

## Healthcare industry BW

### Big Data

# “More data does not automatically imply more knowledge“

**In Germany, one of the big names in evidence-based medicine is Prof. Dr. rer. nat. Gerd Antes, co-director of Cochrane Germany. In the following interview, Antes talks about the hype surrounding big data, warns against false promises and reminds us about what is taken for granted.**

Big data is seen by some as our saviour in the field of medicine. As a representative of evidence-based medicine, do you share this view?



The usefulness of big data does not need to be looked at through the prism of evidence-based medicine. Simply by applying basic scientific criteria, we very quickly realise that the hype behind big data's potential creates a completely one-sided view. Evidence-based medicine acts for the benefit of patients, at the same time as reducing the risk of harm and saving costs. However, in amongst all the praise of big data, it quickly becomes clear that harm and cost have been simply “forgotten”, and no concerns about these issues have been raised at all. This is a fatal error, especially as far as the introduction of new technologies and major changes are concerned. Any technology assessment generally weighs benefit, harm and costs against each other. Even non-experts are aware that these issues have not been raised. But as big data is surrounded by so much hype, no one dares to speak out.

More data does not automatically imply more information. More data may well mean

more false information. Of course, there are beneficial aspects, but they must be identified and specifically funded. Big data does not mean that we need fewer restrictions or more money. What we need is

faster assessment of drugs and methods that have real potential.

Some people believe that prospective randomised controlled trials (RCTs) are the gold standard when it comes to evaluating diagnostic and therapeutic procedures. Others criticise their experimental set-up and question their validity for patients under everyday conditions (real-life data). What should we think about such conflicting views?

The emphasis on real-life data is one of the major undesirable developments in our search for knowledge. Science produces right and wrong, and thus also "false positives". This term comes from the field of diagnostics, but applies to knowledge generation in general. There are statements that big data mechanisms characterise as new knowledge, but which are in fact wrong. Unless we apply the rigorous methodology that Ioannidis calls for (see below), we will end up being overrun by a wave of false positives. Ioannidis' methodology is not error-free either, but much safer than observational studies or huge amounts of data that no longer seem to need validation.

Would it not make sense to also take into account data generated by imaging methods, genome sequencing approaches, registries, observational studies, or secondary data in general?

Yes. Basically, this would make sense. However, it has been 20 years since evidence-based medicine first emerged and we have learned that taking such data into account can be misleading. Let me give you two examples. Health insurance companies tend to analyse their billing data, but due to the billing systems used, these data do not provide any disease-related information. If conclusions on medical treatments are drawn from billing data, it goes without saying that the information will be wrong. The same applies to genome-associated data. It is naïve to think that a genetic "switch" that is identified for a certain disease only needs to be turned off in order to prevent disease symptoms from occurring. (ed. note: such thinking would only apply for monogenic diseases which result from modifications to a single gene.) This kind of information generates a lot of false hope. Why? Because a disease is controlled by many genetic switches rather than just one. Moreover, all these switches interact with each other. Bioinformaticians have been tearing their hair out for years trying to find a way to disentangle such interactions. What happens now? People who see big data as the great saviour claim that this will now be possible in the blink of an eye.

What scientific requirements need to be met by data that is closer to medical routine, i.e., reflects routine medical care?

This must be observational studies. But if you missed out the required initial steps, i.e. providing proof of principle, and immediately launched a trial that gathers information about patients in the social context of their day-to-day lives, you will never work out whether the drug is effective or not. With such an approach, you find yourself on extremely thin ice. Why? Because effectively I am introducing something and asking patients what they think about it. If I ask the "right" people, i.e. those who are happy with the treatment they are being given, I will come to the conclusion that it was "successful".

To what extent could real-life data be integrated into RCTs (randomised controlled trials) and possibly help improve the evaluation of diagnostic and therapeutic procedures? Are there examples of this?

This has been done for 30 years. RWD\* have been generated in the drug development field for a long time. Phase IV studies, which are carried out after marketing authorisation has been obtained, are randomised trials supplemented with market data.

Well, here is what I believe: We have almost everything that we need to be able to integrate real-world data into RCTs. All we have to do now is apply a particular method more rigidly. The method in question is one that has been developed over many years by many bright people since 1932 when randomisation was first introduced. Now we're seeing people coming along and trying to cause trouble. However, the arguments that they put forward are not scientific. There are articles and books that are actually calling for well-established theory to be replaced by data science. Such books claim that the deluge of data is making it much easier to solve all our problems. From a scientific perspective, this is probably the biggest nonsense ever.

Personalised medicine, which relies on omics data, sometimes results in one-patient trials. Is this an alternative to RCTs?

Of course not. For decades, we have been making it clear that a single case cannot be used to generalise about other patients. Scientists are hoping to achieve something like this with precision medicine. They believe that by deciphering all the human genes, this would give us all the information we need to know about human architecture and thus provide the key to switching off disease symptoms. I'd like to mention an article that was written by Gasser on the pathology of Parkinson's (see below). He comes to the conclusion that all that is needed to develop drug therapies is the following: limiting the number of false-positive results, finding the right information faster, being faster in excluding wrong aspects, and minimising errors in general.

Are there any examples where big data has been integrated into studies that followed the principles of scientific evidence?

