

Marie-Caroline Schulte

Evidence-Based Medicine – A Paradigm Ready To Be Challenged?

How Scientific Evidence Shapes Our
Understanding And Use Of Medicine

OPEN



J.B. METZLER

Evidence-Based Medicine – A Paradigm Ready To Be Challenged?

Marie-Caroline Schulte

Evidence-Based Medicine – A Paradigm Ready To Be Challenged?

How Scientific Evidence Shapes Our
Understanding And Use Of Medicine

OPEN



J.B. METZLER

Marie-Caroline Schulte
Hamburg, Germany

Dissertation zur Erlangung des Grades der Doktorin der Philosophie (Dr. phil.) an der Fakultät der Geisteswissenschaften der Universität Hamburg im Promotionsfach Philosophie, vorgelegt von Marie-Caroline Schulte, Hamburg, 2017.



ISBN 978-3-476-05702-0 ISBN 978-3-476-05703-7 (eBook)
<https://doi.org/10.1007/978-3-476-05703-7>

© The Editor(s) (if applicable) and The Author(s) 2020. This book is an open access publication. **Open Access** This book is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this book are included in the book's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the book's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This J.B. Metzler imprint is published by the registered company Springer-Verlag GmbH, DE part of Springer Nature.

The registered company address is: Heidelberger Platz 3, 14197 Berlin, Germany

Content

1	Introduction.....	1
1.1	EBM — a unifying force in medicine?	1
1.2	EBM — one term used for different areas of medicine.....	4
1.3	The necessary division of EBM into ‘evidence-based research’ and ‘evidence-based practice’	6
1.4	EBM as a new paradigm in medicine — or is it?	7
1.4.1	The possible ‘ideological’ base of EBM:	10
1.5	Conclusion.....	12
2	The methodology of evidence-based research.	15
2.1	RCTs and their importance for the methodology of EBM	15
2.1.1	Scientific goals of RCTs	16
2.1.2	Publication of all trial data	19
2.1.3	Introduction of the basic methods of RCTs.....	19
2.2	The different phases of RCTs.....	20
2.2.1	Phases of RCTs according to Benedetti:	21
2.3	Confounding factors and their influence on trial results	23
2.3.1	Randomisation	25
2.3.2	Confounding factors during trials:	29
2.3.3	Blinding — its problems and its virtues.....	30
2.3.4	Blinding in surgical trials	33
2.4	Placebo and the placebo response	34
2.4.1	Passive versus active placebo.....	36
2.5	Stopping clauses.....	37
2.6	External validity	40
2.6.1	External validity and trial design	41
2.6.2	External validity and recruitment.....	43
2.6.3	External validity and n-of-1 trials	47
2.6.4	How can external validity be achieved?.....	47
2.6.5	Publication bias.....	49
2.6.6	Surrogate outcomes versus primary endpoints.....	50
2.6.7	Equipoise.....	51

2.7	Unnecessary trials	52
2.7.1	ECMO — or an example why not all RCTs are equal	53
2.8	Conclusion.....	55
3	Informed Consent and shared decision making in EBM.....	57
3.1	‘Informed consent’ in regular medical practice.....	57
3.1.1	Implied or simple consent	58
3.1.2	‘Informed consent’ and trust	59
3.1.3	Informed refusal	59
3.1.4	Rare limitations to ‘informed consent’	60
3.2	The shift from paternalism to informed consent	62
3.2.1	From the Hippocratic Oath to the Declaration of Helsinki	62
3.2.2	‘Informed consent’ and ‘assent’ for children, teens and physically or mentally incapacitated patients.....	65
3.3	The Tuskegee Experiment.....	66
3.3.1	Consequences of Tuskegee	67
3.4	‘Informed consent’ in clinical research	68
3.4.1	Use of already established data in and for research	69
3.4.2	The issue of trust revisited	70
3.5	How to transmit medical information successfully	71
3.6	Conclusion.....	72
4	Knowledge does not equal evidence — what to do with what we have evidence for?.....	75
4.1	Knowledge versus evidence — why the distinction is important.....	75
4.1.1	Possible definitions of ‘evidence’ and ‘knowledge’ for EBM.....	76
4.2	EBM as a new theory of epistemology in medicine?	80
4.3	Evidence hierarchies	82
4.3.1	Systematic reviews, meta-analyses and RCTs	84
4.3.2	Observational studies	88
4.3.3	Expert judgement, clinical judgement, clinical expertise.....	90
4.3.4	Mechanism and causation	94
4.4	Problems with hierarchies and possible solutions	97
4.4.1	Bench to bedside or knowledge translation.....	99
4.4.2	Too much evidence for the single user.....	101
4.4.3	Mindlines and tacit knowledge, or how evidence can be spread	104
4.5	Conclusion.....	105

- 5 Homeopathy — a case in point why EBM is so important —
or, “the plural of anecdote is not data.” 107**
 - 5.1 Alternatives to medicine? 107
 - 5.2 Historical context of homeopathy 107
 - 5.3 The methodology of homeopathy..... 109
 - 5.4 Homeopathy today 113
 - 5.4.1 A defence of homeopathy — or is it? 116
 - 5.5 Acupuncture 119
 - 5.5.1 Acupuncture under trial 120
 - 5.5.2 Acupuncture and safety 121
 - 5.6 Conclusion..... 122

- 6 Conclusion 123**

- Bibliography 127**

- Appendix..... 143**
 - Kurzfassung der Ergebnisse 143
 - English Version..... 146

List of figures and tables

Figure 1:	ECMO schema, to provide neonatal extracorporeal oxygenation.....	54
Figure 2:	Standardised pyramid of evidence	83
Figure 3:	What the Good Pharma Scorecard wants to achieve and their points of access	87
Figure 4:	The Cochrane Collaboration Logo.....	103
Table 1:	Effects and acceptance of different treatments for infant respiratory distress syndrome.....	102



1 Introduction

“Despite one popular caricature of ‘the philosopher’ as being somehow ‘deep’, the ones I know (coming mostly from what is known as the ‘analytical’ school) make it a point of honour never to have anything profound or clever to say on any matter whatsoever. In fact, they consider it the hallmark of a good philosopher that one is always prepared to ask the sort of naive question that others are too scared to ask, for fear of appearing ignorant.”¹ (Michael Loughlin)

1.1 EBM — a unifying force in medicine?

Evidence-based medicine (EBM) was and is all the rage in medicine and the ‘new way’ to practice and teach medicine in our century. The term ‘evidence-based medicine’ was coined in 1992 by scientists at the McMaster University, Hamilton, Ontario.² Today it is *the* established method in medicine, at least in what is called ‘allopathic’ or ‘conventional’ medicine and the term is by now part of the very fabric of medicine and of medical knowledge. EBM encompasses research, practice and the teaching of medicine, and many articles in peer review journals, and even entire journals, have dealt with the how and why and where of EBM since its appearance on the medical scene. However, it is still not quite clear what exactly EBM is, what exactly it encompasses and why it should be so much superior than other movements in medicine. The definition of EBM offered in the original paper by the ‘Evidence-based medicine working group’ is “the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients. The practice of evidence based medicine means integrating individual clinical expertise with the best available external clinical evidence from systematic research.”³ Sackett and colleagues have since refined the definition: “evidence-based medicine is a systematic approach to clinical problem solving which allows the integration of the best available research evidence with clinical

-
- 1 Michael Loughlin. (2009). “The basis of medical knowledge: judgement, objectivity and the history of ideas.” in *Journal of Evaluation in Clinical Practice*. 15(6): 935.
 - 2 Gordon Guyatt, et.al. for the Evidence-Based Medicine Working Group. (1992). “Evidence Based Medicine: A New Approach to Teaching the Practice of Medicine.” in the *Journal of the American Medical Association (JAMA)*, 268 (17): 2420 - 2425.
 - 3 David Sackett, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. (1996). “Evidence based medicine: what it is and what it isn’t.” in the *British Medical Journal (BMJ)*, 312: 71-2.

expertise and patient values.”⁴ However, and although both of these definitions are still frequently cited, EBM has changed over time. Probably inevitably so because of its own growth factor, but also because EBM and associated terms can be, and are used, to further the cause of pharmaceutical companies and interest groups alike without giving EBM more credence or credibility.⁵

For many reasons EBM has still ‘won’ the claim to be the ‘best’ form of medicine of today against some competition. However, when evaluating EBM historically, it is not quite clear against what type or form of competition EBM had to actually compete. It seems more reasonable to argue that *some* sort of change in medicine was simply inevitable. The technical and inventional progress in medicine over the decades preceding the advent of EBM, including the first randomised controlled trials (RCTs), clinical trials which compare treatments with each other in two groups of patients and which were already performed before the method became part of a medical standard, was so profound that a rethinking of medical research and medical practice was the only possible consequence. Therefore it can be argued that EBM appeared on the scene at the right time and with the right instruments to pick up the different pieces of medicine and to attempt to unite them into one coherent system.

