4 Knowledge does not equal evidence — what to do with what we have evidence for?

> "No evidence without patients! EBM starts with patients and ends with patients!" (Hywel Williams)[216]

4.1 Knowledge versus evidence — why the distinction is important

Evidence and knowledge are often used almost interchangeably in common language. To say that one has something on 'good evidence' to most people means to know something or at least to be pretty sure about the facts. 'Knowledge' in a philosophical context however deals more with the questions "What is knowledge?" and "How do we acquire knowledge?" The last question is the one which is most closely associated with the overarching question of EBM about how we can acquire evidence, and what makes the acquired evidence, "good" evidence? In the medical context it seems to be ambitious to claim to have knowledge, let alone absolute knowledge, since medical facts are changing at a rapid pace. Evidence about disease and their possible cures grows exponentially. It is important to understand how we are supposed to use this evidence, and why medical evidence and medical knowledge are distinct from, but dependent on each other. In one important paper about the topic by Silva and Wyer, titled: "Where is the wisdom?…"[217] the authors even go so far to claim that we need medical 'wisdom' because they ask the question: "how does knowledge lead to wise and just action?"[218] and are thereby encompassing with one question a huge part of the problem of medical ethics. So we have medical evidence, medical knowledge and medical wisdom. The question is how we understand each of these and their importance in the practice of modern medicine?

The term 'evidence' will mainly be questioned in the chapter. However, the full term is 'evidence-based medicine' and therefore the question if medicine can

216 Hywel Williams. (2015). "Reducing avoidable waste in eczema research." Conference talk at the EvidenceLive Conference in Oxford.

217 Suzanna A. Silva and Peter Wyer. (2009). "Where is the wisdom? II-Evidence-based medicine and the epistemological crisis in clinical medicine. Exposition and commentary on Djulbegovic, B. Gyatt, G.H. & Ashcroft, R.E. (2009). "Cancer Control." 16, in Journal of Evaluation in Clinical Practice:15: 158-168.

218 Suzanna A. Silva and Peter Wyer. (2009): 899.

M.-C. Schulte, *Evidence-Based Medicine – A Paradigm Ready To Be Challenged?*, https://doi.org/10.1007/978-3-476-05703-7_4

be *based* on evidence must be asked as well. Some authors, among them Ross Upshur, argue that we should not understand medicine to be *based* on evidence since that would mean a type of philosophical foundationalism which 'evidence' cannot uphold as such. Upshur interprets evidence as too rigid because it is only based on RCTs, leading to results that are not usable for the individual patient. Therefore it cannot, in his view, be a solid base for medical practice.[219] However, as we have seen, the EBM movement acknowledges the deficiency of its early approach and strives to make the evidence-base broad enough to make it applicable to all patients by including many types of research in medical practice.[220] I even go one step further and argue that medical research and medical practice should be separated to solve the problem of bringing the available evidence to the patient by using many more methods than just RCTs. So medicine can be based on evidence if the division between robust and statistical evidence for research on the one hand and robust but fluid evidence based non-randomised studies, expertise, tacit knowledge, values and patients wishes in medical practice is assumed. This means that the evidence base is rather broad, but in an endeavour like medicine, were literally all kinds and types of people need to be included, a broad evidence base can be the only solution. Too narrow a base, as in just allowing the most robust research evidence, reduces all patients to 'averages' for whom it is enough to use rigid guidelines. Personal care would be non-existent in such a scenario and that would be counter-intuitive to medicine being understood as the endeavour to heal individual patients.

The main focus of the chapter however will be on the distinction between evidence and knowledge in and for EBM and why this distinction is so important on the one hand, but can lead to danger on the other, if and when the best knowledge is not similar to the best evidence for the individual patient, i.e. does not lead to 'wise and just action.' Evidence hierarchies will be discussed, because they stand for a certain rigidity in the EBM approach, but also illustrate the flaws of that approach, especially when the overall hierarchy is deemed more important than the robustness of its different steps.

4.1.1 Possible definitions of 'evidence' and 'knowledge' for EBM

Even though it might be easy to accept that it is called *evidence-based* medicine and not knowledge-based medicine, this acceptance does not yet define the term 'evidence' or even gives a good explanation. The importance of separating medical research from medical practice also plays a role for the definition of 'evidence.'

219 Ross Upshur. (2002). "If not evidence, then what? Or does medicine really need a base?" in Journal of Evaluation in Clinical Practice. 8(2): 113-119.

220 Trisha Greenhalgh. et.al.(2014): 1-7.

Medical research produces 'evidence' while medical practice uses it. As has become clear in the previous chapter, solutions to the problem of external validity of research trials play a vital role in making the produced evidence usable for the individual patient. The attempt to define 'evidence' especially for medical practice will therefore look at the problem of external validity again, in the context of epistemology.

The definition of evidence most often used is "grounds for belief" or "good reason for belief".[221] For the sake of clarity, and although there are more definitions available in philosophy of science, I will use this definition as the basis for my argument. 'Knowledge' on the other hand is most often defined as 'justified true belief'.[222] [223] 'Knowledge' therefore contains a truth element which 'evidence' is lacking. This difference is the smallest common denominator on which most philosophers of science can agree upon.[224] And the difference will be significant for EBM and is already manifest in the name. It is not called 'knowledge-based medicine' since contrary to knowledge, evidence is changeable and gradable. Evidence is falsifiable, so there is no inherent truth element. Evidence is also under constant review and change, but based on carefully conducted and completely published research, it is possible to say that the available evidence at that point in time is the best one. Although one has to understand the above sentence in an idealised world. Throughout the dissertation it will become obvious that EBM is a long way away from achieving the ideal of always using the best evidence at any given time because of its many methodological problems.

Since evidence is understood as 'good reason for belief', the question that should be asked is, what 'good' reason actually means and what transforms 'evidence' into 'good' evidence? That precludes that evidence can be either 'good' or 'bad' and the question can be asked. The same question for knowledge would not make sense, since knowledge can ever only be incomplete. It cannot be 'good' or 'bad', neither can it be false since false knowledge would not be knowledge at all. Knowledge seems to me to be value-neutral. It has a truth factor and needs to be

221 Thomas Kelly. (2016). "Evidence" in The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.) https://plato.stanford.edu/archives/win2016/entries/evidence. Last accessed on January 23rd, 2020.

222 Jonathan Jenkins Ichikawa and Matthias Steup. (2017). "The Analysis of Knowledge", The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.) https://plato.stanford.edu/archives/spr2017/entries/knowledge-analysis. Both sources last accessed on January 23rd, 2020.

223 Knowledge as 'justified true belief' is itself a contested notion. Most prominent here is the Gettier problem. Since for my purpose, only the truth condition is of relevance, the definition as such can be used. However, it should be noted that there is disagreement in the philosophical community. Edmund Gettier. (1963). "Is Justified True Belief Knowledge?" in Analysis 23 (6):121-123.

224 Benjamin Djulbegovic, Gordon Gyatt and Richard Ashcroft. (2009): 160.

justified and there is also the possibility to gain and lose knowledge, but knowledge as such does not transfer value judgements. However, even if knowledge itself is presumed to be value-neutral, that does not mean that there are no values attached to knowledge. Values do enter into knowledge for example if and when we appraise the knowledge of a particular person. It would be possible to claim that a person has no knowledge about a particular topic and should give no recommendations accordingly.

It can be argued though, that a physician had bad evidence for a treatment decision. Ben Goldacre gives an example for such a case out of his own medical practice.[225] He had prescribed the antidepressant reboxetine to a patient, after he had consulted the relevant literature and learning that it was better than placebo and equally good to most other antidepressants. Goldacre, together with the patient, opted for this particular treatment. It turned out though, after a meta-analysis was conducted, that the treatment had quite significant side effects. The published data Goldacre has had access to at the time of prescribing the treatment, was based on only one trial which looked favourable but was not representative for all the accrued data concerning reboxetine.[226] This represents a case of publication bias. The evidence which Goldacre had based his decision on was bad, his knowledge concerning the particular treatment however had merely been incomplete.

"Being mistaken is not the same as being unreasonable. To the extent that one respects one's evidence, one is not unreasonable even when one is wrong."[227] Although the statement in and of itself is applicable, this 'being wrong' in medicine can have dangerous consequences. Therefore, the evidence on which medical decisions are based must be as 'good' as possible and it must be 'objective', i.e. not 'one's evidence,' but the best available evidence at the time. Hence, 'good reason for belief' as such is a necessary but not sufficient condition for clinical decisions.

What we do with the available knowledge and how we obtain the knowledge might be harmful, but it does not make the knowledge in itself wrong or bad - just its application. Examples in the medical domain, and ones which will be also of importance in the chapter about informed consent, are the medical experiments during the Third Reich in Germany. Many prisoners, both in prison and in the

225 Ben Goldacre. (2012): 7.

226 Dirk Eyding, Monika Lelgemann and Ulrich Grouven. (2010). "Reboxetine for acute treatment of major depression: systematic review and meta-analysis of published and unpublished placebo and selective serotonin reuptake inhibitor controlled trials." in BMJ; 341.

