

# BIG DATA, BIG EXPECTATIONS? ETHISCHE REFLECTIES

PROF. DR. SIGRID STERCKX

UNIVERSITEIT GENT (BIOETHICS INSTITUTE GHENT)

UNIVERSITY OF OXFORD (CENTRE FOR HEALTH, LAW & EMERGING TECHNOLOGIES)

# INSTEELK / PERSOONLIJKE INTERESSE

Er zijn verschillende redenen waarom en manieren waarop **het 'Big Data paradigma' vragen opwerpt of her-formuleert die ethisch-filosofisch relevant (en uiterst boeiend!) zijn**

**Doel: een aantal van die vragen trachten te schetsen, met focus op besluitvorming m.b.t. gezondheidszorg (op verschillende niveaus)**

# WAAR KUNNEN DIE VRAGEN ZOAL ONTSTAAN?

## 5 aspecten / 'stadia':

- Het **collecteren van data** en het beoordelen van hun kwaliteit
- Het onderzoeken in welke mate die 'data' '**evidence**' kunnen worden
- Het gebruik van RCD door de **regulator** bij de beoordeling van 'evidence' in het kader van de verlening van markttoegang voor én surveillance van therapeutica én medical devices
- Het gebruik van RCD in de context van **geautomatiseerde systemen** voor clinical decision support (of decision-making)
- Het gebruik van RCD om **Shared Decision Making** (arts en patiënt samen) robuuster en helderder te maken

# MAAR: CAVEAT: HET BEGINPUNT IS EIGENLIJK NIET 'DATA'

"This 'mythological' view of Big Data as the 'end of theory' creates ethical concerns regarding justification of increasingly pervasive and unbounded secondary manipulation and aggregation of data when Big Data practices are seen as the future of science and scientific discoveries. Data are (mistakenly) said to "speak for themselves", creating the possibility of science being driven entirely by induction and reduction, or 'data-driven science' without a need for theory or hypotheses ... For proponents, the "sheer abundance of information" is seen as providing a "degree of scientific authority" to Big Data practices, which can seemingly be used to explain any "natural and social phenomena" ... because its meaning is "already there, just waiting to be uncovered" ... This idea of data-led objective discoveries entirely discounts the role of interpretive frameworks in making sense of data which, according to non-objectivist ontologies (e.g. ... Heidegger 1967), is a necessary and inevitable part of interacting with the world, people and phenomena (and thus, data)."

Mittelstadt & Floridi 2016 Science & Engineering Ethics: 320

# KENNISGENERATIE IN GENEESKUNDE IS SUBOPTIMAAL

- In elk 'stadium' in het 'schema' kunnen problemen en kortsluitingen optreden en dan is wat wordt verkregen helemaal geen 'evidence'
- Misschien wordt de situatie dan i.p.v. beter, nog slechter dan nu: "Medical Misinformation Mess":  
"Most physicians and other healthcare professionals are unaware of the pervasiveness of poor quality clinical evidence that contributes considerably to overuse, underuse, avoidable adverse events, missed opportunities for right care and wasted healthcare resources. **The Medical Misinformation Mess comprises four key problems.** First, **much published medical research is not reliable or is of uncertain reliability**, offers no benefit to patients, or is not useful to decision makers. Second, **most healthcare professionals are not aware** of this problem. Third, **they also lack the skills** necessary to evaluate the reliability and usefulness of medical evidence. Finally, **patients and families frequently lack relevant, accurate medical evidence and skilled guidance** at the time of medical decision-making. Increasing the reliability of available, published evidence may not be an imminently reachable goal. Therefore, **efforts should focus** on making healthcare professionals, more sensitive to the limitations of the evidence, training them to do critical appraisal, and enhancing their communication skills so that they can effectively summarize and discuss medical evidence with patients to improve decision-making."

# GENEESKUNDE IS OPAQUE (ONDOORZICHTIG, 'BLIKDICHT')

Zeer boeiend artikel uit 1975: **Samuel Gorovitz & Alasdair MacIntyre, Toward a Theory of Medical Fallibility**, *Hastings Center Report* 5: 13-23:

Hun vertrekpunt: "Medicine as a practice is more opaque than we normally take it to be." (OOK intern!)

**Op het eerste gezicht lijken Big Data die opaciteit (én de opaciteit van de communicatie over geneeskunde) nog groter te maken**: Voorbeelden:

**Voorbeeld 1**: De Tijd, 23 april 2018: interview met de CEO van Philips:

“We zetten sterk in op kunstmatige intelligentie en big data in de gezondheidszorg. De diagnostiek wordt daardoor preciezer en de productiviteit van de gezondheidszorg kan fel verbeteren. Informatica kan nieuwe relaties bouwen tussen patiënten en dokters. ... Het is een markt met veel toekomstmuziek. De wereldbevolking gaat van 7 naar 9 miljard mensen. Die zullen langer leven en meer chronische ziektes hebben.”