EBM was and is embraced by many physicians, clinicians and medical researchers, especially after it transpired that EBM can eliminate faulty medical reasoning that was prevalent over a long time. There are multiple examples available for this phenomenon. Many treatments that were widely accepted in the medical community up until the 1990’s were discontinued after EBM had put them to the test and considered them to do more harm than good.⁶ Prominent examples for this are the hormone replacement therapy (HRT) for menopausal women and the thalidomide scandal, causing birth defects, both of which will be discussed thoroughly in the coming chapters.

It is however important to clarify right from the beginning that there are treatments which will never be tested with the EBM methodology, since it would make no medical sense to do so. Treatments in this category are, for example, the removal of the appendix in the case of acute appendicitis, the Heimlich manoeuvre to unblock a blocked airway in a choking patient or heart fibrillation to restart a

4 David Sackett, Strauss SE, Richardson WS, et al. (2000). *Evidence-based medicine: how to practice and teach EBM*. London: Churchill-Livingstone.

5 Trisha Greenhalgh, Jeremy Howick and Neal Maskrey for the Evidence Based Medicine Renaissance Group. (2014). “Evidence based medicine: a movement in crisis.” in *BMJ*: 1-7.

6 Imogen Evans, Hazel Thornton, Ian Chalmers, and Paul Glasziou. (2011). *Testing Treatments: Better Research for Better Healthcare*. 2nd. Edition. London: Pinter and Martin: 3.

stopped heart.⁷ These treatments, and many more, are proven to be the only effective treatment for the underlying disease. Cause and effect are directly observable and well established. Often however, pure observation is not enough to successfully establish a causal connection. Researchers and clinicians have, over the course of medical history, dealt with spurious correlations which looked plausible in the beginning, but where not so in the end. A wonderful example from the Middle Ages about such a spurious correlation that actually worked to a certain degree and therefore was widely accepted, are the masks of the plague doctors. Plague doctors wore a mask over their entire face with a long beak, a black coat with a cowl and a hat and gloves. The masks contained herbs because the assumption was that the plague was passed on by bad odours and therefore transmitted through the air. Patients suffering from the disease often stank badly because of their open and infected wounds.⁸ In actuality, the plague is passed on through droplet infection. Since the infection was still airborne but based on bodily fluids, the masks and overall get-up helped, because they prevented the plague doctors of inhaling the infected droplets. The reasoning was spurious, the preventive 'treatment' nonetheless useful. If that is the case, it would be arguable that as long as the ends are achieved, the means do not matter. Unfortunately there are a lot of examples in medicine where spurious correlations led to very harmful treatments. One example out of the more recent medical history is about hormone replacement therapy (HRT) for menopausal women. HRT was considered to take away the unpleasant side effects of menopause and to even prevent certain types of cancer. And it seemed to work well. However, after conducting randomised-controlled-trials (RCTs), trials that compared HRT to either other treatments or no treatment, it transpired that HRT did not prevent these types of cancer, nor did it prevent possible heart attacks. On the contrary, HRT has considerable side-effects, ranging from headaches to cancer that were for some time swept under the carpet.⁹ The reason why it was so "successful" was that it was most often prescribed in more affluent areas where many women were already in better shape, lived a healthy lifestyle and where overall more aware of their health and therefore better equipped to quickly deal with upcoming medical problems.¹⁰ EBM and its rigorous methods prevented more women from receiving HRT and therefore prevented the treatment from doing more harm.

As well as EBM seems to function for all these examples, there are still many aspects of it that are not entirely explained, defined or understood, even by staunch

7 Jeremy Howick.(2011). *The Philosophy of Evidence-Based Medicine*. Oxford: Wiley Blackwell, British Medical Journal (BMJ) Books. Introduction: xiii.

8 Jacob L. Kool and Robert A. Weinstein. (2005). "Risk of Person-to-Person Transmission of Pneumonic Plague." in *Clinical Infectious Diseases*, 40(8):1166–1172.

9 Imogen Evans, Hazel Thornton, Ian Chalmers, and Paul Glasziou. (2011): 16.

10 Ben Goldacre. (2008). *Bad Science*. London: Fourth Estate: 108.

supporters of EBM. As with all ,new‘ systems in an established field, EBM was and is not without its critics and their often very valuable criticisms and contributions.¹¹ EBM in general is enthusiastic about criticism, because medicine, like all sciences, advances through criticism and through asking the right questions. As will become obvious throughout the dissertation, some of these criticisms are aimed at improving EBM but some are just phrased to discredit the program altogether. This dissertation will defend EBM, albeit all its shortcomings, as the currently best way to conduct medicine. And it will offer a way forward to produce, understand, and use evidence in medical research, and, even more importantly, in medical practice.

1.2 EBM — one term used for different areas of medicine

EBM is by now sort of an umbrella term for multiple areas in medicine. Related terms are also in use, and most often their aim is to either be more specific or to ‘solve’ an apparent problem of EBM. Examples of such terms are: EBP for evidence-based practice¹² (I don’t like using EBP because it also stands for evidence-based policy and even though both are context specific, there is a real danger of confusing them.) EBP can also refer to social work or education and therefore always has to be specified. EBHC which stands for evidence-based health care, EBN for evidence-based nursing and the number of abbreviations containing ‘evidence’ seem to increase daily. And the abbreviations seem to go in and out of fashion and therefore continue to be confusing. HTA stands for ‘health technology assessment’, and CER stands for ‘comparative effectiveness research’, these are two more terms that are interchangeably used with EBM but which do mean slightly different things and are themselves in need of clarification.¹³ A fairly new term that appears by now in conjunction with, or as a contrast to EBM, is PCHC standing for ‘person centred health care’.¹⁴ PCHC is used in arguments both as an add-on to EBM and as a new way of looking at medicine. Therefore it is yet again not a term that can be used as such but is context-dependent as to its actual use.

11 Robert Smith. (2014). “Medical research—still a scandal.” in the *bmj* opinion. <http://blogs.bmj.com/bmj/2014/01/31/richard-smith-medical-research-still-a-scandal/>. Last accessed on January 23rd, 2020.

12 K. Ann McKibbin. (1998). “Evidence-based practice” in *Bulletin of the Medical Library Association* 86(3):396-401.

13 Bryan R. Luce, et al. (2010). “EBM, HTA and CER: Clearing the Confusion” in *The Milbank Quarterly*. Vol. 88.

14 European Society for Person Centred Health Care. www.pchealthcare.org.uk. Last accessed on January 23rd, 2020

Another example of a presumed ‘alternative’ is narrative-based medicine.¹⁵ It claims that the patient’s narrative, his or her story, is needed to make an informed decision about a medical treatment. However, in every patient assessment the personal history of the patient is taken and used as part of the treatment process. But narratives can never replace evidence. They are necessarily subjective and have to be understood by the physician in the wider context of the patients diagnosis. One improvement that a focus on narratives can bring to the diagnostic side of EBM might be that more time is spend on the wishes and values of the patient, already making the patient feel more comfortable in the medical setting.

