

Evidence-based medicine: zin en onzin

Philippe Koninckx*

'Evidence-based medicine' (EBM) werd geïntroduceerd door epidemiologen die het belang van 'randomized clinical trials' (RCT) en meta-analyse beklemtonen als wiskundig zuivere en dus zekere modellen. In deze enge definitie is er kritiek gerezen over de moeilijkheden van het uitvoeren en interpreteren van de RCT en over de selectiebias bij de meta-analyse. Indien voorgesteld als 'de standaard voor geneeskundig handelen', dreigt een onderschatting van het klinisch beslissingsproces waar context en individualisering belangrijk zijn. EBM is ontegensprekelijk een pluspunt als nieuwe discipline die geneeskunde en statistiek integreert, die de meest adequate proefopzet en statistische analyse onderzoekt, rekening houdend met kostprijs, tijd en haalbaarheid, die een overzicht probeert te houden over de massale literatuur, en die zich voorstelt als een techniek en niet als een norm. Om dit te illustreren, wordt het belang van de observationele studies beklemtoond en, om het geheel historisch te situeren, wordt het parallellisme met rekenkracht besproken.

Inleiding

De naam 'evidence-based medicine' (EBM) werd geïntroduceerd door Britse en Canadese epidemiologen. EBM beklemtoont het belang van 'randomized clinical trials' (RCT) en meta-analyse (zoals theoretisch ontwikkeld door Kuhn in 1962) als basis voor het medisch handelen. Over de inhoud heerst nogal wat verwarring, gezien de term EBM door sommigen eng gedefinieerd wordt als RCT en meta-analyse. Gezien de kracht van het woord is de betekenis voor velen verschoven naar de letterlijke betekenis, namelijk dat geneeskundig handelen best op evidentie gebaseerd is, en dat deze statistisch onderbouwd dient te zijn in tegenstelling tot 'authority-based medicine'. In deze brede betekenis is er geen discussie, maar dan is EBM enkel een nieuwe term zonder vernieuwing. Over EBM in de enge betekenis van RCT en meta-analyse,^{1,2} geconcretiseerd in de 'Cochrane database for systematic reviews', zijn de meningen verdeeld.³⁻⁵ De kritiek varieert van het beklemtonen van de klinische beslissing en de context in tegenstelling tot de harde gegevens,^{6,9} - lees: beklemtoning van de variabiliteit van de individuele patiënt - tot fundamentele kritiek waarbij het belang van RCT in vergelijking tot andere onderzoeksmethoden voor het opstellen van richtlijnen in vraag wordt gesteld.^{10,11} Hierbij wordt retorisch gesteld dat er geen evidentie is voor de superioriteit van de RCT¹² om de kwaliteit van de geneeskunde te verbeteren en dat EBM een nieuwe vorm is van 'authority-based medicine',¹⁰ zonder omzetting in de praktijk,¹³ terwijl anderen het belang van de klinische beslissing beklemtonen.

Het lijkt daarom nuttig te doorlopen wat de impact en het belang van EBM nu eigenlijk zijn, of dit echt een vernieuwing gebracht heeft, waar de grenzen liggen en welke lading de vlag nu juist dekt.

De geschiedenis

De evolutie van de literatuur in de geneeskunde en de evidentie daarvan worden gekenmerkt door een groeiende complexiteit van de gebruikte statistische analyses, en dit loopt volkomen parallel met de toenemende rekenkracht. De geschiedenis van de evidentie in de geneeskunde kan ingedeeld worden in drie tijdperken: die van de rekenlat, van de rekenmachine en later van de (personal) computer. Elk boek over statistiek leert dat de wiskundige onderbouw meestal vele decennia geleden gelegd werd, net zoals het in de fysica vaak verrassend is dat nogal wat theoretische concepten zo vroeg zijn ontwikkeld. Hetgeen ontbrak was de reken capaciteit om dit op een realistische tijd met voldoende nauwkeurigheid uit te rekenen. Men dreigt te vergeten dat tot de jaren zeventig de rekenlat het enige instrument was om te vermenigvuldigen en delen. Zij die dit nog meegemaakt hebben, weten dat de rekenlat enkel een benaderende schatting gaf door de onnauwkeurigheid van aflezen en de afronding tot 3-4 cijfers, komma's inbegrepen. Bovendien was dit erg tijdrovend en waren vergissingen legio, dus absoluut ongeschikt voor complexe berekeningen. De mechanische rekenmachines waren technische juweeltjes, maar hun prestatie was nauwelijks groter dan die van de rekenlat. Zelfs de 'moderne' elektrische rekenmachines van het begin van de jaren zeventig leverden het resultaat van een deling slechts na minstens 30 seconden luid geratel. Het uitrekenen van een eenvoudige *t*-test tussen twee groepen van 20 metingen vroeg minstens 2-3 uur werk. Voor een doorsnee-artikel was dit 2-3 dagen. Bovendien was de mogelijkheid tot narekenen en controle totaal irrealistisch.