# THE SAME OVERRATED PROMISES AGAIN AND AGAIN

## Voorbeeld 2:

“[H]ow about the other pieces of EBM, for example, diagnosis and prognosis and individualizing care? I have had great excitement about the prospects of omics, big data, personalized medicine, precision medicine, and all. Much of my effort has been to put together these efforts with rigorous statistical methods and EBM tools. But I am tired of seeing the same overrated promises recast again and again. [S]everal years ago I gave an invited lecture at a leading institution on the danger of making inflated promises in personalized medicine. Right after my talk, everybody rushed to hear the launch of a new campaign, where the leader of the institution singled out this unique historic moment: that institution would single-handedly eliminate most major types of cancer within a few years. Several years have passed, and none of these cancer types have disappeared. ... [T]here are already millions of unique historic moments where cancer was eliminated. Same applies to neurologic diseases and more. I do not understand why academic leaders and politicians need to make such selfembarrassing announcements now and then.”

Ioannidis, *Journal of Clinical Epidemiology* 73 (2016): 84

# GENEESKUNDE IS OPAQUE (ONDOORZICHTIG, 'BLIKDICHT')

## Terug naar Gorovitz & MacIntyre: Theory of Medical Fallibility:

- Die opaciteit moet worden gereduceerd.
- Eerste deel van hun artikel is een epistemologische analyse van de vraag 'welk soort wetenschap is de geneeskunde eigenlijk?' [ja, ook een kunde of een kunst]
- We moeten weten welk soort wetenschap geneeskunde is opdat we zouden kunnen weten welke soorten fouten kunnen optreden
- Vaak gehoorde visie: fouten in een wetenschap kunnen maar twee mogelijke oorzaken hebben:
  - “ignorance” (stand van de wetenschap is/was nog niet voldoende gevorderd); of
  - “ineptitude” (onachtzaamheid, incompetentie vanwege de wetenschapper)



# GENEESKUNDE IS OPAQUE (ONDOORZICHTIG, 'BLIKDICHT')

Gorovitz & MacIntyre leggen uit dat deze visie verkeerd is, omdat **er een derde oorzaak is van fouten die bij uitstek relevant is voor de geneeskunde als een “science of particulars”**:

“Natural science did not in the seventeenth century discard quite as much of Aristotelianism as its philosophical protagonists supposed. What it retained included an inability to give a plausible account of our knowledge of particulars, of individuals ... For natural science ... the objects of knowledge are universals, that is, the properties of objects classified by *kinds*, and the generalizations that link those properties. The scientist looks for law-like relationships between properties ... To explain the behavior of a particular is nothing else than to subsume its particular properties under the relevant law-like generalization; to predict is to use the same stock of law-like generalizations about the relevant properties.

[However] in the case of particulars one cannot expect to be able to move from a theoretical knowledge of the relevant laws to a prediction of the particular's behavior. **The history of [the particular] is, so to speak, an intervening variable which may always to some degree elude us.**”

# GENEESKUNDE IS NOODZAKELIJKERWIJZE FEILBAAR

Hieruit concluderen zij dat de feilbaarheid van de geneeskunde een noodzakelijke, onvermijdbare feilbaarheid is: “it is inherent in the nature of medical practice that error is unavoidable, not merely because of the present limitations of human knowledge or even the limits of human intellect, but rather because of the fundamental epistemological features of a science of particulars.”

Vraag die mij het meest fascineert i.v.m. Big Data: hoe kunnen Big Data helpen om dat fundamentele epistemologische inzicht over de aard van geneeskunde als een ‘science of particulars’:

- eerst en vooral meer zichtbaar te maken (minder opaciteit – zowel intern als extern); en
- Vervolgens, dank zij die transparantie, ruimte te creëren voor meer accountability & fairness

Zo kan geneeskunde (nog) beter en (nog) ethischer worden

Hoe kunnen we dit doen in het ‘Big Data’ tijdperk? **WELKE VRAGEN MOETEN WE STELLEN EN BEANTWOORDEN OM DIT DOEL TE KUNNEN BEREIKEN??**

# CRUCIALE VRAGEN: 1

(a) Welke '**technische**' hinderpalen kunnen de **KWALITEIT VAN DE DATA** ondermijnen?

(bvb. ontbrekende data, verkeerde data, gebrek aan harmonisering en interoperabiliteit, etc.)

(b) Welke **economische incentives** kunnen dit kwaliteitsgebrek nog verergeren?