All these terms and their underlying approaches and assumptions however are still somewhat part of EBM, because none of them would work without recourse to a solid medical evidence base. In every proposed scenario, from ‘person-centred health care’ to ‘narrative based medicine,’ the clinician nonetheless needs a solid foundation of medical evidence to make a treatment decision, and EBM seems to still be the only solution providing this solid foundation. The clinician can still decide to not use the ‘best’ available evidential treatment either because it does not fit the individual patients needs or it is not available at the point in time. However, this decision is only possible when the treatment options, based on medical evidence, are known. Even if the option is to forgo treatment, the associated benefits and risks are based on solid evidential grounds and can therefore be assessed and calculated, at least to a certain degree. There will always be an element of surprise in medicine, like spontaneous healing or remission, but even this element of surprise will to a certain degree be part of the calculation and decision on treatment, which by its very nature is based on mathematical possibilities.

For clarity’s sake and because many of these terms are referring to EBM, are using evidence as their underlying base or are simply too confusing to use, the term EBM will be used throughout the dissertation. I will do so, however, with the important caveat that the term ‘EBM’ itself is in need of clarification. The most important distinction that should be made explicit when using the term ‘EBM’ is between evidence-based medical *research* and evidence-based medical *practice*, since these two areas differ widely from each other today and hence should be discussed separately, especially where ethical and methodological questions are concerned. The division of the term ‘EBM’ into ‘evidence-based medical research’ and ‘evidence-based medical practice’ will also be one of the key arguments of the dissertation, since it is that division which will make medical practice more person-centred. And making EBM more person-centred again will save it from much of the criticism levelled against it.

15 Trisha Greenhalgh. (1999). “Narrative based medicine: Narrative based medicine in an evidence based world.” in *BMJ* (318).

1.3 The necessary division of EBM into ‘evidence-based research’ and ‘evidence-based practice’

Medical research starts at the molecular level and aims to make inferences from there to the population level and tries to answer questions regarding all patients with a certain disease, illness or handicap. Medical practice, in contrast, deals with the individual patient and has to ask which treatment is the right one for this special patient, with the patient’s individual circumstances, preferences and values taken into consideration.¹⁶ The two areas of medicine, research and practice, are closely related and dependent on each other. But they are not the same and should be discussed in a separate manner and with a slightly different focus with regard to their ethical and methodological advantages and problems.

Although it would be tempting to use the abbreviations EBP, for ‘evidence-based practice’, and EBR, for ‘evidence-based research’, again it seems counter-intuitive to the usual use of the term ‘medicine’. ‘Medicine’ has always encompassed both, practice and research and for the longest time it seems as if the two were so fluidly going hand in hand that a clear distinction between them appeared to be none-sensical. However, today with EBM in full effect and with the incorporation of all the medical advances of the last decades, research and practice are, sort of by necessity, two separate entities in medicine, often with different personnel involved, or even employing entirely different companies to perform research. Pharmaceutical companies are conducting their own research into new drugs and have changed how the science of medicine works considerably by focusing entirely on research and the production of drugs. Physicians are still the ones dispensing the drugs and researchers in a clinical setting are coming up with new drugs and treatments, but nevertheless pharmaceutical companies are a driving force in the production and marketing of drugs.¹⁷

Since medical trials, the evidential basis for EBM, and the hallmark of evidence-based research, are fairly complex to perform and need a special set up, it can be a methodological advantage to remove them from the regular running of a clinic or hospital.¹⁸ Big university hospitals often can and do both, research and regular care, but they tend to be the exemption to the rule. And even if research and clinical care are happening in the same building, they are distinct from each

16 Julian Reiss and Ankeny, Rachel A., (2016). “Philosophy of Medicine” in The Stanford Encyclopedia of Philosophy (Summer 2016 Edition), Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/sum2016/entries/medicine>. Last accessed on January 23rd, 2020.

17 Peter Konrad. (2005). “The Shifting Engines of Medicalization.” in *Journal of Health and Social Behavior*. 46(1): 3-14.

18 Ben Goldacre. (2012). *Bad Pharma: How medicine is broken and how we can fix it*. London: Fourth Estate: 225. Ben Goldacre lobbies for large randomised trials which can happen in everyday medical care by integrating existing patient data. That would be a solution where care and research are combined.

other, because those patients who participate in a trial are not regular patients anymore, but change into participants who are treated differently. Participants agree to be part of a trial, knowing that the medication they receive is part of a trial and its safety not yet fully established. Some participants could be receiving a treatment that might not have any clear benefits and might even be harmful, or receiving a placebo that at least does not harm but also does not directly benefit the participant, above and beyond the placebo effect. Clinicians on their part change into researchers. And researchers are by their very nature interested in the test results on the population level, and not so much in the individual patient. They are blinded to the actual treatment of their participants so that all sorts of biases can be avoided.¹⁹

There cannot be a ‘one-size-fits-all’ approach in medical practice, nor really in clinical research, and hence the term EBM encompassing both can be misleading, but, as said above, it is the most useful term and it can easily be made context specific as to which part of the overall EBM methodology is actually meant. One major critique that is levelled at EBM is that it is too population centred and has lost sight of the individual patient. Later chapters will explicitly deal with this criticism in various ways and they will provide a solution that, for now in simple terms, comes back to the distinction between evidence-based research and evidence-based practice.

1.4 EBM as a new paradigm in medicine — or is it?

The authors of the original 1992 JAMA paper who have coined the term EBM made some interesting claims about the philosophical foundation of EBM. They called it a new paradigm in medicine and based this on Thomas Kuhn’s definition of a scientific paradigm shift or revolution.²⁰ According to Kuhn, a scientific revolution is a shift in scientific methodology and its accompanying worldview, i.e. the old paradigm, triggered by a revolutionary process that leads to a new methodology and worldview, i.e. the new paradigm. This process or shift starts if and when the old paradigm cannot successfully explain and incorporate new problems anymore because by doing so, it would inevitably collapse. In order to move on in science, a new paradigm has to appear that can successfully incorporate and explain these anomalies. This ‘new’ paradigm will work until it too will be unable

19 The actual set up of regular RCTs and their related problems and biases will be explained in detail in the next chapter. The reason why a short description here is necessary, is to set the stage for all discussions later in the text, since the division of EBM in research and practice will be the main topic and focal point of the entire dissertation.

20 Thomas S. Kuhn. (1996) *The Structure of Scientific Revolutions*. Third Edition. Chicago and London: The University of Chicago Press.

to explain upcoming anomalies. The old and the new paradigm are, according to Kuhn, “incommensurable” because, or so he claims, scientists working in different paradigms do not even speak the same scientific language anymore and therefore cannot communicate meaningfully with each other. A valid criticism to this definition of a paradigm in science is that if the language would really be that ‘incommensurable’ then Kuhn, and everyone else who tries, should not even be able to explain different paradigms in a coherent and comparative manner.²¹

Kuhn claims that “Since new paradigms are born from old ones, they ordinarily incorporate much of the vocabulary and apparatus, both conceptual and manipulative, that the traditional paradigm had previously employed. But they seldom employ these borrowed elements in quite the traditional way.”²² That means for Kuhn that they do not share a common measure.²³ The terms the scientists are using do not have to same meaning from one paradigm to the next. Kuhn centres his incommensurability theory on specific examples like the shift from the Newtonian to the Einsteinian understanding of space and time. Admittedly, the shift here was a major scientific one and left some physicists, who were still believing in the Newtonian world view, stranded. The progress in medicine however does not seem to be comparable in magnitude to the Newton-Einstein example, or the shift from the Aristotelian to the Copernican understanding of the solar system.²⁴ Progress in medicine happened and still happens gradually and often slowly and not in seismic shifts which would make the old non-comparable to the new.

Trisha Greenhalgh, one of the proponents of EBM, claims, based on Kuhn, that the paradigm shift can also be triggered by a young generation of scientists who are not accustomed to the established paradigm and are starting to question its premises. Since these questions are not appreciated by the senior scientists, the young group branches out and establishes a new paradigm.²⁵ However, even if a younger generation is branching out and establishing a new view on the medical

21 Scott R Schon and Donald E Stanley. (2003). “A philosophical analysis of the evidence-based medicine debate” in *BioMedCentral Health Services Research* 3(14).

22 Thomas Kuhn. (1996): 149.

23 Alexander Bird. (2013). “Thomas Kuhn”, *The Stanford Encyclopedia of Philosophy*. Ed. Edward N. Zalta, <https://plato.stanford.edu/archives/fall2013/entries/thomas-kuhn/>. Last accessed on January 23rd, 2020.