Dit veranderde fundamenteel met de introductie van de rekenkracht van de computer. De eenvoudige beschrijvende statistiek en tests gebeuren bijna automatisch en meer complexe analyses zoals de niet-parametrische tests en de multivariate analyse werden mogelijk. Bovendien ontstonden

een reeks tests waarvan de meeste onder ons wel de naam kennen doch waarvan de exacte betekenis, het belang en het gebruik veel minder duidelijk zijn. Enkele die in ons vakgebied belangrijk zijn, zijn cluster- en 'discriminant' analyse, logistische regressie- en 'life-table' analyse. Elk van deze tests – zoals alle statistische tests – hebben specifieke beperkingen en zijn pas geldig indien aan een reeks veronderstellingen voldaan is. Bovendien bestaat nu reeds de wiskundige onderbouw voor tests waarvoor de huidige rekenkracht van de PC onvoldoende is voor grotere bestanden zoals bij de 'hazard' analyse.

Zeer recent werd een radicale nieuwigheid geïntroduceerd, eveneens gebaseerd op rekenkracht. Het onderzoek was tot nog toe bijna altijd 'hypothesis-driven', waarbij statistiek de hypothese diende te bevestigen. Recent ontstond in de moleculaire biologie het 'non-hypothesis-driven' onderzoek met de introductie van technieken als microarrays, SAGE en proteomics. Deze technieken leveren zo ontzettend veel data dat nieuwe vormen van een analyse dienden te worden ontwikkeld, de zogenaamde bio-informatica. Merk de omkering van de volgorde: menselijk denken en hypothese die geconfermeerd wordt door statistiek versus statistische analyse geïnterpreteerd door menselijk denken. Men zou deze evolutie in de analyse kunnen vergelijken met de bommen van de Tweede Wereldoorlog, later de 'slimme' bommen en anderzijds 'carpet bombing'.

Evidence-based medicine

De term EBM werd een tiental jaren geleden geïntroduceerd door epidemiologen zodra de rekenkracht het toeliet de wiskundige onderbouw daterend van de jaren zestig te implementeren.

'Authority-based medicine' heeft ontegenzeggelijk vergissingen begaan. Het is echter makkelijk te glimlachen bij de hemiclectomieën als behandeling voor hystericie, omdat men dacht dat gassen opstegen naar de hersenen, net zoals wij nu glimlachen bij de verhalen over het draaien van de aarde rond de zon. Deze houding dreigt echter te doen vergeten hoeveel accurate kennis verzameld werd met deze observationele geneeskunde. Hierbij denk ik niet in de eerste plaats aan anatomie, of aan de eerste geschriften rond endometriose bij het begin van deze eeuw, maar aan meer recente ontwikkelingen als de antibiotica en de orale contraceptie, die uit het rekenlatticeperk stammen. Orale contraceptie was een revolutie voor de vrouw. Het was tevens het eerste grootschalige gebruik van medicatie over lange periodes. Dit creëerde terechte bezorgheid en stimuleerde de ontwikkeling van de epidemiologie, die moest toelaten uit deze massa aan gegevens conclusies te trekken over laag frequente neveneffecten op korte en lange termijn. Het was duidelijk dat observationele en cohortstudies verbanden konden aantonen, doch dat het moeilijk zo niet onmogelijk was een causaal verband te leggen. De 'case-control' studie probeerde hieraan te verhelpen, maar de mogelijkheden tot bias blijven zeer groot. De enige wiskundig en

wetenschappelijk juiste manier om het antwoord te geven over één specifiek effect zijn de RCT met in hun verlengde de meta-analyse, die de conclusies van meerdere RCT samenbrengt. Dit kreeg de naam EBM die zich opwierp als promotor van grote prospectieve gerandomiseerde trials in de epidemiologie.