(bvb. 'upcoding' om meer geld te kunnen innen, wetende dat de data als bron voor facturatie worden gebruikt)

# CRUCIALE VRAGEN: 2

Welke factoren kunnen (terecht of onterecht) **TOEGANG TOT DE DATA (EN DUS GEBRUIK VAN DE DATA) belemmeren?** (a) vormen van controle en eigendom van de data en producten gerelateerd aan de data; (b) privacy & gegevensbescherming

(a) bvb. geheimhouding van algoritmen via **fabrieksgeheim** (trade secret protection = IP) belet validering van automatische systemen en transparantie (bvb. ‘obligation to explain decisions’); bvb. voorstellen voor nieuwe soorten eigendomsrechten op data (cf. Max Planck Instituut München: voorstel voor exclusieve ‘data rights’ om downstream gebruik te kunnen controleren)

(b) Privacy & data protection: **vaak te enge focus op health data**, terwijl we ook zouden moeten nadenken over de bescherming van **‘quasi-health data’** en **‘emergent medical data’**

Health data: zie bvb. aankondiging Facebook begin april ivm samenwerking met ziekenhuizen om ‘geanonimiseerde’ patiëntendata over diagnoses en voorgeschreven geneesmiddelen te bekomen om te combineren met data van FB gebruikers en zo digitale profielen van patiënten te construeren

## CRUCIALE VRAGEN: 2

Health data: zie bvb. rechtszaak ingeleid op 15 maart 2016, nu in Beroepsfase:

**UNITED STATES DISTRICT COURT NORTHERN DISTRICT OF CALIFORNIA**

WINSTON SMITH; JANE DOE I; and JANE  
DOE II, on behalf of themselves and all others  
similarly situated, Plaintiffs,

v.

FACEBOOK, INC.; AMERICAN CANCER SOCIETY, INC.; AMERICAN  
SOCIETY OF CLINICAL ONCOLOGY, INC.; MELANOMA RESEARCH  
FOUNDATION; ADVENTIST HEALTH SYSTEM; BJC HEALTHCARE;  
CLEVELAND CLINIC; and UNIVERSITY OF TEXAS – MD ANDERSON  
CANCER CENTER, Defendants.

## CRUCIALE VRAGEN: 2

“Plaintiffs’ cancer and other sensitive health-related Internet communications with these medical websites were divulged to Facebook and acquired by Facebook along with the Plaintiffs’ personally-identifiable information. In addition, Facebook acquired, tracked, and used the Plaintiffs’ sensitive medical information collected through medical websites and the Facebook website for purposes of direct marketing.

The disclosures, tracking, and use of their sensitive medical information for direct marketing were all done without Plaintiffs’ knowledge or consent in violation of their privacy rights under federal and state law.

Defendant Facebook failed to disclose to its users that it (a) tracks, intercepts, and acquires user communications in violation of other websites’ privacy policies, (b) tracks, intercepts, and acquires user communications with medical websites, including the websites of medical providers subject to HIPAA and other medical privacy laws, and (c) uses the personal information it gathers from its users, including sensitive medical information, to place its users into medical categories for purposes of direct marketing.

# CRUCIALE VRAGEN: 2

## 2<sup>e</sup> categorie: Quasi-health data:

- Voornamelijk uit health & fitness apps en wearables: minder specifiek dan health data, maar ze hebben een duidelijk verband met gezondheid – vallen niet onder bescherming van HIPAA; wellicht ook niet onder GDPR ?
- Zie bvb. Malgieri & Comandé (2017), Sensitive-by-distance: quasi-health data in the algorithmic era, *Information & Communications Technology Law* 26: 229-249.

## CRUCIALE VRAGEN: 2

3<sup>e</sup> categorie: ‘emergent medical data’: via AI (machine learning) infereert FB gevoelige medische informatie over al zijn gebruikers

Mason Marks, Yale Law School, onderzoekt de toepassing van AI op clinical decision making): blog over hearing Mark Zuckerberg in het Amerikaanse Parlement: “few Members of Congress seemed to understand that the value of data often lies not in the information itself, but in the inferences that can be drawn from it.” Zuckerberg verklaarde dat FB geen “user data” verkoopt, maar op de vraag of FB “information that it infers from user data” verkoopt, kwam totaal geen duidelijk antwoord.

Mason Marks: “[There is] a **new type of medical information that is not protected by existing privacy laws**. I call it Emergent Medical Data (EMD) because at first glance, it has no relationship to your health. Companies can derive EMD from your seemingly benign Facebook posts, a list of videos you watched on YouTube, a credit card purchase, or the contents of your e-mail. A person reading the raw data would be unaware that it conveys any health information. Machine learning algorithms must first massage the data before its health-related properties emerge.”