24 Kuhn uses the example of Einsteinian and Newtonian physics and although for the most part that shift in the sciences really did happen, the Newtonian system is still used for example to explain gravity. In the realm of the microcosmos, Newtonian physics still holds. (Brian Greene. (2003). *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*. New York: Vintage Books.)

25 Trisha Greenhalgh. (2012). “Why do we always end up here? Evidence-based medicine’s conceptual cul-de-sacs and some off-road alternative routes.” in *Journal of Primary Health Care*; 4:92-7.

problems, that is not a paradigm shift in the Kuhnian sense, since the overall medical language is still the same. The groups can communicate in a meaningful way, even if new terms are incorporated, like EBM itself, which has emerged as a new term. There is a lot of criticism levelled at Kuhn for his theory of scientific change. The problem with Kuhn's theory, especially for medicine is the 'incommensurability theory'. EBM in the early 90's brought something new to the medical world, a new approach to look at all the available evidence, to produce new evidence and to assess the available evidence. However, that 'something new' was not a radical change or an anomaly in an established practice that needed to be included in a new and different way. The 'new' approach was to acknowledge that not all of medical practice is and was actually based on the most robust evidence and that treatments are and were commonly used which could be outdated and might even be dangerous. The proponents of EBM also realised that medicine was not unified. Prescription and treatment habits did not only differ between countries, but very often between surgeries in the same country and sometimes even within a single surgery, if multiple clinicians were employed and were part of the decision making process.²⁶ The prescription and treatment habits were often based on previously learned approaches which were not questioned over time. One goal of EBM was to make these differences disappear and, in consequence, to provide all patients with the same, and if possible excellent, level of care.

Therefore it is arguable if EBM is and was far less a paradigm in the Kuhnian sense, but was far more an inevitable measure at the time, and to a degree even today, to incorporate the rapid medical scientific progress between the 1950's and the 1990's into everyday medical practice and teaching of medicine. Medicine was shifting because of the progress and the accompanying changes, and a unifying movement, like EBM, made it possible to shift it in such a way that medical research as well as patient-centred care became equal parts of the fabric of medicine by asking the right methodological questions but without devaluing everything that medicine had achieved so far.

EBM still works by using the available information technology of our time, i.e. the internet and all the medical search engines, online journals and medical publishing corporations that are coming with it. The authors of the original paper had already predicted that the amount of available information would be growing exponentially in the future and hoped that there would be a workable solution for the problem. Today one of the biggest challenges of EBM and all other approaches

26 Gordon H. Guyatt. et.al. for the Evidence-Based Working Group. (2000). "Users' Guide to the Medical Literature: XXV. Evidence-Based Medicine: Principles for Applying the Users' Guides to Patient Care." in JAMA 284(10): 1290-1296.

in medicine is exactly the problem of too much information coupled with publication bias and industry funding which can distort the evidence base, since still not all information is available to everyone.²⁷

EBM questioned the available treatments and medications at the time and demanded that many of them were put under rigorous testing for safety and efficacy. As seen in the first paragraph, not all treatments were put under test, because their causality to the underlying disease was well established. This alone could be enough to argue against any incommensurability based on language. The language of medical research and medical practice changes necessarily over time and that change is predictable and accepted. However, it changes fluidly and slowly, not in an instant, and practitioners can always converse meaningfully with each other.

The era of heroic medicine were treatments like blood-letting were used had already ended well before the advent of EBM.²⁸ Therefore it is save to say that medicine before EBM was also based on research. But it was lacking in an overall, globally usable structure and because of the ever increasing methods of communicating research results globally, it needed a type of foundation on which it was possible to utilise the available information wherever it was needed. Most vaccines were already established for example and the language regarding immunology did not change abruptly from one day to the next, and neither did the authors of the JAMA paper seem to expect that. They used the term Kuhnian revolution or paradigm shift, it seems not really in a philosophical sense, but as a convenient catch phrase to demonstrate how big the shift that they were attempting actually was, and apparently how important it was to give the new movement a solid base, not only a scientific one but also a philosophical and an ideological one. Sehon and Stanley for example understand the term ‘paradigm’ as used by the original authors as a metaphor, however a poorly chosen one, which, as we have seen, does not work according to the Kuhnian definition of ‘paradigm’ and ‘paradigm shift.’²⁹

1.4.1 The possible ‘ideological’ base of EBM:

The scientific and the philosophical base of EBM will be the two main foci of the dissertation. However, it is interesting to at least briefly discuss the ideological

27 Ben Goldacre. (2012): 27. Goldacre and others show how difficult it can be to get hold of complete trial protocols and clinical study reports. Many companies claim confidentiality issues and withhold the information. As Goldacre and others rightfully claim, patient data could be anonymised, but so far the companies are not legally bound to disclose all available information.

28 P. Stavrakis. (1997). “Heroic medicine, bloodletting, and the sad fate of George Washington.” in Maryland Medical Journal. 46(10): 539-540.

29 Scott R Sehon and Donald E Stanley. (2003).

base of EBM. The ideological base is the leaving behind, or at least the challenging, of authority. Because EBM is based on research facts, authority in the form of senior clinicians telling everyone else, including the patients, what to do and thereby acting paternalistically, will be undermined and vanish in due time, or so the authors predicted. The original JAMA paper claimed that a junior physician was in a better position to judge a patient, after using the methods of EBM rather than after asking and relying on his or her superior. The authors were aware that that approach could undermine hospital hierarchies and could threaten the value of experience. In fact, although the authors specifically state that this outcome is not intended by them,³⁰ the undermining of hierarchies and the lack of use of clinical experience is one of the main criticism towards EBM. In a little side note, it is interesting to observe that, at least in Germany, patients are willing to pay extra to see the head physician and to attain special treatment, although it is acknowledged that the head physician of a department might be the one member spending the least amount of time on the ward.

Following the original thinking of EBM, the junior doctor would be precisely as well equipped to make the right diagnosis and be as successful as the senior one. What the authors undermine with this assumption is the fact that it is also a part of clinical experience to be able to interpret the results from a trial, from diagnostic tests or from a literary search about the symptoms or the diagnosis. They claim the opposite, but the idea that a junior physician can utilise the available evidence in the same way in which a senior physician would be able to do, already implies the unimportance of clinical expertise. However, clinical expertise and experience is needed to ask the right questions and to realise that a ‘minor’ symptom mentioned by the patient can be the actual indicator for a major disease. Even if the right diagnosis is found, clinical experience is needed to decide which treatment to choose in this particular instance. In order for the patient to be in the centre of the medical practice, clinical experience must be understood as vital and should be understood in today's view of EBM as being outside the evidence hierarchy altogether, informing and being part of the decision making process a) about which research to conduct, and b) about which treatment to administer to the patient.

Research results, notwithstanding their robustness, are only half of what is needed to successfully treat an ailing patient. Other forms of evidence in the form of expertise by the physician might be more important for this patient and it might even be the case that not treating a patient could be an option. Some patients, especially towards the end of their lives, prefer to only receive palliative care and one would not treat such a patient with antibiotics against for example, an inflamed lung, whereas a young patient presenting with the same symptoms but otherwise healthy would definitely be treated with antibiotics. The basic research, i.e.

30 Gordon Guyatt, et.al. (1992): 2423.

how to successfully treat an inflamed lung, is the same for both cases. The individual patient's history and situation however decides the actual treatment.³¹

The authors are careful to point out that colleagues can learn from colleagues, and that the actual practice of medicine cannot be taught exclusively through books and articles but must also be taught through the greater expertise of these clinicians who have a lot of experience in their field. It seems as if the authors themselves would have been slightly uneasy in devaluing expert opinion of clinicians completely.

One main theme of EBM is the literary search, i.e. finding and assessing the important medical literature for a given medical case. At McMaster Hospital, where most of the authors of the original JAMA paper worked, each floor had a search computer with a simple Medline³² search tool. All through the following chapters, it will become obvious that the simple search approach that was heralded in 1992 is just not feasible anymore in 2017, and for multiple reasons. The most compelling of these would be the sheer volume of literature in any given medical field and the number of search engines that provide access to medical articles and peer reviewed journals. The same article can be accessible via multiple links, some of them with open access, some only accessible through research institutions.