De definitie van EBM is over de jaren progressief verbreed tot de discipline die zich bezighoudt met de meest adequate statistiek voor het beantwoorden van klinische vragen. In deze zin is EBM niet verschillend van wat men in de voorbije decennia probeerde te doen. Wel laat de toenemende rekenkracht het gebruik toe van meer complexe analyses, die de kennis van de doorsneeclinicus overstijgen, zodat EBM in deze definitie een nieuwe discipline is die zich situeert tussen geneeskunde, epidemiologie en statistiek. Voorbeelden hiervan zijn de RCT voor evaluatie van een behandeling, de vergelijkende cohortstudie voor diagnose, de 'inception cohort' voor diagnose en de case-control voor een oorzakelijk verband.

Een tweede verbreding van de definitie van EBM is die van de meta-analyse tot de discipline die zich bezighoudt met het vinden van de beste evidentie in de huidige massa aan literatuur. De noodzaak om op een geordende manier de groeiende massa literatuur te blijven overzien, neemt niet weg dat deze discipline zich beweegt in het spanningsveld tussen de review, veel data, volledigheid maar interpretatiebias door de auteur en anderzijds het reduceren van de literatuur tot enkel RCT met de mogelijkheid tot het uitvoeren van een meta-analyse.

Randomized clinical trials

De RCT zijn wiskundig waterdicht. Praktisch echter zijn er beperkingen. Het is duidelijk dat de RCT de meest geschikte methode is om het effect van een behandeling te evalueren. Het probleem dient zich vooral voor bij grotere trials met een groot aantal patiënten waarbij kleine neveneffecten erg belangrijk zijn voor de volksgezondheid. Laat ons aannemen dat door grote inspanning een perfecte randomisatie en blinde evaluatie kan worden uitgevoerd. Fundamenteel zijn de kritieken over de indicatie tot het uitvoeren van de RCT, over de interpretatie en vooral extrapolatie van de resultaten en over de kostprijs, tijd en inspanning in verhouding tot de verkregen resultaten. De indicatie tot het uitvoeren van RCT situeert zich tussen twee uitersten: ofwel is de bestaande evidentie reeds zo groot dat het bijna onethisch wordt de trial uit te voeren, ofwel is het te verwachten effect zo klein dat het klinisch irrelevant dreigt te worden. De interpretatie en extrapolatie zijn misschien het grootste probleem. Doorgaans worden stricte inclusiecriteria gehanteerd om variabiliteit te vermijden en resultaten beter zichtbaar te maken. Dit impliceert echter dat de conclusies enkel geldig zijn voor deze beperkte groep en eigenlijk niet geëxtrapolleerd mogen worden. Zo is de recente HERS-studie (Heart and Estrogen/progestin Replacement Study) voor hormonale substitutie uitsluitend toepasbaar op oudere

obese Amerikaanse vrouwen behandeld met één specifiek oestrogeen. Of dit ook zo is voor andere groepen patiënten, kan worden bediscussieerd. Dit leidt tot het aspect kostprijs en tijd. Rekening houdend met het feit dat op slechts één vraag één antwoord gegeven wordt bij een specifieke en doorgaans geselecteerde populatie, lijkt de RCT een weinig performante test, die het best gereserveerd wordt om essentiële vragen met een belangrijke impact op de volksgezondheid te beantwoorden. Bovendien zijn de RCT minder geschikt voor een reeks klinische problemen. Voor chirurgie is een stricte standardisatie van techniek, chirurg, indicaties en geassocieerde problematiek vaak onverzoenbaar met de aantallen die noodzakelijk zijn voor het beantwoorden van sommige vragen met voldoende zekerheid. Bovendien is de evolutie in technologie doorgaans zo snel dat trials vaak technisch obsoleet worden voordat voldoende aantallen zijn verkregen.