227 Thomas Kelly. (2014) "Evidence", The Stanford Encyclopedia of Philosophy, Edward N. Zalta (ed.), http://plato.stanford.edu/archives/fall2014/entries/evidence/. Last accessed on January 23rd, 2020.

concentration camps, were subjected to medical experiments.[228] Most prisoners that were 'chosen' for these experiments had some special feature, a disease that was interesting or a bodily feature which distinguished them. Some concentration camp prisoners even opted to be part of these experiments in the hope to survive longer because they were needed. Especially fiendish were the experiments conducted on children, one very notable example being the "twin study" conducted by Joseph Mengele, a student of Otmar von Verschuer who was one of the Reichs geneticists and huge beneficent of the concentration camp medical experiments.[229] The knowledge which these 'physicians' had obtained through their experiments is not in itself bad, most of it is medical knowledge which is still in use today, but the way in which it was obtained was intrinsically evil, because it reduced human beings to guinea pigs for whom it was acceptable to die if they had fulfilled their role. There were no ethical guidelines that controlled these experiments and no control to save those who were experimented on. The atrocities committed during the Second World War are stark reminders why it is so important today to ethically check and approve all experiments and to insist on informed consent by the patients to participate in medical research. In medical practice it is also important to accept that a patient might opt out of a treatment, even though it is deemed to be the best one for him or her. The patients consent, or lack thereof, should trump all other considerations.

Knowledge, even though it is not gradable and is notoriously hard to define, nevertheless plays a significant role in EBM,[230] as it is part of what makes the evidence usable. But knowledge cannot be generated or appraised quantitatively, and EBM is based on the quantitative generating of evidential facts and numbers. It is based on statistics using the population level. Thus, evidence as such does not include the individual knowledge of the physician, nor of the patient. So one goal of medical research must be to produce evidence which is robust and yields a good reason to belief that the treatment under test is better than placebo and/or has some advantage to an already established standard treatment. In research it does not matter that the evidence is not geared toward one particular patient. In practice, evidence also needs to be robust, but research evidence can merely be informing medical decisions. The evidence used for a particular patient must be more than just robust, it must be fluid enough to include the expertise of the physician and the values and wishes of the patient. It must also be fluid enough to incorporate a broad range of evidence, not only the results of the most rigorous tests, if they are

228 Ernst Klee. (2014). *Euthanasie im Dritten Reich. Die Vernichtung lebensunwerten Lebens.* Frankfurt: Fischer Taschenbuch Verlag:

229 William E. Seidelman. (1995). "Mengele Medicus: Medicine's Nazi Heritage." in International Journal of Health Services, 19(4): 599-610.

230 A fact that is also underscored by the myriad of literature about epistemology and evidence in EBM. Much of the literature however is based on the same premisses and therefore quite repetitive.

appropriate for the particular patient. The hierarchy of evidence that plays such a crucial role in medical research must be overcome in medical practice.

4.2 EBM as a new theory of epistemology in medicine?

Since the first JAMA paper from 1992, EBM has not only be called a 'new paradigm' but also, sometimes only implicitly, a new theory of epistemology. The idea seems to be that since EBM is so rigid in its production of 'robust' evidence, its methodology necessarily must be usable in other sciences and lead to a new way of not only arriving at knowledge, but also at defining knowledge. Djulbegovic and his colleagues, including two co-authors of the famous first EBM paper, however argue against this definition, claiming that "EBM enthusiastically draws on the major traditions of philosophical theories of scientific evidence. However, EBM does stress the importance of reliable, unbiased observation over theory."[231] To what the authors allude here is the debate in science if evidence can be neutral or if all observations are automatically theory-laden, following Popper here, since we would not be able to make sense of them otherwise.[232] A detailed discussion of the two sides would go beyond the scope of this work, but it is important to make clear that EBM favours neither the one view nor the other exclusively. In good scientific tradition, hypotheses, based on already accepted theories are a good starting point for research. The occurring results are necessarily then theory-laden. However, some treatments were and are used solely because they were observed to be successful, without looking for a valid theory which could underwrite the observation. Neither approach makes evidence in itself more robust or reliable.[233]

Djulbegovic and colleagues are arguing that "EBM makes a normative claim about when some kinds of medical knowledge can genuinely be taken as knowledge."[234] And they even argue that it is not only a theoretical normative claim, but also a practical one "It [EBM] also makes a normative claim about medical practice: Wherever possible, the choice of diagnostic test, preventive measure, or treatment should be based on the best available evidence about the available interventions." However, again EBM is not called knowledge-based medicine. By using the term 'evidence' in the first place it it is implied that the 'grounds for belief' which are assumed at the exact time the evidence is used, are subject to constant and continuous change. On the contrary to the authors claim, it would be far more prudent to accept that there is very little absolute knowledge in medicine,

231 Benjamin Djulbegovic, Gordon Guyatt and Richard Ashcroft. (2009). "Epistemologic Inquiries in Evidence-Based Medicine." in Cancer Control: Journal of the Moffitt Cancer Centre. 16(2): 164.
232 Karl Popper. (1992). *The Logic of Scientific Discovery.* London, New York: Routledge.
233 Benjamin Djulbegovic, Gordon Guyatt and Richard Ashcroft. (2009): 163.
234 Benjamin Dulbegovic, Gordon Guyatt and Richard Ashcroft. (2009): 159.

especially since this acceptance would lead to a constant questioning of the science of medicine, and hopefully to constant progress in its practice. Silva and Wyer argue in a response to this paper that "the issue posed by EBM is not the 'relationship between theory, evidence and knowledge' but rather the relationship between theory and practice, which means the relationship between 'what we know' (knowledge) and 'what we do with what we know' (wisdom)."[235]

It seems to me, and I agree with Silva and Wyer here, that the question about what knowledge actually is, is much less important than the question, what to do with the medical evidence we have, in practice. Since I claim that the term 'medical knowledge' is contestable, I also contest to/the? use of 'wisdom' in medicine, but would rather use 'clinical experience' which informs clinical practice on a daily basis and should inform medical research by asking the right questions to guide research along.

In medical practice it seems that Silvas and Wyers questions are still important, but, I argue, would need to be reformulated into 'what do we have evidence for?' and 'what we do with the available evidence?'

Even though I aim at a different terminology, Silva and Wyer formulate it best in their paper and therefore I will use the entire quote:

> "Rather the first epistemological challenge, forced by the 1992 proposal, is how inferences regarding the likely ranges of true average effects and frequencies across study populations can and should impact upon the process of delivering health care to individuals….Hence the 'evidence' stemming from clinical research, although *direct* with respect to the task of predicting population effects and outcomes, and perhaps with respect to evaluation of practice patterns of individual or groups of clinicians, is necessarily *indirect* evidence with respect to the decisions, actions and general clinical care of an individual patient."[236] [my emphasis].

So again, what makes evidence 'good' evidence and for whom is it applicable when?

A new theory of epistemology would have 'knowledge' at its very centre. EBM has evidence at its core and not knowledge, hence can we talk about EBM as being a new theory of epistemology? EBM makes use of theories of evidence and also of theories of knowledge and transfers them to a practical setting. It is attempting to use it specifically for the individual patient but in any case producing evidence which is at the very least directly usable on the population level and in lesser form on the individual level. But EBM does not give rise to a new and unique theory of epistemology, since because EBM is unique as being neither a pure science nor a pure art and therefore most of the methodological theories developed for and within EBM cannot be successfully used in other sciences. A new

235 Suzanna A. Silva and Peter Wyer. (2009): 900.
236 Suzanna A. Silva and Peter Wyer. (2009): 901.

theory of epistemology however should be transferable and usable in other sciences as well. A valid critique to the last argument is that it would not matter if it would be a distinct 'new' theory of epistemology for EBM. However, since medicine and especially EBM are drawing on so many other natural sciences, as in biology, chemistry and physics, it seems advisable to agree on a theory of epistemology which then holds for all, since it would be easier to use the aforementioned fields of science in conjunction with medicine, sparing one translational step on the way.

4.3 Evidence hierarchies

EBM uses a system of evidence hierarchies to show which forms of evidence are methodological superior to others. Evidence hierarchies are everywhere, and there are many different hierarchies in published literature about EBM, but they only portray an idealised version of quantitative evidence, not its usability. In the following sections I will sketch a typical EBM hierarchy, to illustrate why they are useful for generating evidence, but to also illustrate why they are not useful for the individual patient.

Evidence hierarchies most often have the highest ranking form of evidence on top and the lowest on the bottom. Meta-analyses (the statistical aggregation which produces a single effect size) and systematic reviews (the process of selecting the studies)[237] are usually on the very top, followed by RCTs. Those are followed by cohort studies, case control studies, case series and at the very bottom, expert opinion and mechanistic reasoning and causation. The usual diagram for this hierarchy is the pyramid. [238]

237 M. Hassan Murad. et.al. (2016). "New evidence pyramid." in ebmed. Journal for evidence-based medicine. 21(4): 125.

238 Bob Phillips. (2014). "The crumbling of the pyramid of evidence" in BMJ Blogs. http://blogs.bmj.com/adc/2014/11/03/the-crumbling-of-the-pyramid-of-evidence/. Last accessed on January 23rd, 2020.