EBM today is often accused of being "Cookbook medicine"³³, however, in the original paper, the proponents of EBM claim the exact opposite, saying outright that cookbook medicine has its appeal, because it is quick and easy and EBM is not, but that EBM in the long run will provide by far the better base for medicine because it is more rigorous and better designed to help the actual patient.³⁴

1.5 Conclusion

The advent of evidence-based medicine should not be understood as a paradigm shift in the Kuhnian sense. The different fields and areas of medicine still have a meaningful way to communicate with each other and research results and medications that were used before EBM were not simply abandoned but scrutinised and either tested, kept or abandoned, according to their safety, efficacy, and effectiveness. But even if EBM is understood as 'just' inevitable change because of the advances in science and medicine itself, some of the criticisms levelled at EBM

31 Gordon H. Guyatt, et.al. for the Evidence-Based Working Group. (2000): 1291.

32 MEDLINE/PubMed Resources Guide. www.nlm.nih.gov. Last accessed on January 24th, 2020. Medline has merged since the 90s and naturally has expanded and is one of many search engines for medical articles.

33 RB. Darlenski, et. al. (2010) "Evidence-based medicine: Facts and controversies." in *Clinics in Dermatology*, Elsevier:554.

34 Gordon Guyatt, et.al. (1992): 2423.

are still very valid and need to be addressed. Not only for EBM to progress further, but also to make it more patient-friendly and to produce valuable research which is important to future patients and not just profitable.

EBM can be used as a unifying force in medicine to bring different areas of medicine closer together by providing a solid base for the use of medicine and for shared decision making between physician and patient in order for both sides to profit from it. The solid base needs to be the available evidence which informs not only the patients and physicians but also shapes new research since evidence can never be complete.

Therefore, instead of proposing all sorts of alternatives to EBM, it should be clear that the solution to many of EBMs problems have to come from within EBM. And one such solution is to divide EBM into ‘evidence-based practice’ and ‘evidence-based research.’ In the former, the patient with his or her values, concepts, wishes and goals should be the centre of all considerations. The physician has the task to individualise all the available evidence and to use it in ways appropriate to every particular patient. In ‘evidence-based research’ patients and physicians have to be aware that their roles are changing and that they both are participating in the generating of new evidence. Individual needs and wishes do not play the active part anymore that they play in evidence-based practice.

Research evidence needs to be robust, quantifiable and reproducible. EBM uses a hierarchy to rank evidence according to these criteria, but as will become apparent, this hierarchy is only as useful as are the results of its individual ranks and the highest ranking evidence might not be the best evidence for the individual patient. However, it is necessary to take a close look at how the most high ranking evidence in evidence-based research is achieved.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





2 The methodology of evidence-based research.

2.1 RCTs and their importance for the methodology of EBM

EBM is heavily based on quantifiable research and its results with the goal to achieve qualitative health care on the population level and on the level of the individual patient. The step from the population level to the individual in health care is a fairly complicated one and will be analysed in the following chapter, because it is also the most important one to free EBM from the criticism that it is not person-centred enough. In order to do so, it is necessary to look at how quantitative results are achieved and how the actual methodology of generating ‘good’ and ‘usable’ evidence functions. The methodology of evidence-based research is based on a hierarchical view of the quality of evidence. The actual evidence-hierarchy will be explained in detail in the chapter about evidence and epistemology. This chapter will focus on randomised controlled trials (RCTs) as those medical first-in-man trials which play a major role in the generating of evidence. RCT’s are medical experiments that are designed to test safety, risk, efficacy and effectiveness, normally in this order, of novel treatments like drugs, surgery, and even acupuncture and physiotherapy.

The results of RCTs inform most medical and health policy decisions, since they are often considered to be the best option to arrive at robust and usable evidence. The following chapter will closely look at RCTs and will assess if they really are as fail-safe in producing ‘good’ evidence as they are made out to be.

On the very top of the overall hierarchy of evidence, including steps to *assess* evidence, are systematic reviews, or meta-analyses, because they pool and assess the available data, generated from RCT’s and often from well-conducted cohort and observational studies, and make it comparable. The best-known institution for systematic reviews and meta-analyses is the Cochrane Collaboration and the Cochrane Library.³⁵ Archie Cochrane could be called one of the founding fathers of EBM and epidemiology. He argued that it was vital, especially with sparse resources, that only those treatments are used which have shown beyond a doubt to be effective. He greatly favoured RCTs as a means to arrive at this goal.³⁶ Cochrane reviews are instrumental in making evidence-based practice manageable

35 Cochrane Collaboration. <http://www.cochrane.org>. Last accessed on January 23rd, 2020.

36 Cochrane, trusted evidence, informed decisions, better health. <http://www.cochrane.org/about-us/our-name>. Last accessed on January 23rd, 2020.

for the evidence user. The literature about medical evidence is so vast that it is vital to have succinct reviews which pool the most robust evidence and are able to give short and precise statements about the pros and cons of the treatments in question. Meta-analyses and systematic reviews are not without faults however, since they can only be as good as the initial data provided. The Cochrane Collaboration, and like-minded organisations, strive diligently to regularly assess new data about topics of already conducted reviews in order to keep the evidence as up to date, and as manageable, as possible. Lesser known organisations are the Joanna Briggs Institute based in Australia, the “Centre for review and dissemination” and the EPPI Centre.³⁷ Individual scientists and authors can also conduct systematic reviews and have them published. Yet again, even the amount of meta-analyses and systematic reviews about each single diagnosis is staggering and not manageable in its entirety for any physician or clinician. However, because of their special training, most clinicians can assess the relevance of the research pretty quickly and focus on the most important parts for their individual patients. Increasingly, these skills are already taught in medical schools.³⁸

In this chapter the focus lies on the methodological and ethical problems of RCT’s, since they are still viewed as the most important tool in the arsenal of conducting meaningful medical research and are therefore the cornerstones of EBM. For the longest time it was argued that the results from RCTs were the only admissible evidence in EBM.³⁹ In the last decade however, with growing criticism towards EBM and its methods, it has been acknowledged that other forms of medical evidence can play an equally valuable role, especially when and where the individual patient is concerned.^{40 41} But not all evidence is automatically ‘good’ evidence and even if it is considered to be ‘good’ the question remains: for whom?

2.1.1 *Scientific goals of RCTs*

The overall scientific goal of RCTs is to establish the absolute risk reduction, ARR, or inversely, the number-needed-to-treat, NNT, of patients with the novel treatment compared to standard or placebo. These numbers stand for the efficacy of the novel treatment.

37 Library Guides: Subjects, Services and Resources. <http://libguides.rgu.ac.uk/c.php?g=536793&p=4389919>. Last accessed on January 23rd, 2020.

38 Jorgen Nordenstrom. (2007). *Evidence-Based Medicine in Sherlock Holmes’ Footsteps*. Oxford: Blackwell Publishing.

39 Gordon Guyatt, et.al. (1992): 2420.

40 Jo Rycroft-Malone, Kate Seers, Angie Titchen et.al. (2004). “What counts as evidence in evidence-based practice?” in *Journal of Advanced Nursing*. 47(1): 81-90.

41 Trisha Greenhalgh, et.al. for the evidence-based renaissance movement. (2014).

I am using the example for NNT and ARR of the Centre for Evidence-Based Medicine Oxford here due to its conciseness: “The absolute risk reduction or absolute effect is the amount by which your therapy reduces the risk of the bad outcome. For example, if your drug reduces the risk of a bad outcome from 50 per cent to 30 per cent, the ARR is:

$$\text{ARR} = \text{CER} - \text{EER} = 0.5 - 0.3 = 0.2 \text{ (20 per cent).}^{42}$$

CER stands for Control Event Rate and EER for Experimental Event Rate.