Ook de meta-analyse is wiskundig waterdicht. Het probleem is echter dat de selectie van artikelen doorgaans vrij arbitrair is, omdat artificieel een keuze gemaakt wordt voor 'goed genoeg', omdat 'perfect' bijna niet bestaat. Een beknopte samenvatting van enkele honderden artikelen tot een tiental RCT en de verkondiging van de waarheid via de meta-analyse geven soms de indruk van een gemakkelijksheidsoplossing, eerder dan het eerlijk zoeken naar de waarheid.

Voorbeelden lopen altijd mank en hebben slechts een anekdotische waarde. Wel kunnen zij sommige zaken verduidelijken. De RCT en meta-analyse van humaan menopauzaal gonadotrofine (HMG) versus recombinant follikel-stimulerend hormoon (rec-FSH) kunnen worden geïnterpreteerd als een bewijs dat de resultaten met rec-FSH beter zijn. De conclusie dat wat meer activiteit per ampul beter is, is echter even aannemelijk.

Observationele studies

Het voordeel van observationele studies is dat er weinig selectiebias is en dat alle gegevens voor analyse gebruikt worden. Het nadeel is dat de interpretatie niet toelaat causale verbanden met zekerheid vast te leggen. Voor snel evoluerende disciplines als chirurgie blijft het zeer de vraag of de eerlijk descriptieve multivariate analyse van resultaten niet de meest performante manier is om gegevens te verzamelen, zeker indien men rekening houdt met het feit dat de resultaten van andere groepen gemakkelijk kunnen worden vergeleken. Op een later tijdstip kunnen belangrijke vragen dan door RCT worden ingevuld. Belangrijk is dat men zich realiseert dat de observationele studies de gegevens leveren voor de noodzakelijke schattingen om de RCT adequaat te plannen, dat de observationele studies de meerderheid van de informatie levert, terwijl de RCT enkel deelaspecten confermeren, en dat de observationele studies de enige zijn die

informatie verschaffen over de extremen als bijvoorbeeld de zeer obese patiënt.

Besluit

De geschiedenis van de rekenapparatuur loop parallel met het soort statistische analyse dat gebruikt werd en wordt. Het belang van de observationele studies blijkt uit het verleden, terwijl de huidige rekenmogelijkheden voor grote databases deze observationele studies aan belang doen toemen. Als dusdanig kan dit worden vergeleken met het 'non-hypothesis-driven' onderzoek. Observationele studies hebben weinig selectiebias, een lage kostprijs in verhouding tot de verkregen informatie, en laten een praktisch continue evaluatie toe. Deze aspecten worden belangrijker voor de extremen van de populatie of voor een techniek die nog in volle evolutie is zoals de endoscopie. Alhoewel men altijd getracht heeft statistisch zo accuraat mogelijk informatie te verzamelen, ligt het belang van de EBM in het benadrukken van randomisatie en van de RCT. De RCT zijn theoretisch wiskundig ontegensprekelijk het meest aannemelijke onderzoek voor het effect van een behandeling te evalueren. Al te vaak dreigt men echter te vergeten dat de RCT, opgezet voor het opsporen van kwantitatief kleine effecten bij grote groepen patiënten, vaak moeilijk organiseerbaar zijn, vaak lang duren en duur zijn in verhouding tot de gegenereerde gegevens. Bovendien zijn de resultaten slechts toepasbaar voor de beperkte groep die samenvalt met de selectiecriteria. Daarom dreigt het ongenueanceerd gebruik van EBM, eng gedefinieerd, als RCT en meta-analyse een verarming te zijn.

EBM zou nuttig en belangrijk zijn indien deze gedefinieerd wordt als een nieuwe discipline die zich situeert tussen geneeskunde, statistiek en epidemiologie, die de meest adequate studieopzet en tests beoogt zodat maximale informatie wordt verkregen op de kortst mogelijke tijd tegen de laagste kostprijs en met de grootste zekerheid, en die zich tot doel stelt methoden te ontwikkelen om een overzicht te houden op de ganse literatuur en hieruit de best beschikbare evidentie te halen. In deze definitie vervaagt het belang van de RCT tot een test tussen de velen. De observationele studie blijft de basis van het geneeskundig onderzoek met daarna specifieke tests waaronder de RCT die sommige aspecten hiervan nader onderzoeken.