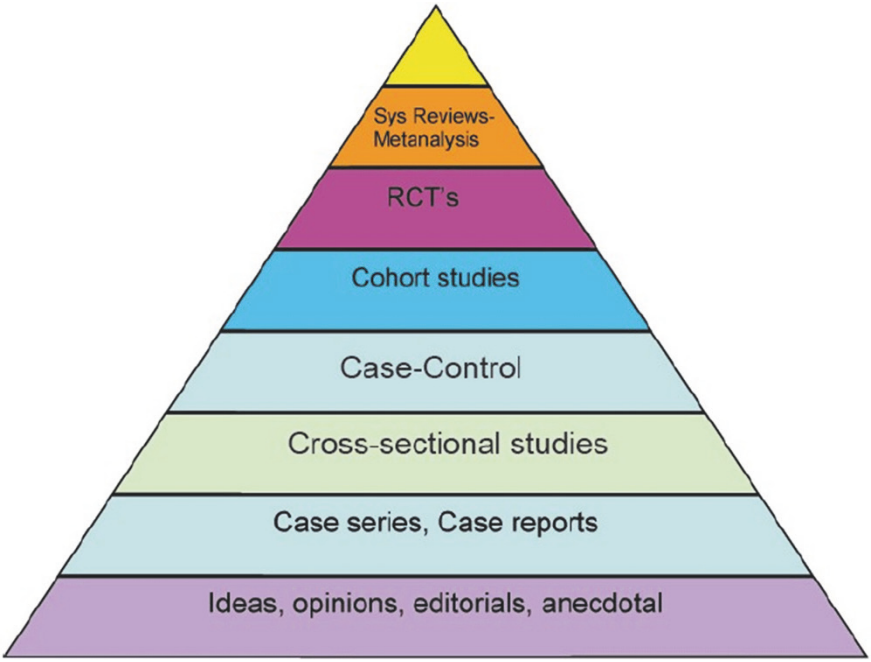

Figure 2: Standardised pyramid of evidence

Source: https://blogs.bmj.com/adc/2014/11/03/the-crumbling-of-the-pyramid-of-evidence/. Last accessed on November 14th, 2019.

This is one of the hierarchies which was for the longest time favoured by most EBM proponents. Almost all hierarchies look the same at the top, but can differ on the bottom. Some include on the very bottom, right next to "expert opinion", "laboratory and animal research", some 'mechanistic reasoning', some dissect the different types of observational studies and rank them according to perceived robustness. Therefore, the pyramid form is actually slightly misleading, since everything below RCTs is often clustered into non-robust evidence and/or at the very least to be of much lesser value than RCTs. Authors of some early papers explaining EBM even went so far as to advise their readers to stop reading medical papers, if their results were based on anything other then RCTs.[239] Since a couple of years however, it seems to be understood and accepted, that other forms of evidence, such as cohort and case control studies can be just as good, if they are as

239 Gordon Gyatt, et.al. (1992). "Evidence Based Medicine: A New Approach to Teaching the Practice of Medicine." in JAMA, 268 (17): 2420 - 2425.

well conducted as a RCT, and that badly conducted RCTs only provide 'bad' results and therefore 'bad' evidence.[240]

> "Although it is common to talk about "the" hierarchy of evidence, there are actually multiple hierarchies…For example, some hierarchies explicitly say that RCTs included in a meta-analysis must have similar characteristics (e.g. medication dosages, inclusion and exclusion criteria), and some subdivide the level of "observational" studies into cohort and case-control designs."[241]

The above might be a minor point, however it shows how much has changed from what the fathers of EBM originally wanted and how it is used and understood today. Sackett understood the hierarchy of evidence as a tool to compare and assess evidence and to come to a consensus. But it seems already to be too complicated to arrive at a consensus about which hierarchies to use. It seems that most hierarchies agree that meta analyses and RCTs belong somewhere at the top of the pyramid, while clinical expertise is either relegated to the bottom, or taken out completely. "In 2002, the AHRQ [Agency for healthcare research and quality][242] reported 40 systems of rating in use, six of them within its own network of evidence-based practice centers….The GRADE Working Group,[243] established in 2000, is attempting to reach consensus on one system of rating the quality and strength of the evidence. This is an ironic development, given that evidence-based medicine sees itself as replacing expert group consensus judgement."[244] Miriam Solomon here makes a reference to the method that was used before EBM, the so called consensus conferences in which experts tried to arrive at a consensus about treatments based on their experience. However, many of these consensus conferences stalled when every expert had explained his or her method. Hence, something like EBM had to happen to push medical science forward.[245]

4.3.1 Systematic reviews, meta-analyses and RCTs

On top of most hierarchies are systematic reviews and meta-analyses. In conjunction, and on their own to a lesser degree, these are considered to be the ultimate

240 Jeremy Howick, et.al. (2009). "The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation' contribute? in Journal of the Royal Society of Medicine. (102): 186.

241 Robyn Bluhm and Kirstin Borgerson. (2011). "Evidence Based Medicine." in *Handbook of the Philosophy of Science. Vol. 16. Philosophy of Medicine.* Ed: Fred Gifford. Oxford: North Holland, Elsevier: 210.

242 AHRQ Agency for healthcare research and quality. https://www.ahrq.gov. Last accessed on January 23rd, 2020.

243 The GRADE working group. From evidence to recommendations – transparent and sensible. http://www.gradeworkinggroup.org/. Last accessed on January 23rd, 2020.

244 Miriam Solomon. (2015). *Making Medical Knowledge.* Oxford: Oxford University Press:113.

245 Robyn Bluhm and Kirstin Borgerson. (2011): 210.

solution for the accumulation and overall analysis of all the available evidence. However, "different meta-analyses of the same evidence can reach contradictory conclusions….A frequent goal of using meta-analysis is to discover causal relationships and to determine the magnitude of an effect for a particular magnitude of a purported cause."[246] Jacob Stegenga continues by arguing that if RCTs are supposed to be the 'gold standard' in EBM, meta-analyses are claiming the title 'platinum standard' for EBM. Stegenga argues against this approach and I concur with him. RCTs per se are not the gold standard, as we have seen in the previous chapter, and nor are meta-analyses the platinum standard per se. Since meta-analyses are using the results of RCT's, their results are also only based on a population level average and are again not viable for the individual patient. Publication bias also plays a role here. If not all data about a trial is published and not made available to those researchers conducting the meta-analysis, the results of the analysis can be as flawed as the results of the original RCTs.

RCT's are almost purely about effectiveness on the population level. "They are not designed to discover how health care interventions work (when they do work), or to come up with new ideas about mechanisms, new theories about disease processes, or new technologies for medical interventions."[247] Solomon continues by criticising that even RCTs with known methodological flaws are ranked higher than a high-quality observational trial. Because of these 'flaws' evidence hierarchies can be rendered unreliable.

The CONSORT statement[248] and the GRADE Working Group[249] are focused on standardising evidence hierarchies and make them more reliable and even include variation. However, these organisations have the same problems as described above. Hierarchies are rigid by their very nature and it seems to be almost impossible to make them reliable on the one hand, and fluid on the other, all at the same time and having to work with the same facts but interpreting and using them differently.

Trials sponsored by Big Pharma should be automatically ranked lower on the evidence hierarchy, according to Miriam Solomon. RCTs sponsored and/or conducted by pharmaceutical companies, because of funding and publication bias, have consistently more favourable results than those from comparable but independent trials. Solomon's solution to that problem is that those trials are supposed

246 Jacob Stegenga. (2011). "Is Meta-Analysis the Platinum Standard of Evidence?" in Studies in History and Philosophy of Science Part C 42 (4): 497.

247 Miriam Solomon. (2015):117.

248 David Moher, Kenneth F. Schulz and Douglas Altman. (2001). "The CONSORT statement: revised recommendations for improving the quality of reports of parallel group randomised trials." in BMC Medical Research Methodology. I:2.

249 GRADE working group. gradeworkinggroup.org. Last accessed on January 23rd, 2020.

to be ranked much lower on the hierarchy and that they should only be reconsidered if their reliability is improving. One problem here seems to be that, by now, many more trials are sponsored by Big Pharma than are conducted independently. And even if they are 'independent' it still renders the question, 'independent' from what or whom? University researchers also have an interst in publishing positive results. I tend to concur with Solomon that trials done by Big Pharma are more prone to bias, but it is not enough to simply push them down the hierarchy. A solution to the problem should be found already when RCTs are initiated by pharmaceutical companies. These companies have a necessary interest in RCTs when their products are under test and they have an interest in positive outcomes. Often negative or questionable results are still not published or made available to independent researchers. A possible solution to prevent industrial bias would be to make it mandatory to outsource the trials to independent clinics in which they can be performed. But even if they are conducted in-house, there are possibilities to establish a type of self-control of the companies, especially since pharmaceutical companies do not want to lose their trustworthiness. In the United States, Jennifer Miller, professor at Harvard, has established the Good-Pharma-Scorecard on which pharmaceutical companies are ranked according to their successfulness in publishing all relevant trial data.