The NNT is the inverse to the ARR and is always rounded up to the nearest number since whole patients are treated, not fractions.

$$\text{NNT} = 1/\text{ARR}$$

For the example above that means that:

$$\text{NNT} = 1/\text{ARR} = 1/0.2 = 5$$

The NNT stands for the number of patients that need to be treated in order to improve one patient. Therefore the ideal NNT is 1, as in one patient. The higher the NNT is, the less patients benefit from the novel treatment. However, even a treatment with a high NNT can be useful in particular circumstances, especially when no other treatment for this particular patient is available. It is very context dependent if a treatment with a high NNT can still be deemed acceptable. The experience of the clinician can guide the judgement call if a treatment with a high NNT might still be useful in a particular case. In trials about disease prevention, the NNT is often allowed to be higher than in trials testing singular treatments or drugs.⁴³

Additionally to the efficacy of the novel treatments, RCTs are also, and even more importantly, conducted to assess the safety of a novel treatment. This is the special focus of early phase RCTs. If the safety cannot be established, then no further RCTs will be performed. Efficacy does not trump safety! ‘Safety’ however should not be confused with ‘overall risk’ of a treatment. The risk/benefit factor needs to be established in a trial. If the actual benefit of a treatment outweighs its risk/s, then further RCTs should be conducted to assess how far they actually do so and if it is still reasonable to use a treatment even though it contains a certain amount of risk.

42 CEBM: Centre for Evidence-Based Medicine, Oxford. “Definition of ARR and NNT.” <http://www.cebm.net/2014/03/number-needed-to-treat-nnt/>. Last accessed on January 23rd, 2020.

43 Katja Suter, Matthias Briel and Judith Günther. (2015). “Number needed to treat (NNT) and Number needed to harm (NNH): weitere Abkömmlinge der Vier-Felder-Tafel.” in *Medizinische Monatszeitschrift für Pharmazeuten* 38(3): 103-106.

RCTs come in multiple forms and with various research protocols attached to them, always depending on what it is that is to be tested. Trials testing non-drug treatments are slightly different to those testing drugs, and those testing preventing measures like vaccines are yet again slightly different. However, there are some aspects that they all have in common and these will be the focus of the discussion. RCTs are so-called ‘first-in-man trials and are conducted in different phases and with a different amount of participants necessary, exactly to assess overall safety first and subsequently, efficacy, risk and effectiveness.

RCT’s are set up in a way in order to arrive at “ideal” results, i.e. results that are reliable and robust. Since it is good scientific practice never to rely on the results of single trials, most treatments are tested in multiple clinics, in multiple RCTs, sometimes with differing, sometimes with the same research protocols. The results of one RCT need to be reproducible, and preferably more than once.⁴⁴ Reproducibility of results leads to scientific acceptance, and to the results to be possibly externally valid and therefore of relevance to the actual target population. However, a multitude of trials and trial data, often running over hundreds of pages, is not usable as such for the actual evidence user.⁴⁵ That is why systematic reviews and meta-analyses are so important, because they assess the available data, dismiss badly conducted trials and make the results easier to ‘digest.’ Therefore, they are a vital part of EBM, because they make evidence-based practice possible in the first place.

If RCTs are set up correctly, they are considered to be internally valid, which means that they yield correct and rigorous results about the treatment for the actual trial population under test and have used the correct methods to do so. They are therefore comparable to an experiment in a laboratory setting, insofar as that they adhere to the same form of rigorousness. However, that internal validity does not yet make RCTs also viable for the actual target population needing that specific treatment for a that specific ailment, let alone for the individual patient. On the contrary, a very high internal validity can lead to a quite low external validity.⁴⁶ If and how this specific problem can be solved will be discussed later.

The quality of the results of RCT’s is based on the validity of the trials and the validity of the RCT’s is based, more often than not, on successful randomisation and on successful blinding. Randomisation and blinding are the two main methodological features of RCTs that set them apart from trials like cohort studies and observational studies, both of which will be evaluated as part of the hierarchy

44 J. Shao and SC. Chow. (2002). “Reproducibility probability in clinical trials.” in *Statistics in Medicine*. 21(12):1727-42.

45 Jorgen Nordenstrom. (2007):

46 Nancy Cartwright. (2007). “Are RCTs the Gold Standard?” in *BioSocieties*. 2 (1) (Special Issue: The Construction and Governance of Randomised Controlled Trials). 11-20.

discussion, and they are the two features which need to be maintained throughout the trial to really render robust results.

2.1.2 *Publication of all trial data*

In order to successfully assess RCTs and to conduct meta-analyses, the results of all RCTs need to be published in a coherent and complete form, which unfortunately they are not, at least in many cases.⁴⁷ To establish reporting standards, the CONSORT Group, a consolidated group of two research groups, both aiming at a standardisation of trial reporting, merged in 1996 and published the first CONSORT Statement in 2007 and a revised form in 2010.⁴⁸ The CONSORT statement and tool consists of a flow chart and checklist for authors of RCTs to control if all necessary information is written up and all data is included. Many academic journals prefer that the trial write-up is done according to CONSORT or even stipulate the necessity for authors to adhere to CONSORT before publishing any research results. CONSORT however does not judge the methodology of RCTs. It only aims at good reporting standards to prevent reporting bias. The value of the methodology of RCTs however lies in their set-up and in the quality of said set-up which starts with randomisation.

2.1.3 *Introduction of the basic methods of RCTs*

As the name already indicates, the most prominent and important feature of RCTs is ‘randomisation.’ Randomisation means that trial participants are divided into two different, but equal, groups. There are different methods of randomisation available and these will be discussed throughout the chapter. In its most basic, but also most usual form one group receives the treatment under trial and the other group receives either the standard treatment or a placebo. Placebos can be either inactive or active and both pose their own problems, ethical and methodological, as do placebo-controlled trials as a whole. These ‘special’ trials will be discussed at length later on.

The main question here is if randomisation really has the methodological virtue to make the groups equal and what role confounding factors play for answering that question. Blinding is the second most important step in the set-up of RCTs and means that at least two or more groups involved in the trial do not know in

47 Ben Goldacre. (2012): 81.

48 The Consort Statement. <http://www.consort-statement.org/> Last accessed on January 23rd, 2020.

which groups the participants are randomised into.⁴⁹ Blinding is supposed to control for confounders such as selection - and allocation bias, i.e. the researcher consciously or unconsciously putting participants into either the control or the treatment arm based on where they will do better. The other case which is to be avoided is that the participant knows which treatment he will receive in order to minimise expectations toward the treatment. The difficult part in any blinded trial is to maintain the blinding. At least one party can be blinded but the goal is to blind all parties involved in a trial. In trials where treatments such as pills or injections are under test, blinding of a large group of people involved is possible, even to the extent that the outcome assessors, those who receive the raw data and evaluate it, are blinded. In trials concerning surgical procedures it is much more complicated to maintain the blinding, since the surgeon at least needs to know if he is supposed to perform a real or a sham surgery. The same is true for all trials testing treatments where at least the dispensing physician or health practitioner needs to know which treatment he or she is administering. Acupuncture would be such an example where it is impossible to blind the acupuncturist.⁵⁰

2.2 The different phases of RCTs

Before a drug or treatment is tested on human beings, it is developed and rigorously tested in the laboratory. Often, in the later stages of the development, these tests and trials involve laboratory animals, such as rats or monkeys for example. The necessary requirement is most often that the DNA has a special similarity to that of humans, or that some other feature is close enough to humans to make the results of animal studies usable in later human studies. Animal studies however suffer the same problems that RCTs in humans do. They can be flawed through observation or publication bias and through the lack of external validity. It is acknowledged that they are and were necessary to further the understanding of the mechanisms of disease but are less well equipped to reliably inform about the effectiveness and safety of treatments.⁵¹

In these animal trials, randomisation and blinding is not absolutely necessary and often hard to perform because the sample sizes can be too small. It is however important to mention that in theory, and even sometimes in practice, RCTs can be performed in animal studies where the groups are sufficiently big enough and can

49 Jeremy Howick. (2011): 63.

50 Edzard Ernst and Simon Singh. (2008). *Trick or Treatment? The undeniable facts about alternative medicine*. London: Transworld Books: 67.

