversos autors alemanys començaren a competir en estadística; finalment, el 1874, Julius Hirschberg plantejà dubtes sobre el manteniment d'un nivell  $\alpha$  del 0,0047, que representava unes possibilitats (*odds*) de 212 a 1. Després d'uns quants càlculs, Hirschberg va proposar un estàndard força més baix, amb unes possibilitats d'11 a 1, el qual fixava essencialment  $1 - \alpha$  a 0,916 i  $\alpha$  a 0,084. El 1877, Carl Liebermeister, que "havia definit la qüestió com una disputa entre les disciplines de la matemàtica i la medicina", pensava que fins i tot un requeriment més baix però satisfactori, d'utilitat pràctica, seria una probabilitat de 5 : 1, això és una  $\alpha = 0,17$ .

Matthews mostra un exemple excel·lent de la manera en què científics clínics esplèndids desconfiaren o s'inhibiren de la demanda probabilística per a una  $\alpha$  "dura". Quan Joseph Lister va introduir el mètode antisèptic en la cirurgia, va mostrar que la taxa de mortalitat havia caigut de 16/35 (46%) a 6/40 (15%). Preocupat per la grandària relativament modesta dels grups, Lister va concloure que "aquests nombres són, sens dubte, massa petits per satis-

fer una comparació estadística”. No obstant, no totalment intimidat, Lister també digué que “quan es consideren els detalls, són altament valorables per a la qüestió que s’està considerant” (si hagués estat accessible la prova de khi quadrat, hauria mostrat que la comparació de 46% *versus* 15% produeix una  $\chi^2 = 8,50$ , segons la qual  $p < 0,005$ , un valor que compleix fàcilment amb la demanda de Poisson-Gavarret de “significança”). Per sort, els èxits de Lister foren acceptats clínicament i l’antisèpsia quirúrgica va prevaldre, malgrat l’aparent absència d’una “comparació estadística satisfactòria”.

Malgrat que Matthews no discuteix més l’argument sobre un nivell apropiat d’ $\alpha$ , el seu límit es va establir arbitràriament més tard a 0,05 quan R. A. Fisher es convertí en la figura principal de l’estadística en la primera meitat del segle xx, quan els valors de  $p$  passaren a ser els mètodes dominants per expressar estabilitat numèrica en assolir “significació estadística”. Tanmateix, la disputa, quant a les expressions d’estabilitat, ha reaparegut en els darrers anys quan els oponents dels valors de  $p$  insten a què siguin reemplaçats pels intervals de confiança. Pel fet que la demanda no s’ha acompanyat de l’establiment d’un nivell  $\alpha$  per calcular els intervals, els estadístics poden explorar lliurement diferents nivells d’ $\alpha$  per obtenir sigui quin sigui l’interval de confiança que els agradaria tenir. Els partidaris del valor  $p$  argumenten que la seva utilització evita la manipulació arbitrària d’ $\alpha$ .

Pels historiadors del futur, la característica més cridadora de la controvèrsia sobre un límit per  $\alpha$  sigui, probablement, l’absència d’atenció a  $\delta$ , que és la magnitud d’increment o d’una altra diferència observada que es considera clínicament rellevant i que s’anàliza posteriorment segons valor  $p$  o interval de confiança. Els clínics moderns han acceptat silenciosament i de manera plena les variables recomanacions estadístiques sobre valors  $p$ , intervals de confiança i límits per a  $\alpha$ , però han fet poc o res per dur a terme les seves pròpies responsabilitats per establir nivells de  $\delta$  clínicament sensibles. A causa d’aquesta mancança, la literatura mèdica de les recents dècades ha contingut moltes afirmacions de resultats “estadísticament significatius” relativament espúries, obtingudes amb grandàries de mostra enormes i per a diferències que eren essencialment petites o trivials.

## 5. Qualitat de les dades bàsiques

En la mesura que la medicina evolucionava al segle XIX, un altre objectiu científic fou obtenir dades que tinguessin alta qualitat en ser objectives, fiables i precises. Aquest objectiu frustrà probablement l’èxit dels deixebles de Louis més que qualsevol disputa sobre els nombres petits i la inadequada individualitat. Les “observacions” en què Louis es va basar per descriure persones vives eren sovint percepcions subjectives com el dolor i altres símptomes, que depenien dels comentaris dels pacients, o sorolls respira-

toris, els qual depenien de les troballes auscultatòries dels metges. Louis havia assenyalat el problema de la variabilitat entre observadors en l’auscultació en dir que “la diferència entre un crepitant, més aviat tosc, i [...] un subcrepitant, més aviat fi [...], pot haver ‘enganyat’ fins i tot experts com Laennec, malgrat els seus ‘sentits altament cultivats [...] (i) delicada [...] oïda’ ”. Per contrast amb tots aquests esdeveniment clínics transitoris, observats subjectivament, la necròpsia ofería teixits que podien ser preservats per a inspeccions repetides i els mètodes de laboratori oferien l’aparent objectivitat de les dades dimensionals.