"Our Good Pharma Scorecard (GPS) ranks large pharma companies and every new FDA approved drug on key ethics, human rights, and public health criteria. We focus on 5 areas:[250]

250 Bioethics International. The Good Pharma Scorecard. http://bioethicsinternational.org/good-pharma-scorecard-overview/. Last accessed on January 23rd, 2020.

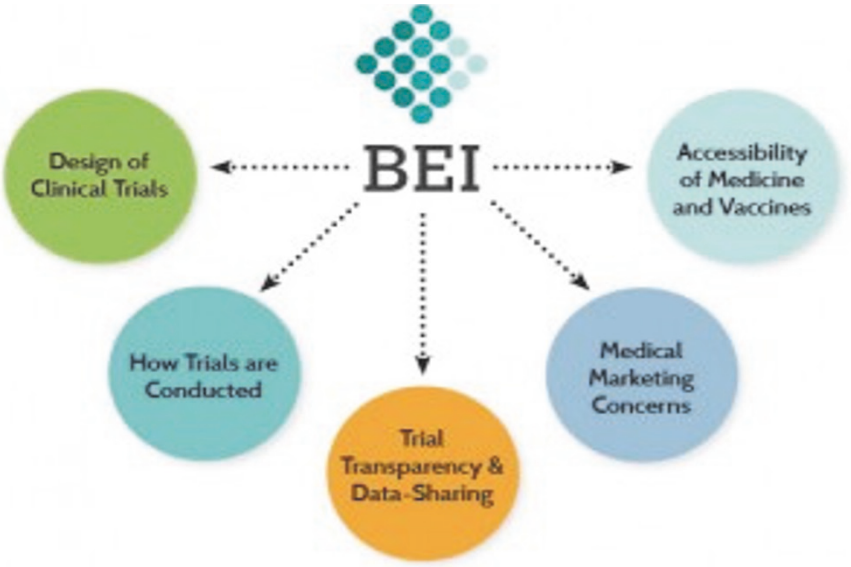

Figure 3: What the Good Pharma Scorecard wants to achieve and their points of access

Soure: Bioethics International. The Good Pharma Scorecard. http://bioethicsinternational.org/good-pharma-scorecard-overview/. Last accessed on September 25th, 2017.

The AllTrials campaign initiated in the UK also tries to persuade all pharmaceutical companies to register their trials and to disclose all trial data.[251] Full disclosure would lead to the possibility to question the trial and maybe to a form of 'obligation' to produce 'good' data, meaning honest data, from the very beginning. However, all of these are voluntary measures. Neither pharmaceutical companies nor individual researchers have a legal obligation to publish trial data. They can only be ethically held accountable for their work.

Systematic reviews and especially meta-analyses can only be as good as the data they are working with. Therefore, if the data accrued by RCTs or other studies is flawed, so are meta-analyses. In themselves therefore meta-analyses cannot solve the problem of making evidence usable for the individual patient. They can and do help to make the available evidence more manageable and they can dismiss evidence that is obviously flawed, but they are powerless against hidden flaws. So again, the solution does not lie in more methods of appraising the same data, the solution lies in producing better data to begin with.

251 AllTRials campaign. http://www.alltrials.net/. Last accessed on January 23rd, 2020.

4.3.2 Observational studies

One step below RCTs, but included in most hierarchies are observational studies.[252] Observational studies do not randomise their participants but only observe them over time, without actively testing for specific results. Observational studies can have the positive effect that they are fairly easy to perform, can be longitudinal and can include many patients with minimal costs. And in some cases, observational studies are the ones that lead to new research questions and a new focus on a certain disease, making them imperative for medical progress. In order to successfully utilise the results of these studies, STROBE has been developed. STROBE stands for "Strengthening the Reporting of Observational Studies in Epidemiology" recommendations.[253] Since observational studies are ranking so low on the evidence hierarchies, they are especially prone to publication bias. STROBE wants to correct for possible publication bias for all forms of observational studies so as to make the results robust and reliable and to really inform future medical research.[254]

Even though observational studies are slowly gaining more importance, they are not without problems. Because they are not using randomisation, possible confounders can lead to a misrepresentation of the accumulated data. Additionally, it is very complicated or even impossible to conduct observational studies in a blinded setting. However, as seen before, 'blinding' is the only method to prevent selection bias and observer bias. Therefore observational studies are prone to suffer from both of these biases. Blinding in observational studies is only possible if and when the 'to be collected data' is either comprised of a laboratory test or of a radiograph. Direct patient observation is impossible to blind. This fact alone renders observational studies far less robust in the eyes of strict EBM adherers. A well conducted and open observational study however can be more robust, even without randomisation and blinding, than a sloppily conducted RCT. Most observational studies also differ from RCTs because they do not look at novel treatments or drugs but on disease progression over time, given certain parameter, as in treatment versus no treatment, general health, regression to the mean of illnesses and

252 W. Yang, A. Zilov, P. Soewondo, O.M. Bech, F. Sekkal, P.D. Home. (2010). "Observational studies: going beyond the boundaries of randomized controlled trials." in Diabetes Res Clin Pract. 88 Suppl 1:S3-9.

253 STROBE Statement: Strengthening the reporting of observational studies in epidemiology. https://www.strobe-statement.org/index.php?id=strobe-group0. Last accessed on January 23rd, 2020.

254 Jan P Vandenbroucke, Erik von Elm and Douglas G Altman, et.al. (2007). "Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration" in PLoS Medicine.

patients quality of life. Observational studies in EBM are divided into cohort studies and case control studies and both can, when done correctly, yield robust results.[255]

Cohort studies are purely observational studies which focus exclusively on the causes of disease. Cohorts are groups of patients, who are observed over time. These groups can be compared among each other and chosen in a way as to have a cohort with a certain disease, medication or health problem and one cohort without. However, these cohorts are not randomised but are sorted purely by the existence of the above factors. An example for a cohort study is the Nurses' Health Study,[256] a long running observation of women's health in general which started in 1976 and was renewed in 1989. At the time of writing this chapter in 2017, the NHS is recruiting for a third phase of the study which is already running since 2010. Because of its longitude and the large number of participants, the Nurses' Health Study is an excellent example of a robust cohort study yielding very robust results which should be ranked above many RCTs about the same topics, because of these features.

Another form of observational studies are case-control studies. These studies are retrospectively comparing patients, separated in two groups, one with the disease in question, one without. Retrospectively means that the patients are 'observed' after a certain outcome has already occurred. The difficulty of case-control studies is that most people do not reliably remember their symptoms over time. Equally they might have forgotten if they had taken all the necessary medication all the time or if they had lapsed in the intake or when these possible lapses might have occured. Data can get lost and not everyone might follow-up. But case-control studies do have the advantage that a large number of patients can be recruited into them and that they can be conducted over a lengthy period of time. And they are fairly quick and painless to perform.[257]

Not all observational studies are longitudinal though. Sometimes, to achieve a kind of 'snapshot' of a certain symptom or to study prevalence, cross-sectional studies are used. These are in most cases not usable to establish causal connections, but are quick and easy to perform and multiple outcomes can be studied. 'Cross-sectional' means, that for example four groups can be compared over a very short time frame and the results are then collated. Four groups are often used because it is then possible to compare for three variables. For example age and cholesterol and how and if exercise can make a difference. In this scenario it would be possible to create four groups, two in each age range, one with high cholesterol levels, one

255 Jae W. Song and Kevin Chung. (2010). "Observational studies: Cohort and Case Control Studies." in Plastic and Reconstructive Surgery. 126(6): 2234.

256 Nurses' Health Study. http://www.nurseshealthstudy.org/. Last accessed on January 23rd, 2020.

257 Jae W. Song and Kevin Chung. (2010): 2238.

with normal cholesterol levels. Participants in both groups are then to perform light exercise.[258] After a short amount of time, the results are collated and a 'snapshot' is created if there are any short term inferences to be had.

Right below these observational studies are, in most hierarchies, case series and case reports. These are usually called 'observational' as well, but they are not scientific studies but descriptive reports about groups of patients, in case series; or a single patient, as in a case report.[259] These are only used if patients showed any unusual or diverting symptoms from the usual disease progression. Because of their purely observational status and the lack of a comparison, as for example in cohort studies, these case observations can be more prone to bias and can by their very nature not be as robust as is desirable for statistical evidence. However, they are important because they can lead to new research questions, since they are almost exclusively conducted when an anomaly occurred.[260]

4.3.3 Expert judgement, clinical judgement, clinical expertise

The lowest rank of almost all evidence hierarchies is occupied by 'expert judgement', 'clinical judgement', or 'clinical expertise'. For some reason, these vital skills in medicine are ranked fairly poorly. One reason might be that they are so called soft skills. They are not quantifiable and no numerical or statistical value can be attached. Additionally it might be that because of these being soft skills, they are prone to biases and faults. Humans make mistakes and so do experts. Soft skills are in themselves not evidence, but they are necessary to assess evidence and to ask the right questions. Therefore these soft skills are needed on every step of the evidence hierarchy. Without experts asking the right questions and performing the necessary trials there would be no evidence to begin with. So they should not rank the lowest on the evidence producing hierarchy, but should be outside of it, informing all ways of producing evidence.