We need to look at how big data is defined or rather, not defined. This is what has led to the confusion. On top of that, there are books like "Das Ende des Zufalls. Wie die Welt vorhersagbar wird" (The End of Chance. How the world becomes predictable) that cannot be taken seriously from a scientific point of view and stop people from really thinking about and critically evaluating these kinds of issues. That's an example of how not to approach big data but there are many positive examples as well. It would be a mistake, not to mention virtually impossible, to condemn progress; for example, registers with complete datasets that can be used to answer medically very relevant questions. All this is available, but that's not what we consider big data.

Big data is the promise that if we only had access to all data, we would be able to address any issue with entirely non-transparent tools. A negative example of this is IBM's supercomputer Watson, which IBM claims can make the world a better place. The supercomputer is not living up to the expectations IBM created for it. There is no evidence for that it is achieving what it set out to do. Moreover, MD Anderson, one of the largest UC cancer hospitals in the USA, has just dumped Watson.

The established principle of orthodox methodology used in medical decision-making is to minimise systematic mistakes (risk-of-bias) and thus minimise the intrinsic risk of generating systematically wrong results. Artificial intelligence, however, does not recognise the issue of systematic errors at



Colour representation of the activity of a Wikipedia bot over a prolonged period of time: typical example of how big data is visualised.

©

[CC license](#)

all because, as its supporters claim, by default, it does everything right. However, evidence for this ambitious claim is largely lacking. Relevant publications always refer to the same examples when highlighting the medical success that can be achieved with big data. One such claim is that Google was believed to have successfully used an algorithm to estimate flu pandemics. However, this claim has already been refuted (Nature 494, 155–156 (14 February 2013) DOI:10.1038/494155a). The tracking of seasonal flu levels worked for two years, but not for the third. Why? Google just got lucky in the first two years. Despite the evidence, Google’s ability to estimate peak flu levels is still advertised as a success. Similar success stories largely go back to successful sales promotion by way of personalised advertising, a concept which I do not think should be taken as a model for the field of medicine.

What needs to happen so that the much-vaunted potential of big data in the field of medicine can really benefit the patient?

People have to remember that the patient is the focus. If you look at funding programmes, some are downright harmful. I have been observing the biologisation of medicine for many years, which means that the patient no longer has a role to play. Billions of euros are invested in the wrong thing, and the knowledge acquired is not turned into treatment methods or implemented in a targeted way. Despite calling for research results to be translated into practice, we are

experiencing an extreme asymmetry in the use of funding in favour of big data and digitisation, while the implementation of findings made with humans has to be subject to cost-effectiveness analyses. It is essential that the final step of studies is based on valid methods, and the efficacy of a particular treatment on rigidly defined empirical methods. Success must be measured by patient-relevant results, not by the number of databases or the digitisation level of clinics and medical practices. Big data must be seen as a tool for medicine, not as the objective.

Ed. note: RWD = real-world data

#### References:

<http://www.zeit.de/2017/25/medizin-studien-forschung-fehler-pharmaindustrie-john-ioannidis/>  
Thomas Gasser, Das Zittern hat eine Geschichte, in: FAZ, 17.5.2017,  
[http://www.ghst.de/fileadmin/images\\_redesign/neurowissenschaften/Alterserkrankungen/FAZ\\_Gasser.pdf](http://www.ghst.de/fileadmin/images_redesign/neurowissenschaften/Alterserkrankungen/FAZ_Gasser.pdf)

#### Papers published by Prof. Dr. Gerd Antes:

<http://www.laborjournal.de/editorials/981.lasso>  
<https://www.aerzteblatt.de/archiv/175874/Big-Data-und-Personalisierte-Medizin-Goldene-Zukunft-oder-leere-Versprechungen>

#### Further reading:

Evidence-based Medicine Network: <http://www.ebm-netzwerk.de/>  
Mehr Daten kann auch mehr falsche Infos bedeuten: Meng, XL /Xie, X.: I got more data, my model is more refined, but my estimator is getting worse! Am I just a dumb?, *Econometric Reviews*, Vol. 33, 2014,  
<http://dx.doi.org/10.1080/07474938.2013.808567>  
Traue keinen Anekdoten: <http://www.students4bestevidence.net/1-2-anecdotes-are-unreliable-evidence/>  
Casey Ross, Ike Swetlitz: IBM pitched its Watson supercomputer as a revolution in cancer care. It's nowhere close, *STAT*, 5.9.2017,  
<http://www.statnews.com/2017/09/05/watson-ibm-cancer/>

---

## Article

13-Mar-2018

Walter Pytlik

© BIOPRO Baden-Württemberg GmbH

---

## Further information

Prof. Dr. Gerd Antes  
Co-Direktor  
Cochrane Germany  
Universitätsklinikum Freiburg  
E-mail: [antes@cochrane.de](mailto:antes@cochrane.de)

► [Cochrane](#)  
[Germany](#)

---

**The article is part of the following dossiers**



Evidence-based medicine – between magic formula and a doctor's strait jacket

---



Data mining: new opportunities for medicine and public health

---



Big data - the big promise of the new digitised world

bioinformatics

clinical  
trial

expert interview

telemedicine

big data