## CRUCIALE VRAGEN: 3

- Als de voorwaarden vervuld kunnen worden om RCD als ‘evidence’ te kunnen beschouwen, hoe kan die evidence dan uitmonden in voordelen voor patiënten door: **RCD te gebruiken om op een meer verantwoorde manier GENEESMIDDELEN EN MEDICAL DEVICES OP DE MARKT TOE TE LATEN EN TE MONITOREN?**

Zie bvb. het ‘**Adaptive Pathways**’ project van het **EMA** om markttoegang te versnellen door de facto Fase III studies te vervangen door een conditionele markttoelating plus post-marketing verzameling van bewijs op basis van RCD : Hoe geloofwaardig, transparant en afdwingbaar gaat dit zijn?; Gaat dit meer zijn dan een ‘verschuiving’ van risico’s van ontwikkelaars van geneesmiddelen naar artsen en patiënten? (cf. duidelijk weinig handhaving van post-marketing verplichtingen m.b.t. ‘accelerated access’ programma’s in de VS (Cifu & Prasad 2015, *Ending Medical Reversal*); Wat met de aansprakelijkheid van artsen / verzekeraars die onder een Adaptive Pathway conditioneel goedgekeurde geneesmiddelen (niet) voorschrijven / terugbetalen?

# CRUCIALE VRAGEN: 4

- Als de voorwaarden vervuld kunnen worden om RCD als ‘evidence’ te kunnen beschouwen, hoe kan die evidence dan uitmonden in voordelen voor patiënten door: **RCD te gebruiken als ‘input’ voor GEAUTOMATISEERDE BESLUITVORMING M.B.T. GEZONDHEIDSZORG?**

Hoe gaat men de gigantische uitdagingen i.v.m. validiteit en betrouwbaarheid van dergelijke tools aanpakken? Of gaat men die gewoon veronderstellen? Als de algoritmen theorie-vrij zijn (data-driven) ontberen ze elke vorm van domeinkennis en genereren ze massa’s ‘spurious’ correlaties. Als ze niet louter data-driven zijn bestaat het risico dat allerlei vormen van discriminatoire biases erin sluipen en dat de besluitvorming meer unfair wordt.

Wat met de juridische aansprakelijkheid van artsen en ziekenhuizen voor het gebruik van dergelijke tools? Wat met hun ethische verantwoordelijkheid voor de besluitvorming? Hoe kunnen zij controle uitoefenen over dergelijke systemen?

Welke onbedoelde effecten kan dit hebben? (volgende)

# CRUCIALE VRAGEN: 4

- Welke onbedoelde effecten kan dit hebben? Bvb. Cabitza et al. 2017 JAMA 318(6):

“[A] study of 30 internal medicine residents showed that the residents exhibited a decrease in diagnostic accuracy (from 57%to 48%) when electrocardiograms were annotated with inaccurate computer aided diagnoses. Further research is needed to better understand whether the overreliance on [machine learning systems] could also cause a subtle loss of self-confidence and affect the willingness of a physician to provide a definitive interpretation or diagnosis.”

“[It can also] lead to focusing more on what can be rendered as text (ie, data) at the expense of other elements that are more difficult or impossible to easily describe. Relying on ML-DSS requires considering digital data as reliable and complete representations of the phenomena that these data are supposed to render in a discrete and trustworthy form. This may be a problem when the clinical context is not represented ... Such lack of information may lead to partial or misleading interpretations of ML-DSS diagnostics and therapeutic or prognostic outputs.”

# CRUCIALE VRAGEN: 5

- Als de voorwaarden vervuld kunnen worden om RCD als ‘evidence’ te kunnen beschouwen, hoe kan die evidence dan uitmonden in voordelen voor patiënten door: **RCD te gebruiken om MEER EN BETERE SHARED DECISION MAKING te bewerkstelligen?**

SDM = balans tussen EBM kennis, klinische expertise van de arts, en waarden en voorkeuren van de patiënt

Staat of valt met begrip van de informatie, dat vaak ontbreekt doordat de statistische geletterdheid van patiënten én artsen erg laag is.

Tegelijkertijd wordt in heel wat settings het aantal behandelingsopties groter.

- Hoe kunnen RCD bvb. worden geïncorporeerd in grafische representaties die dit probleem kunnen aanpakken en dus zowel de arts als de patiënt in staat te stellen tot betere besluitvorming?

DANK VOOR UW AANDACHT

sigrid.sterckx@ugent.be

Dank aan Prof. Johan Decruyenaere, Prof. Wim Van Biesen, Dr Johan Steen, Lisa Diependaele