And as soon as evidence is to be used for the individual patient, these skills are of vital importance. "The view that experts have special access to knowledge goes back to Plato. In medicine this view has been particularly influential: experienced clinicians are often believed to possess tacit knowledge and intuition that cannot be reduced to mechanical rules."[261] Junior doctors, next to their studies,

258 Institute for Work and Health. At Work: Issue 55, (2009). "Cross-sectional studies." www.iwh.on.ca. Last accessed on 15.07.2015.

259 Luca Ansaloni, Fausto Catena, and Ernest E Moore. (2007). "WJES and case reports/case series." in World Journal of Emergency Surgery. 2: 11.

260 Jan Vandenbroucke. (2001). "In defense of case reports and case series." in Annals of Internal Medicine. 134(4): 330-334.

261 Jeremy Howick (2011): 158.

have to work under expert supervision in clinics, to learn the skills that are necessary for the practice of medicine. Howick calls them apprentices. He goes on to argue that already the fact that this type of learning is part of their schooling proves that knowledge transfer from those with more experience to those with less experience is understood to play a vital role in the teaching and learning of medicine.

It can seem as if the EBM community has forgotten about the importance of these soft skills for medical practice. It is not enough to find the appropriate evidence, the physician also has to question it, to see if it does fit the individual patient. Critical thinking as a skill in medicine should have become more, instead of less important, within EBM. In order however to be able to critically question the available evidence, physicians must have a thorough knowledge; and here the term 'knowledge' is appropriate because beyond its intrinsic meaning it also stands for a vital soft skill without which the physician would not be able to even do his job in a meaningful way; knowledge about disease, the human body and diagnostics. The physician needs experience, to be able to question the evidence. And knowledge is what fuels experience. Guyatt, et. al. in the original EBM paper claim that the junior physician does not need that experience, and that it is enough to look up and understand the available evidence and to use it in the individual case. However, since the initial paper from 1992, the medical database has grown exponentially. Every physician could spend multiple weeks if not months or years reading through the literature of one single diagnosis, and by the time he would be done, a lot of the information would be already outdated. To illustrate this point, here are a few numbers.

> "More than 15 million medical papers have been published.
>
> The number of medical journals is in excess of 5000.
>
> It has been estimated that only some 10-15% of what is published today will be of lasting scientific value. It has been estimated that half of today's medical knowledge base will be out-of-date, erroneous or irrelevant in 10 years."[262]

Again knowledge and expertise are needed to filter all the available information, look for the best evidence in the circumstances and to do all this in a timely manner. Physicians therefore often have to use a short-cut or heuristic approach in decision-making, based on the evidence presented in a particular case, but not necessarily based on the best available evidence, since that might not be known to the clinician at that point in time and there is no time to search for it. Examples for this are emergency situations, in which a patient needs to be treated in a very short time frame. Since the clinician does not have the time to critically reflect his decisions, his "base" for the decisions cannot be "theoretical evidence", nor can it be

262 Jorgen Nordenstrom. (2007): Introduction.

"tacit knowledge" on its own. A definition by Milos Jenicek says: "clinical expertise is an amalgam of several things: there must be a solid knowledge base, some considerable clinical experience, and an ability to think, reason and decide in a competent and well-calibrated fashion."[263] He should have added 'quick' to his list, because in most clinical or private practise settings, the physicians are pressed for time and in a way have to 'think on their feet' in order to get to all patients or deal competently with emergency situations.

Jenicek's definition of clinical expertise can be reformulated to incorporate the EBM language. Clinical expertise then should include: a solid evidence-base, tacit clinical knowledge, research-based clinical knowledge, and the ability to apply these different forms of knowledge and the available evidence in a short amount of time, focusing on the individual patient or situation.

If clinical expertise does fulfil these criteria, then it is not only the starting point for the actual treatment of the individual patient, but also the be-all and end-all of EBMs two sides, namely research and practice, because this form of expertise is needed for both.

Another reason why clinical expertise is not in such high regard seems to be that clinicians themselves seem to underestimate their abilities to quickly absorb and incorporate new evidence and to overestimate what they already know and do and perceive as successful. Before EBM, and still used today, are consensus conferences and Trisha Greenhalgh calls these the GOBSAT (Good Old Boys Sat Around a Table) method.[264] And at these conferences is it were most of the overestimation of the single clinician's expertise takes place. Greenhalgh's GOBSAT method stands for the inherent problems of clinical expertise, and not only at these or other medical conferences, but also in the hallways of clinics and doctors offices. Experts are human and therefore seldom perfect. One single clinician very often does not convincingly know if an observed effect is based on a drug, on a placebo effect or on the resilience of the human body. In order to find that out, drugs need to be tested.

Diagnostic skills are also soft skills, but they are not impossible to quantify. Given the right information, computers can do a lot of diagnostic work. Howick writes about examples where computers were on average more accurate in their diagnosis than clinical experts, in those cases where a computer based formula was available.[265] However, that does not stretch as far as the computer being able to prescribe the right treatment in case of multi-morbidities and to then dispense that

263 Milos Jenicek, Pat Croskerry, David Hitchcock. (2011). "Evidence and its uses in health care and research: The role of critical thinking." in Medical Science Monitor: 13.

264 Trisha Greenhalgh. (2014). *How to read a paper: The Basics of Evidence-Based Medicine. Fifth Edition.* London: Wiley BMJ Books: 7.

265 Jeremy Howick (2011): 169.

treatment in a compassionate fashion. But it does show that expertise can be questioned and should be questioned and that if experts agree to such a type of scrutiny, their individual results would probably be that much more reliable.

Howick argues that clinical judgement/expertise should not be used as evidence, and I agree with him on this point. Clinical judgement and expertise can lead to the right evidence by asking the right questions, in research and in clinical practice, but in and of itself it should not be regarded as evidence, but as a part of clinical knowledge. Howick reformulates the "description of EBM from 'EBM requires the integration of the best research evidence with our clinical expertise and our patient's unique values and circumstances' to 'EBM requires clinical expertise to integrate the best research evidence with patient values and circumstances.'"[266]

David Sackett, the father of EBM, solved the 'problem' of being an expert in his own way by stopping to write and lecture about EBM. In "The Sins of Expertness and a proposal for redemption"[267] he writes "…experts…commit two sins that retard the advance of science…Firstly, adding our prestige to our opinions gives the latter far greater persuasive power than they deserve…The second sin…is committed on grant applications and manuscripts that challenge current expert consensus…in 1983 I wrote a paper calling for the compulsory retirement of experts and never again lectured, wrote, or refereed anything to do with compliance."

Dave Sackett does not talk or lecture about EBM since 2000. He believes that it would hinder progress in medicine, and especially in EBM if he and other 'experts' would go on talking about their expertise. He claims that it makes much more sense to refocus ones career when a certain level of expert knowledge is reached in order to make way for new ideas in the field of ones own expertise and to develop new ideas in a new field. If all experts would follow Sackett's advice, then GOBSAT would not be a problem anymore, because experts would stop being experts as soon as a new research question is asked. It might sound like a trivial point in the grander scheme of things regarding EBM, but experts are much more important in medicine, research and practice than EBM allows, but they are less important than they sometimes themselves seem to believe, by priding themselves on their own expertise.

As a German comedian and physician, Dr. Eckhart v. Hirschhausen has said, physicians only get feedback from those patients who return, they never hear from those that stay away. They would however, learn much more from the latter group.[268]

266 Jeremy Howick (2011):177.

267 David Sackett. (2000). "The Sins of Expertness and a proposal for redemption." in BMJ; 320:1283.

268 Eckhart von Hirschhausen. (2008). *Die Leber wächst mit ihren Aufgaben: Komisches aus der Medizin.* Berlin: Rowohlt Verlag.

4.3.4 Mechanism and causation

Next to expert judgement, or expertise, mechanistic reasoning is also relegated to the bottom of the hierarchy. Mechanistic reasoning tries to establish if there is a mechanism linking a putative cause to a putative effect, or if a correlation of two facts was simply due to possible confounders.[269] It is on the lowest rank of the evidence hierarchy, because mechanisms are difficult to establish beyond a doubt. As we will see, there are examples in medical history where the mechanistic reasoning did function as usable evidence, and there will be examples where it is not clear 'why' some treatment works, it is just clear that it does and that is enough reason to use it. One mechanism in medicine which is fairly well understood is how oral medication reaches its target in the body. The process is called ADME (mechanisms for absorption, distribution, metabolism, and excretion.)[270] This overall mechanism is regularly used for medical research, but it does not in itself constitute medical evidence. Therefore it is not part of the evidence hierarchy, especially on the rank of mechanistic reasoning, but an important part of the methodology of medical research. And it is part of the chain of mechanistic reasons that can lead to patient-relevant outcomes.