Tot slot dreigt het risico dat een ongenueanceerd gebruik van EBM sommigen laat stellen dat alles wat niet met zekerheid is aangetoond, niet zou moeten worden gebruikt. Dit is misschien waar voor de toekomst, indien men goed beseft dat het zeer moeilijk kan zijn statistisch de afwezigheid van effect met zekerheid uit te sluiten. Ik meen dat het voorzichtiger is te stellen dat, vooraleer men de ervaring van jaren geneeskundig handelen overboord gooit, het minstens met zekerheid zou moeten worden aangetoond dat dit niet nuttig is.

Summary

Evidence-based medicine (EBM) has been introduced by epidemiologists to emphasize the randomized clinical trials (RCT) and meta-analysis as providing the best evidence, since they are mathematically pure. Defined as such, we have seen criticism in the literature regarding feasibility and interpretation of the RCT and selection bias in meta-analyses. If proposed as the standard for medical practice, EBM runs the risk of disregarding the importance of individualization and context for medical decision-making. EBM can also be defined as a new discipline, encompassing medicine, statistics and epidemiology. The aim is to define the best available design and analysis, taking time, cost and efficacy into account and to develop methods to make a comprehensive summary of the vast literature, considering quality of information. Regarded as a tool in medicine rather than the standard, this can only be welcomed. To illustrate this, the importance of observational studies has been discussed, while the parallel between computing power and statistical testing is a lesson in modesty.

Referenties

1. Raspe H. Evidence-based medicine: Modischer Unsinn, alter Wein in neuen Schläuchen oder aktuelle Notwendigkeit? (Evidence-based medicine: popular nonsense, old wine in new bottles or current necessity?) *Z Ärztl Fortbild Jena* 1996; 90: 553-62.
2. Sackett DL, Rosenberg WM. The need for evidence-based medicine. *J R Soc Med* 1995; 88: 620-24.
3. Charlton BG, Miles A. The rise and fall of EBM. *QJM* 1998; 91: 371-4.
4. Fugelli P. Clinical practice: between Aristotle and Cochrane. *Schweiz Med Wochenschr* 1998; 128: 184-8.
5. Egger M. Unter dem Metaskop: Möglichkeiten und Grenzen von Meta-Analysen. (Under the meta-scope: possibilities and limits of meta-analyses). *Schweiz Med Wochenschr* 1998; 128: 1893-901.
6. Collins JA. Clinical research evidence and clinical practice. *Hum Reprod* 1997; 12: 1847-50.
7. Fletcher RH. Practice guidelines and the practice of medicine: is it the end of clinical judgment and expertise? *Schweiz Med Wochenschr* 1998; 128: 1883-8.
8. Greenhalgh T, Worrall JG. From EBM to CSM: the evolution of context-sensitive medicine (editorial). *J Eval Clin Pract* 1997; 3: 105-8.
9. Jacobson LD, Edwards AG, Granier SK, Butler CC. Evidence-based medicine and general practice (see comments). *Br J Gen Pract* 1997; 47: 449-52.
10. Couto JS. Evidence-based medicine: a Kuhnian perspective of a transvestite non-theory. *J Eval Clin Pract* 1998; 4: 267-75.
11. Sondanaa K, Nesvik I, Solhaug JH, Soreide O. Randomization to surgery or observation in patients with symptomatic gallbladder stone disease. The problem of evidence-based medicine in clinical practice. *Scand J Gastroenterol* 1997; 32: 611-6.
12. Vennin P. Evidence-based medicine (EBM) ... la médecine de l'évidence ... une évidence? (Evidence-based medicine (EBM)...an evidence?). *Bull Cancer* 1998; 85: 358-59.
13. Aveyard P. Evidence-based medicine and public health. *J Eval Clin Pract* 1997; 3: 139-44.

* Prof. dr. Ph. Koninckx
 Dienst Verloskunde-Gynaecologie
 UZ Gasthuisberg
 KU Leuven
 3000 Leuven
 België