51 H. Bart van der Worp, David Howells, et.al. (2010). "Can Animal Models of Disease Reliably Inform Human Studies?" in PLoS Med 7(3): 1.

be meaningfully randomised and the observer could even be blinded as to the intervention, just recording the observations without attaching any results.⁵² Since 2010 the ARRIVE guidelines, as an equivalent to the CONSORT guidelines, have been adopted to make animal research and the publication of animal research more transparent and to make it liable to ethical considerations.⁵³

However, when there is talk of RCTs in the medical setting, it is necessary to assert that it is almost always based on trials involving humans. So when we talk about RCTs here, it is about phase 0 to IV trials, those following the laboratory stage of the research, involving human beings. The most common RCTs are performed in phases I to III, phases 0 and IV are rarely performed. However, especially phase IV trials, those which are conducted after the market approval of the treatment, are of vital importance for the safety assessment of a drug, because they can be conducted over a lengthy period of time, in many patients of the actual target population, and side effects can be more easily detected.⁵⁴

2.2.1 Phases of RCTs according to Benedetti.⁵⁵

Phase 0 trial: These trials are fairly rare, mostly conducted in cancer research and are only involving a very small number of participants who are usually suffering from the disease. The aim of phase 0 trials is to establish safety and the potential of the drug to reach the target area depending on the dosage. Therefore it is essential that the participants are suffering from the disease in question. The novel treatment is rarely tested against a placebo.

Phase I trials: These are for most drugs and treatments the actual first-in-man trials. They are conducted, usually, on a very small number of healthy participants. The main questions are ‘safety of dosage’, possible side effects, and the bodies reaction to the drug. The novel treatment is rarely tested against placebo.

Phase II trials: These trials try to answer the same questions as in a phase I trial, but are conducted with a larger number of participants, often patients with the actual disease or illness in question. Since a base-safety of the drug is established after successful phase I trials, efficacy of the drug plays a bigger role than safety.

52 Beverly Muhlhausler and Frank Bloomfield, et.al. (2013). “Whole Animal Experiments Should Be More Like Human Randomized Controlled Trials.” in *PLoS Biology*, 11(2):. 1.

53 Beverly Muhlhausler and Frank Bloomfield, et.al. (2013).

54 Fabrizio Benedetti. (2014). *Placebo Effects*. 2nd Ed. Oxford: University of Oxford Press, Kindle Version: Chapter 1.2.1 Placebos are the tenet of the randomised, double-blind, placebo-controlled trial design.

55 Fabrizio Benedetti. (2014): Chapter 1.2.1 Placebos are the tenet of the randomised, double-blind, placebo-controlled trial design.

Phase III trials: These are the last trials before a possible licensing and market approval of the drug. They are preferably conducted with a huge number of participants and the novel treatment should be tested against the standard treatment, and only against placebo if no standard is available. Safety, efficacy and effectiveness of the novel treatment are again the main questions for the researchers. Side effects are observed over some length of time.

Phase IV trials: These trials are conducted after the licensing and market approval of the novel treatment. They are most often observational studies which gather information of the long-term risk and safety of the drug as it hits its actual target population and therefore play a part in establishing external validity. Unfortunately phase IV trials are fairly rare, even though they could be conducted, even as RCTs, quite easily, if GPs and clinicians were allowed to gather the appropriate data in their daily practice.⁵⁶

The trials as they are described above are idealised versions of RCTs. The number of participants per trial, especially if it is a small number, contributes in a big way to the problems of RCTs. Phase II and III trials require, to be performed ‘correctly,’ a fairly large number of participants. The more participants the better, since the stratification of possible confounders, factors that influence the trial results, is better guaranteed and even rare side-effects are picked up more easily, *if* they are picked up at all in the given time. The obvious problem is the recruitment of this larger number of participants for many different reasons which will be discussed throughout the chapter. Sometimes the disease is simply so rare that there are just not as many patients available as a trial would actually need. In these cases provisions are done for smaller trials and the smaller number of participants is not interpreted as a flaw of the trial.⁵⁷ In most cases however there are just not enough volunteers. Either because patients are not made aware that a trial is conducted for which they would be eligible or they are not ‘ideal’ enough, which means they are too ill or have too many co-morbidities. Or they simply do not want to participate in a trial, fearing that they receive the lesser treatment or lesser care.⁵⁸ However, in many trials the care is even better in the research setting than it would be in the regular setting, since the participants are under close supervision and are controlled much more regularly than a patient in a normal GP practice would be. That is especially true, and alarmingly so, in trials that are conducted in the developing

56 Ben Goldacre. (2012): 225-241.

57 Niklas Juth (2014). “For the Sake of Justice: Should We Prioritize Rare Diseases?” in Health Care Annals.

58 Yvonne Brandberg, Hemming Johansson, Mia Bergenmar. (2016). “Patients’ knowledge and perceived understanding —Associations with consenting to participate in cancer clinical trials.” in Contemporary Clinical Trials Communications 2: 6-11.

world. Trials are less expensive to conduct there and participants are presumably easier to find, because the level of overall care is so bad that research settings provide the one possibility for many patients to even have some form of treatment.⁵⁹

2.3 Confounding factors and their influence on trial results

Confounding factors are those factors, like age, gender, overall health, weight, pregnancy, existing illnesses, chronic diseases, etc., but also many forms of bias, that can severely change the outcome of a trial. The attempt to control for confounding factors is what makes randomisation and blinding so important for medical trials. Confounding factors are divided into known and unknown confounders and both types can differ significantly from trial to trial, depending on the treatment or drug under test.⁶⁰ Therefore, there is no exhaustive list of possible confounders and every investigator has to conclude from previous research which confounders are relevant for the trial and subsequently need to be controlled for. This before-hand control however can only work for known confounders. Unknown confounders make control mechanisms, such as randomisation and blinding even more necessary.

Howick explains confounding factors by labelling them with three properties. First: “the factor potentially affects the outcome.” Second: “the factor is unequally distributed between experimental and control group” and third: “the factor is unrelated to the experimental intervention.”⁶¹ The most important feature of confounding factors is that “each confounding factor provides a potential alternative explanation for the results of a clinical trial.”⁶² This chapter will therefore aim to explain why the control for possible confounders is on the one hand important for the overall validity of the test results and therefore desirable and on the other hand not always possible and not always necessary, even if the above explanation of confounding factors is correct.

Sometimes, mostly in phase IV trials, does it become apparent that confounding factors were responsible for the overall outcome of previous trials. In some cases the deviation between the results of a drug or treatment before market approval and afterwards is so big that the drug will be, and needs to be, taken off the

59 Sonia Shah. (2008). *Am Menschen Getestet! Wie die Pharmaindustrie die Ärmsten der Welt für Medikamententests missbraucht*. München: Redline Wirtschaft.

60 Jeremy Howick. (2011): 34.

61 Jeremy Howick. (2011): 77. Howick modifies this third property with regard to side effects. In the moment I will let it stand as is and will come to the point about side effects later. “the confounding factor must be unrelated to the positive characteristic effects of the experimental treatment on the target disorder (as opposed to side effects).”

62 Jeremy Howick. (2011): 36.

market. A well-known example of such a case, albeit without a phase IV trial was running, is the thalidomide scandal. Thalidomide was a drug manufactured in the 1950s and 1960s in Germany and given to many expecting mothers to treat nausea and fatigue.⁶³ After the drug was on the market for quite some time it was discovered that it caused malformation of the limbs, called phocomelia, in newborns. Thalidomide was subsequently taken off the market, but only after a huge scandal ensued in Europe, the United States and Canada.⁶⁴ It is now in use again for certain types of cancer and for rare complications in leprosy, but pregnant women are specifically warned not to use the drug.⁶⁵ The Thalidomide scandal has influenced, and changed, how drugs and treatments are tested today and how the marketing approval of new drugs is governed to prevent similar ‘mistakes’ in the future.

Medical scandals show how necessary it is to test the safety of a new treatment or drug, long before it reaches the open market, and to stay diligent even after it is approved. And in order to securely arrive at a conclusion about safety, all the other factors that could influence the outcome must be eliminated in the best possible way. That makes it so important to control for all confounders, known and unknown, as good as possible. As will become obvious, randomisation and blinding are far from perfect to always and reliably control for unknown confounders and many scientists seem to agree about their imperfectness without presenting valid alternatives. Therefore it is important to study the shortcomings of these control mechanisms and to try and improve them, instead of vilifying RCT’s in general, as some seem to do.⁶⁶

So in order for RCT’s to be internally valid and to arrive at “ideal” results, they have to control for confounders and use randomisation and blinding to do so.⁶⁷ The two methods go hand in hand with each other but need to be examined somewhat separately because they each have their inherent problems and strength.