High quality mechanistic reasoning in medicine would mean that the entire mechanistic chain of reasoning is known. Howick defines mechanistic reasoning as such, and claims further that it is imperative not only to know the actual mechanism, but to also understand how that mechanism, and every link in the chain will change due to treatment.[271] Since most mechanisms in the body are fairly complex and so are the changes due to treatment, mechanistic reasoning is questionable as a confident source of evidence in most cases. However, there are examples where mechanisms could be proved and used to advantage. And there are cases in which statistical evidence was not enough to convince the medical community of a treatment before the mechanism behind it was not known. A well-known example for the latter case is the Semmelweiss hypothesis that puerperal fever can be reduced by increased hygiene on the part of the physicians, especially hand washing. The method was only fully adopted after the death of Ignaz Semmelweiss and although he had shown through extensive statistics that his method worked. Only after the germ theory of disease was accepted did the Semmelweiss hypothesis take hold in clinical practice.[272]

A classic example for a working mechanistic reasoning is Robert Koch's effort to prevent future cholera outbreaks. This effort was stimulated by a serious

269 Brendan Clarke, Jon Williamson, Federica Russo, et.al. (2014). "Mechanisms and the Evidence Hierarchy." in Topoi. 33(2): 339-360.
270 Jeremy Howick. (2011): 127.
271 Jeremy Howick. (2011): 130.
272 Jeremy Howick. (2011): 130.

outbreak of cholera in Hamburg in 1892. Hamburg has a neighbouring city, Altona, further down the river Elbe, but curiously Altona was nearly free of cholera. What made this more surprising was that Hamburg's sewage was carried down the Elbe to Altona. Altona however, because of the sewage problem, already used slow sand filters to filter its water supply, long before the cholera outbreak. Hamburg did not filter its water. This evidence of correlation strongly suggested that slow sand filtration prevented cholera. However, this conclusion was not generally accepted and was, in particular, rejected by Koch's opponent Max Joseph von Pettenkofer.[273]

Koch had isolated the cholera vibrio in 1884, and suggested that it was the cause of cholera. Using this hypothesis, he now proposed a mechanism, namely that slow sand filtration removed the cholera vibrio. This mechanism could be tested out by bacterial counts before and after slow sand filtration. The results strongly confirmed the correctness of Koch's mechanism. When this evidence of mechanism was added to the earlier evidence of correlation, Koch's view became generally accepted, and was adopted by the German government in its efforts to prevent further cholera outbreaks.[274]

The above example includes evidence of correlation and evidence of mechanism, which are both necessary to make it a valid claim according to the Russo-Williamson Thesis (RWT). "In order to establish that A is a cause of B in medicine one normally needs to establish two things. First, that A and B are suitably correlated—typically, that A and B are probabilistically dependent, conditional on B's other known causes. Second, that there is some underlying mechanism linking A and B that can account for the difference that A makes to B."[275]

Causation and mechanistic reasoning are not two different kinds of evidence, just two different ways of looking at the evidence. It seems as if in common medical practice, correlation is higher regarded than mechanisms, although correlations are often more spurious. The mechanism on the other hand, if correctly understood, is that which gives ultimate proof, since a known mechanism gives absolute reason to believe something, i.e. 'good' evidence. Therefore it should be ranked higher, but it is much harder to come by, because mechanisms must be proven beyond a doubt. The quality of the mechanistic reasoning must be high in order to qualify. And most mechanisms are not easy-to-go one-step accounts, but are dependent on a chain of evidence linking the different mechanistic steps. "Each link in the inferential chain should be based on sufficiently strong evidence, perhaps (but not necessarily) from high-quality comparative clinical studies." [276]

273 Stephen B. Turner. (2013). *The politics of expertise.* Oxford: Routledge: 131.
274 Stephen B. Turner. (2013): 133.
275 Brendan Clarke, Donald Gillies et.al. (2013). "Mechanisms and the Evidence Hierarchy." in Topoi: 5.
276 Jeremy Howick. (2011): 144.

Even if a mechanistic link between the cause, as in giving a treatment, and effect, as in change in symptoms, can be detected, this link does not need to be the same for every patient. Treatments, especially drugs, can have quite massive side-effects, both negative and positive.[277] Not every patient experiences these side-effects and not every patient benefits from the drug in question. The causes for these idiosyncrasies can be many, but again it means that a causal link or a known mechanism might not be enough evidence to render a drug or treatment beneficial for the individual patient.

Since mechanistic reasoning is strongly linked to causality, it should be accepted that for mechanistic inferential chains the causal law of a cause preceding an effect has to hold. A curious case of correlation where the cause did not precede the effect is the Leibovici trial. Leibovici initiated a trial about "remote retroactive intercessionary prayer" for patients who were already discharged from the hospital. The patients were divided into two groups, one was prayed for, one was not. The trial results showed that patients who were prayed for, and I stress here, retroactively, left the hospital earlier than the patients in the control group. The absolute results however, were, as expected, statistically insignificant and no causal connection could be established.[278]

Some authors, such as Goldacre, also use the example of homeopathy as a spurious correlation. Homeopathy is, time and time again, under test to establish if an underlying mechanism can be found. So far, only spurious correlations have been detected. And these spurious correlations are, most of the time, based on anecdotes. The 'normal' progression of an illness is a slow to quick ascend, peak, and then sometimes a rapid decent which would have happened with or without medication. So the reasoning used by many patients is that whatever you did while your symptoms were at their worst, is what made them disappear. The fallacy behind that is called 'post hoc ergo propter hoc' fallacy, meaning 'after this, therefore because of this'.[279]

Since homeopathic remedies however do not contain any active ingredients, it is impossible to find a working mechanism. Goldacre however, again using homeopathy, provides us with an argument against putting too much weight on mechanisms. "We should remember, though, that the improbability of homeopath's claims for how their pills might work remains fairly inconsequential, and is not central to our main observation, which is that they work no better than placebo. We do not know how general anaesthetics work, but we know that they do work,

277 Jeremy Howick. (2011): 141.

278 Leonard Leibovici, (2001). "Effects of remote, retroactive intercessory prayer on outcomes in patients with bloodstream infection: randomised controlled trial." in BMJ; 323(7327): 1450–1451.

279 John Woods and Douglas Walton. (1977). "Post Hoc, Ergo Propter Hoc." in The Review of Metaphysics. 30(4): 569-593.

and we use them despite our ignorance of the mechanisms."[280] The latter is a very valid and most important point. Mechanistic reasoning can only replace other types of evidence when its quality is exceedingly high and for some treatments knowing the underlying mechanism would not make a difference in their use. Again, evidence hierarchies are too static to accommodate these differences. For the evidence user it is crucial to have enough experience and 'knowledge' to use the available evidence and to distinguish the quality of the different methods with which it was obtained. And even high quality mechanistic reasoning might not be applicable to the individual patient. Howick argues that, next to assume that the quality is sufficiently high of the mechanistic reasoning that "second we must assume that the mechanisms operating in the study population operate in the same way as the mechanisms operating in the individual who presents him or herself to the practice."[281] If these assumptions are taken for granted, then mechanistic reasoning can be part of every step of the evidence hierarchy, when it is established in high enough quality. And if it is of low quality it does not belong on the hierarchy at all, because it than can neither inform research nor be helpful for the individual patient in medical practice.

4.4 Problems with hierarchies and possible solutions

The most significant problem concerning evidence hierarchies is that they are perfectly suited for the production of evidence, but not very well suited to the actual use of evidence. I agree with Robyn Bluhm here who says that the "term hierarchy of evidence is a misnomer: the hierarchy is actually a hierarchy of methodologies."[282]

RCTs especially, but also all other studies in which a large number of patients is involved, provide statistical results about certain treatments. The population under test is most often not equivalent with the actual target population. The results are therefore mostly applicable on the population level, but most often not viable for the individual patient.

When assessing hierarchies, Howick talks about the necessity of a minimum effect size.

> "Yet a categorical ranking of randomised trials over observational studies leads to the paradox of effectiveness, whereby best evidence does not seem to support the effects of our most dramatically effective therapies. The paradox can be resolved by replacing strict hierarchies with a requirement that comparative clinical studies reveal an effect

280 Ben Goldacre. (2009): 34.
281 Jeremy Howick. (2011): 148.
282 Robyn Bluhm. (2005). "From Hierarchy to Network: a richer view of evidence for evidence-based medicine." in Perspectives in Biology and Medicine, 48(4): 535-547.