L’interès en l’avaluació de tractaments declinà probablement també perquè els tractaments d’aquell període no eren particularment efectius, però també perquè el focus de la medicina havia canviat. Els investigadors alemanys i Claude Bernard emfatitzaren que els experiments de laboratori i l’explicació fisiològica farien de la medicina una ciència més “determinista” més que no pas “una ciència conjectural basada en estadístiques”. Així com la base nosològica de les malalties humanes havia canviat a principis del segle xx, des de la malaltia clínica a la capçalera del llit a la malaltia morfològica del dipòsit de cadàvers, el principal desafiament científic per als metges va passar a ser la utilització de mètodes clínics per fer diagnòstics, en pacients vius, que podien ser posteriorment confirmats en l’examen *post mortem*.

Quan l’atenció estadística es va tornar a girar cap a la teràpia, principalment amb els assajos aleatoritzats inaugurats després de la Segona Guerra Mundial, els investigadors volien concentrar-se en dades fiables: mort, malaltia, demografia i proves de laboratori, que havien estat comprovades acuradament pel “control de qualitat”. Aquesta restricció a “dades dures” va eliminar informació clínica important “tova” dels assajos aleatoritzats, però també els va donar uns resultats estadístics excel·lents i fiables que, posteriorment, serien acusats d’estar “deshumanitzats”.

Els mètodes d’ensenyament focalitzats més en l’anàlisi que en l’obtenció de dades, junt amb l’èmfasi analític en “dades dures”, apartà els investigadors estadístics d’una major conscienciació dels problemes de la pròpia informació bruta. El coeficient kappa i els test de McNemar per a la variabilitat entre observadors en dades categòriques van ser desenvolupats per psicòlegs clínics, no per estadístics i els mètodes pragmàtics, desenvolupats per W. J. Youden i W. E. Deming, respectivament, per al control de qualitat en les dades de laboratori i en la manufactura van ser, durant molts anys, menystinguts o oblidats en els centres acadèmics del lideratge estadístic.

Matthews inclou una història sobre el control de qualitat, tanmateix, que il·lustra les contribucions i avenços que es podien haver donat si els investigadors haguessin estat més alerta cap als problemes. Una controvèrsia famosa sobre la inoculació antitifoïdal sorgí el 1904 entre el

bacteriòleg Sir Almroth Wright i el prominent estadístic britànic Karl Pearson, assistit pel seu protegit Major Greenwood. La major part dels metges, avui en dia, recordant la disputa, que va ser enriquida en l'obra de George Bernard Shaw "El dilema del metge", creuen que el principal problema estava en els nombres petits. Els "petits nombres" de Wright sobre els efectes protectors de la immunització han estat fins i tot utilitzats pedagògicament per demostrar que uns resultats aparentment impressionants podien no ser estadísticament significatius.

Matthews assenyala, no obstant, que un origen més important de la disputa era la fiabilitat de la prova de laboratori, l'índex opsònic, desenvolupat per Wright com a guia per al diagnòstic i tractament. Després d'analitzar acuradament els resultats estadístics, Greenwood defensava que el mètode de Wright de comptatge cel·lular en l'índex opsònic era poc fiable. "Les distribucions de freqüència... eren marcadament asimètriques" i les mitjanes no es podien utilitzar per a anàlisis bàsiques o inferències.

Independentment dels mèrits de l'argument de Greenwood i les respostes de Wright, l'episodi indica que la qualitat de les dades pot ser un problema científic major fins i tot en les proves de laboratori. El problema és encara un tema espinós avui en dia, però es dona principalment com una variabilitat no reconeguda i no corregida entre observadors en els radiòlegs, histopatòlegs i citopatòlegs. Un altre assumpte fèrtil per als futurs historiadors serà l'àmplia confiança que es dona actualment, en les estadístiques del càncer i d'altres malalties, a resultats sobre tipus cel·lulars i imatges anormals, que s'han demostrat inconsistents i sovint incorrectes quan aquestes dades morfològiques s'han avaluat adequadament.

#### REFERÈNCIES BIBLIOGRÀFIQUES

1. Vandenbroucke JP. Medicina basada en l'evidència i "la medicina de l'observació". *Annals de Medicina*. 2021;104:87-91.
2. Matthew JR. *Quantification and the quest for medical certainty*. Princeton University Press; 1995. Hi ha una versió en castellà: Matthews JR. *La búsqueda de la certeza. La cuantificación en medicina*. Madrid: Triacastela; 2007.