> size that outweighs the combined effect of plausible confounders. This requirement would allow observational studies to provide equally strong evidence to randomised trials in some cases, and would also be more exacting of certain randomised trials. Rather than displaying some (statistically significant) benefit, randomised trials would have to reveal a minimum effect size before being accepted as sufficiently strong evidence. Likewise, observational studies whose effect size outweighs the combined effect of plausible confounders can provide strong evidential support."[283]

The above quote, although rather lengthy, describes the problem with evidence hierarchies perfectly. What is actually judged is the overall value of a method, not the significance of the results. If the methods were used in a 'perfect', that is in a robust and unfailing way, then the ranking in such a hierarchy would be equally robust. However, since the methods itself are flawed, so is the hierarchy. A possible solution to the problem would be to understand evidence hierarchies as research guides that start with an idea about what to look for, for example the expert who asks the right questions. The question is then followed through, via research from the more 'simple' to the more 'complex', as in RCTs, and arrives at robust research results which can be used as a basis for medical decision making. And the results of trials and studies can be used in a less linear fashion when the individual patient is concerned.

It seems to be that the biggest obstacle to a compassionate treatment of patients is not so much too much or too little evidence, but all the paperwork that is required today and that keeps physicians away from their patients. Instead of having the time to spend on the bedside, they have to fill out forms and charts and because of the sheer number of patients, there is often just five minutes for each patient left. Five minutes, or even ten, are not enough to really establish a meaningful relationship to someone. The problem here is not a question of knowledge, wisdom or evidence but of administration versus humanity in principle.

Another problem concerning evidence hierarchies and EBM in general is what is often called 'guideline medicine.' Today there are a huge number of guidelines about patients, disease and treatments available. Guideline Central for example is an internet search tool for the United States which collates all available guidelines.[284] In the UK they are published among others by the NHS and are given out to GP practices as well as hospitals. As it turns out, guideline medicine is most often practiced in hospitals, whereas in GP practices guidelines are seen lying around but seldom adhered to.[285] The reasons that are given for this are that

283 Jeremy Howick. (2011): 187.

284 Guideline Central. https://www.guidelinecentral.com/summaries/. Last accessed on January 23rd, 2020.

285 Steven H Woolf, Richard Grol, Allen Hutchinson, Martin Eccles, Jeremy Grimshaw. (1999). "Clinical guidelines Potential benefits, limitations, and harms of clinical guidelines." in BMJ 318: 529.

doctors in GP practices most often claim that their particular patient does not fit the description or that the guideline is too narrow to treat a patient with multiple ailments.[286]

The term 'guideline medicine' is not meant in a neutral way. Guideline medicine is most often contrasted with patient centred medicine, because especially in clinical settings, instead of closely assessing the patient, after a quick examination, the 'relevant' guideline is used, no matter how applicable it is for the actual patient. Iona Heath argued at the 2015 Evidence Live Conference in Oxford that "We should never have produced guidelines. Instead we should have done summaries of the available evidence."[287] Guidelines have brought the fear of litigation to young doctors. Even though, some are defending guidelines, because they appear to be providing clinicians with the possibility to quickly "know" which evidence is important.

Guidelines are not per se bad, and they can be very helpful in quickly assessing a patient and having the most relevant information to hand in a short and precise manner. They are informative. But they are not more than that and should not be confused with good diagnostic skills or the necessity to look at every patient individually. They should only be a quick and easy 'go to' guide in the first instance of a diagnosis, but not taken as a treatment plan. It is obvious that guidelines can be very helpful for quickly assessing a situation but they are just guides, not more, and a conscientious practitioner should always question their usability for the individual patient. Clinical expertise and experience are again relevant to use guidelines in an appropriate way for the individual patient.

4.4.1 Bench to bedside or knowledge translation

A new approach that is heralded as an innovation trying to make EBM and its strict adherence to evidence hierarchies less severe, is called "bench to bedside" or "translational medicine"[288] and wants to solve the problem of using population based data for the individual patient. "Bench to bedside" however is not really a novel concept in medicine. In the early days of medical research, results were immediately used for and on the patient. The clinicians doing the research were the ones treating the patients. This "simple" approach is neither practical nor advisable

286 Steven H Woolf, Richard Grol, Allen Hutchinson, Martin Eccles, Jeremy Grimshaw. (1999): 530.

287 Iona Heath. (2015). "Eminence or evidence-based medicine: why this question is still relevant today." Conference talk at the Evidence Live Conference, Oxford.

288 Miriam Solomon. (2015):157. The two terms are often used equivocally, although depending on the user, they might mean different things and are given a different importance. Solomon acknowledges that phenomenon without further explaining it.

anymore, since medical research has become increasingly complex which makes it vulnerable to mistakes which can be understood and solved through medical trials. Therefore 'bench to bedside' and how it was practiced before EBM does not sound viable anymore today. The evidence hierarchies could therefore be understood as functioning as a safeguard to eliminate faults, flaws and mistakes.

The main problem with 'bench to bedside' or 'translational research' again is the rather vague definition of the terminology. It is clear that both terms, especially since they are often used interchangeably, are simply pertaining to a method to make laboratory results usable for the patient. However, it is not defined if the method is supposed to do so for the individual patient, thereby solving parts of the problem of external validity? Or if the method again only seeks to make results applicable on a population level, using the hierarchy of evidence production for its purpose? Every author has to define if a narrow or broad approach is discussed, which renders the terms as such difficult for discussion, because the two different approaches would lead to entirely different outcomes.[289]

A good example for the 'bench to bedside and back' approach is the development of penicillin. In animal trials, penicillin was successful, but in first human trials it was not. Going back and forth between the laboratory, animal trials and human trials in the end brought about the right dose in humans to cure.[290]

Solomon argues that EBM has a limitation in producing exactly that medical knowledge which is needed for this back and forth approach in medical science. "In particular it is a method that devalues mechanistic reasoning, in vitro and animal studies, and indeed everything except for high-quality clinical trials. But the high-quality clinical trials that characterise evidence-based medicine are in fact the final stage of the research process, which begins with mechanistic reasoning and laboratory trial and error and continues with the design of the high-quality clinical trial."[291] I concur with Solomon here, but still *understanding* 'bench to bedside' or 'translational knowledge' seems to be not sufficient where the solution to the problem of external validity is concerned. It seems as if an important component is missing, yet again, in the discussion about 'bench to bedside' and 'translational knowledge' and that component is the clinical expertise of the one who has to do the translating.

As has become already clear, clinical expertise should play a much more significant role in EBM than it does so far. Expertise is that skill which makes knowledge translation even possible. However, this expertise needs a solid theo-

289 Anna Laura van der Laan and Marianne Boenink. (2015). "Beyond Bench and Bedside: Disentangling the Concept of Translational ResearchHealth Care." in Health Care Analysis, 23: 32–49.
290 Miriam Solomon (2015): 163.
291 Miriam Solomon. (2015): 169.

retical background. One challenge in knowledge translation is the difficulty between knowing something and being able to explain. Having a skill does not necessarily mean that one has theoretical knowledge about it. If that theoretical knowledge is not present, then the possessor of the skill will not be able to explain it or to mentor and help others in acquiring it. Anja Silja, a German opera singer famously made that point in an interview.[292] She claimed that she could never teach singing, since she learned it intuitively. Most singers have a profound knowledge about how the voice functions. The role of the vocal chord, the larynx, and the different techniques to open and close the voice is taught in academies and singers can use that knowledge to at least explain their skill. If however, like Anja Silja, a singer has only learned to sing intuitively, without the technical background knowledge, it is almost impossible to explain the skill. Those professionals hear the mistakes that students make, but they would not be able to correct them. A sort of similar example used by Polanyi is the difference between the skill to drive a motorcar and the knowledge about why a motorcar is even able to be driven. An engineer is able to explain the workings of the machine, but that does not necessarily make him a better driver. Knowing about particulars and successfully using them are two different skills.[293]

4.4.2 Too much evidence for the single user

As we have seen, medical evidence grows exponentially every year. However, it is still expected from every clinician that he or she is up to date with all the available information, which is impossible due to the sheer amount of evidence. And even if it were possible, the 'naked' evidence is not all that plays a role in clinical decision making. A clinician undoubtedly has an opinion about possible treatments, and this opinion has informed the search for and the appraisal of the available evidence. Evidence might be able to change such an opinion or to inform it differently from the previous held belief, but it may as well not and the clinician most often will pass that 'unsaid' opinion along. This might be called clinician bias and seems to be an almost unavoidable one in clinical practice.

It seems to be not far from the norm that treatments are accepted very differently depending on their effect, their marketing, and their novelty. Greenhalgh uses the example of premature babies with a breathing difficulty due to the lack of the substance surfactant that is lacking in underdeveloped lungs, also called 'infant

292 Anja Silja. (1999). Television interview conducted by August Everding. "Da Capo". Accessed on youtube.com. https://www.youtube.com/watch?v=LZE4C_uzR8M. Last accessed on January 23rd, 2020.

293 Michael Polanyi. (1974). *Personal Knowledge: Towards a Post-Critical Philosophy.* Chicago: University Press: 20.

respiratory stress syndrome'. Since the early 1970's, women in premature labour received the steroid drug dexamethasone that accelerated the maturity of the lungs of the unborn babies. However, this specific treatment was not widely accepted. Surfactant treatment once the baby was born however was accepted almost immediately. I will use a table from Greenhalgh showing the effects and the reasons for the different acceptance rates for both treatments.[294]

Table 1: Effects and acceptance of different treatments for infant respiratory distress syndrome

	Surfactant treatment	Prenatal steroid treatment
Perception of mechanism	Corrects a surfactant deficiency disease	Ill-defined effect on developing lung tissue
Timing of effect	Minutes	Days
Impact on prescriber	Views of effect directly (has to stand by ventilator)	Sees effect as statistic in annual report
Perception of side effects	Perceived as minimal	Clinicians' and patients' anxiety disproportionate to the risk
Conflict between two patients	No (paediatrician's patient will benefit directly)	Yes (obstetrician's patient will not benefit directly)
Pharmaceutical industry interest	High (patented product; huge potential revenue)	Low (product out of patent; small potential revenue)
Trial technology	New' (developed in late 1980's)	Old' (developed in early 1970's)
Widespread involvement of clinicians in trials	Yes	No

The above table accumulates many of the problems associated with knowledge and evidence in EBM. Prescriber, patient and pharmaceutical industry interests favoured one and not the other, without a good reason to do so. Albeit the prenatal steroid treatment having been there first and proven to be successful, it was not widely accepted and many preventable death occurred, because of the reluctance to use the best available treatment based on the best evidence at the time. Available evidence needs to be implemented to be useful. Unused evidence is a waste of money, time

294 Trisha Greenhalgh. (2014). *How to read a paper: The Basics of Evidence-Based Medicine.* 5th Edition. Oxford, New York: Wiley BMJ Books.

and resources and ultimately health. And the failure to implement a treatment is often based on a lack of knowledge, demonstrated clearly in this case, where the evidence was present. There can be different reasons for this lack of knowledge.

1. Lack of information — the information is available in print but due to time constraints or poor adherence to print evidence, has not been spread wide enough, let alone be implemented.
2. Lack of understanding — the information is known but not understood properly and hence not valued as important.

The same treatment used by Greenhalgh as an example is also used in the Logo of Cochrane.[295]

Figure 4: The Cochrane Collaboration Logo

Source: https://www.cochrane.org/about-us/difference-we-make. Last accessed on November 14th, 2019.

> "The horizontal lines in the logo represent a series of trials that tested the benefits of a short inexpensive course of corticosteroids for women who were ready to give birth prematurely. The outcome of interest was infant mortality due to complications of immaturity."[296]

The horizontal lines signify each trial. Those touching the vertical line show no or very little effect, those on the left hand side show positive effects. The shorter the line, the more precise the results. The diamond "represents the combined effect of the treatment in all studies."[297]

Although these trials were known and published, only very little changed. Even after a meta-analysis was done and published, the practice of giving corticosteroids was not widely adopted. It needed a consensus statement by the NIH (Na-

295 Cochrane Collaboration. Our Logo. http://www.cochrane.org/about-us/our-logo. Last accessed on January 23rd, 2020.
296 Jeremy Howick. (2011): 18.
297 Jeremy Howick. (2011): 18.

tional Institute of Health, UK) to widely adopt the use of corticosteroids. Apparently many clinicians, especially in this case obstetricians and paediatricians, did not see eye to eye and the latter thought that the treatment ideas expressed by the further were just voiced to make their life harder.[298]

In the case of corticosteroids, evidence should have superseded any form of 'knowledge' the clinicians had assumed they had. The evidence for the superiority of the treatment was there, in abundance, and in this case 'abundant evidence' is good and should have easily been recognised as 'good' evidence with convincing data. The consensus statement, or GOBSAT to use Greenhalgh's term, was a solution here, but essentially would have been superfluous, because the evidence already had been there for quite some time. Using this example it might be fair to say that the 'wisdom' that Silva and Wyer are asking for in medical decision making was lacking, since the clinicians were not questioning their believes but assumed 'knowledge' where they should have assumed their 'knowledge' to merely be changeable 'evidence.'

4.4.3 Mindlines and tacit knowledge, or how evidence can be spread

A couple of times already 'tacit knowledge' has been mentioned and plays a special role in conjunction with 'mindlines' and with how knowledge is processed and used within the individual. The philosopher Michael Polanyi coined the term 'tacit knowledge' and used it as an argument against the value-free ideal that was prevalent in philosophy of science and the sciences in the 1970's.[299]

Tacit knowledge is knowledge that cannot be transferred by writing it down or by explaining the necessary skills. Tacit knowledge is inherent in every person, sometimes even without the person being consciously aware of possessing it. Like the example of the opera singer Anja Silia who is a marvellous classical soprano, but who cannot teach singing. Examples for tacit knowledge, as already explained, are skills like skiing or riding a bike. It is possible to technically explain these skills, but to actually learn them, the technical explanation is not enough. The skill has to be learned through trial and error. However, tacit knowledge is not the same as empiricism. According to Polanyi it is inherent and motivated by passions.[300]

Gabbay and LeMay in their influential book about practice-based medicine, call the process of interactive knowledge communication between practitioners 'mindlines'[301] and these mindlines are heavily based on Polanyis tacit knowledge.

298 Jeremy Howick. (2011): 163.
299 Michael Polanyi. (1974).
300 Michael Polanyi. (1966). *The Tacit Dimension*. Chicago: University Press.
301 John Gabbay and Andrée LeMay. (2011). *Practice-Based Evidence for Health Care: Clinical Mindlines.* London: Routledge.

Instead of reading up and statically following guidelines, the process of establishing and sharing mindlines is much more fluid. It is based on reading and assessing evidence, but only in the first instance and only in small amounts. The actual knowledge transfer is achieved by talking about the evidence and assessing it together with colleagues. However, this is not GOBSAT again, but meant in the daily practice of hospital medicine where colleagues communicate with each other, especially when and where 'special' cases are concerned. The approach sounds a bit like the translational medicine described above, however it seems to me to go beyond it, because mindlines do not start with a bench-to-bedside approach. It is more about how evidence is incorporated into every day clinical and medical practice. Gabbay and LeMay describe a combination of EBM and tacit knowledge. It seems as if neither is deemed sufficient on its own. This picture vehemently contradicts the idealised version of clinical practice which was described in the original EBM paper in which a junior doctor was able, by recourse to the available literature, to 'overrule' the opinion of the more senior member of staff. Gabbay and LeMay seem to portray a much more realistic picture of actual clinical practice in which the senior clinician is still adhered to and in which there are 'consensus meetings' happening in the hallway. Mindlines are growing from experience and are coming from people that are trusted.[302] Mindlines take patient preferences into account and therefore could be used as a step to making evidence-based practice more patient-centred. They are not directly usable for medical research, because although they lead to questions, they do not necessarily lead to research questions. They might do in special circumstances, but the power of understanding mindlines and tacit knowledge lies in their use for medical practice.

4.5 Conclusion

Evidence-based medicine is not a new theory of epistemology. It uses parts of epistemological theories where those are applicable for the special use in medicine, but it cannot for itself claim to establish a new 'theory of knowledge'. It is important to understand EBM to be based on 'evidence' and not on 'knowledge' and thereby to acknowledge that the base on which medicine is put in the case of EBM is constantly changing, incorporating new evidence and discarding 'bad' evidence as robust research results are generated and updated on a continuous basis. Therefore the definition of evidence as 'good reason for belief" or "grounds for belief" can be upheld. Albeit with the special addition that in order to claim 'good reason to belief', for example in the validity of a treatment, all data concerning

302 John Gabbay and Andrée LeMay. (2004). "Evidence based guidelines or collectively constructed 'mindlines?' Ethnographic study of knowledge management in primary care." in BMJ (329): 4.

that special piece of evidence has to be available, so that a real informed decision can be made on part of the physician. So that even if the evidence proves to be 'wrong' or 'bad' through later research, it still is possible to maintain that it was the best available evidence at that exact point in time. Therefore, for the definition of evidence it is permissible to lack the 'truth condition' that is an integral part of the definition of 'knowledge.' Medical knowledge is that what is needed to render the available evidence useful in clinical practice.

Evidence hierarchies are useful for the production of this 'robust' evidence but they need to be challenged in medical practice where the evidence needs to be used for the individual patient. In the latter case, the 'best' evidence for a particular treatment at the given time must be contrasted with the 'best' evidence for a treatment for the individual patient. And these treatments might be very different from each other, because the patient might exhibit idiosyncrasies which are not compatible with the 'best' treatment on the population level.

The problem of 'too much evidence' can only be solved by accepting the challenge that not all evidence can be known by all physicians at every point in time, and that the dialogue among colleagues is important to maybe partially close the resulting gap. Mindlines and tacit knowledge are coping mechanisms in medical practice to handle the amount of evidence and to partially solve the problem of external validity as well, because both 'soft skills' are necessarily used to interpret research results for the individual patient.