63 Imogen Evans, Hazel Thornton et.al. (2011): 4.

64 Jack Botting. (2002). “The History of Thalidomide.” in *Drug News & Perspectives*. 15(9): 604.

65 Teru Hideshima, Dharminder Chauhan et. al. (2000). “Thalidomide and its analogs overcome drug resistance of human multiple myeloma cells to conventional therapy.” in *Blood*, Vol. 96:2943-2950.

66 John Worrall. (2010). “Do we need some large, simple randomized trials in medicine?” in M. Suarez, M. Dorato and M. Redei (eds). *EPSA Philosophical Issues in the Sciences*. Dordrecht: Springer.

67 Jeremy Howick and some other philosophers of science would prefer to call “blinding” masking, because blinding is a derogatory term towards visually impaired patients. Since most articles and books about RCT’s still use the term “blinding”, that is what I will use, without any derogatory intent.

2.3.1 Randomisation

Randomisation will be explained first, since it is understood as playing the most important role in the control for possible confounders.⁶⁸ Randomisation in its most simple form means that patients who are willing to participate in a trial are selected into two groups, the treatment group that receives the novel drug or treatment and the control group that receives the standard treatment or a placebo. The selection process is randomised, most often according to a numerical pre-specified code that is not known to the researcher or the participant. Therefore, neither know in which treatment arm the participant is randomised into. Eligible participants are most often randomised via an independent agency which has produced the randomisation codes. The researcher receives a number for the patient that is matched either for control or for treatment. The random numbers are generated via a computer program and coded before they are given out. Although it sounds very complicated, the method is very close to throwing a dice or flipping a coin.⁶⁹ The main reason that computer generated numbers are used is that there is less of a possibility to manipulate the process. The method of computer based randomisation can be compared to a lottery. Most often in practice, after the computer-generated lottery has run, envelopes are prepared which “contain randomly generated instructions about which group to assign the next patient.”⁷⁰ Here blinding and randomisation go hand in hand. The researcher receives the envelope but since he does not have sufficient information to decipher the code, he is blinded to the intervention that the participant is randomised into.

There are a number of softwares available on the internet which are fairly easy to use and can be used for multiple types of randomisation. One of those often used by medical statisticians is <http://www.graphpad.com/quickcalcs/index.cfm>. Graphpad even goes beyond that and offers different statistical methods to arrive at and calculate results.

In EBM both single and cluster randomisation can be used.⁷¹ In evidence-based policy (EBP), cluster randomisation is most often the method of choice, since policies are never implemented on an individual level. Cluster randomisation means that entire groups of people are randomised as groups, not divided into individuals. Single randomisation simply means randomisation at the individual patient level, with the methods described above. In EBM cluster randomisation could

68 John Worrall. (2004). “Why there’s no cause to randomize.” Technical Report in Causality: Metaphysics and Methods. CPNSS: 3.

69 K.P. Suresh. (2011). “An overview of randomization techniques: An unbiased assessment of outcome in clinical research.” in *Journal for Human Reproductive Sciences*. 4(1): 9.

70 Jeremy Howick. (2011): 43.

71 BetterEvaluation: Sharing information to improve evaluation. <http://betterevaluation.org>. Last accessed on January 23rd, 2020.

for example mean that the participating clinics would be randomised, and the patients would be treated with either the novel or the standard treatment, depending in which clinic they would be treated at. Another form of cluster randomisation is also called block randomisation and is sometimes used when it is important to have equal sample-sizes. Block-randomisation however is not very good in eliminating confounding factors.⁷² Cluster randomisation can achieve that, if enough patients in each clinic participate in the trial.

To eliminate known confounders before the trial, especially when the number of participants is high, stratified randomisation can be used. It starts with a type of block randomisation in which participants are assigned to different blocks depending on the confounder. “After all subjects have been identified and assigned into blocks, simple randomization is performed within each block to assign subjects to one of the groups.”⁷³ This method however only works when all participants are selected before the trial. If they are selected on a continuous basis while the trial is running, then regular single randomisation should be used.⁷⁴ Cases in which stratification on the individual level is asked for are normally involving certain aspects of a specific disease. An example here can be any cancer that can appear with or without metastases. A novel treatment can yield positive results in both cases, but the participants should be randomised according to the presence or absence of metastases, so that the control and the treatment arm contain a fair number of both types of cancer patients. Therefore both ‘types’ of cancer are used as strata and the participants within the strata are then randomised into control and treatment arms.⁷⁵ However, an approach like this is obviously only workable with very few strata, otherwise the subgroups would be too small. Machin claims that “For continuous prognostic variables such as age, stratification can only be carried out when these variables are divided into categories. [blocks] ... Although age (or some other continuous variable) may be prognostic for outcome, it is usually preferable not to stratify for this but to record the information for each patient and take account of this in a retrospective sense at the analysis stage.”⁷⁶ However ‘controlling’ for confounders during the analysis stage, when all the data is gathered, can be very complicated and can lead to false results. Therefore it would be better to use some pre-stratification and thorough randomisation in order to control for known confounders right from the beginning of the trial.

72 K.P. Suresh. (2011): 10.

73 K.P. Suresh. (2011): 10.

74 BetterEvaluation Pg. 2.

75 David Machin and Peter M. Fayers. (2010). *Randomized Controlled Trials, Design, Practice and Reporting*. Oxford: Wiley-Blackwell: 101.

76 David Machin and Peter M. Fayers. (2010): 102.

In systematic reviews, all types of randomisation can be compared, but it should be specified which type of randomisation was used in order to better evaluate the outcome. An example where specification was necessary is a meta-analysis of trials in low- and high-income countries about antenatal care for pregnant women. The women in the high-income countries were randomised individually and the women in the low-income countries were randomised in clusters. The trials were set-up to evaluate how much care was necessary for women with a low-risk pregnancy and if the number of antenatal visits could be decreased.⁷⁷ One main reason between the differences in randomisation in this example is the difference between high- and low-income countries. In high-income countries, ‘contamination’ is less likely. ‘Contamination’ can occur when participants communicate with each other and find out in which arm they belong and what they potentially miss out on. Contamination can lead to participants leaving trials and seeking for a higher level of care elsewhere. Even though pregnant women speak with each other and might compare their levels of care in high-income countries, the overall level of care is high and being in the test arm of a trial, especially in this case, does not pose a greater risk to either mother or child, as was shown in the sub-group analyses of high-income countries in which no more perinatal death occurred than in standard care.⁷⁸ In low-income countries the risk of contamination was deemed higher. Cluster randomisation of the clinics was used to prevent this, since all patients in one clinic are either treated with the novel or the standard treatment. However, possible hidden ‘individual’ confounders are not controlled for in cluster randomisation. Therefore, it is impossible to directly compare trials with individual and with cluster randomisation in most cases, simply because possible confounders cannot be equally controlled for. One obvious confounder in the example is already the difference in income and overall care given. Perinatal death were higher in the test-arm in low income countries, because conditions which would have led to the admission into neonatal intensive care were not recognised at all or too late. One outcome of the trials and the subsequent meta-analysis was that all women were less satisfied with fewer prenatal visits, regardless of low- or high-income countries.

Even single randomisation can be problematic, if it is not done correctly, since the different methods of single randomisation, if they are not done as described above, can be easily manipulated. One very questionable method is the Zelen method of randomisation, because participants are only asked for their consent to participate in a trial after the randomisation has already happened.⁷⁹ The

77 T. Dowswell and G. Carroli. (2010). “Alternative versus standard packages of antenatal care for low-risk pregnancy.” in Cochrane Database of Systematic Reviews, Issue 10. Art.

78 T. Dowswell and G. Carroli. (2010).

79 Marvin Zelen. (1979). “A new design for randomized clinical trials.” In *New England Journal of Medicine* 300: 1242-1245.

method can lead to obvious selection bias and to early drop-out rates if and when the participants are questioning the trial and opt out to use the standard treatment. A choice denied to those patients in the control arm of the trial, because there were not informed at all about their trial participation. When Zelen wrote his randomisation proposal, such an approach would still have been possible. Today however, participants have to consent to the use of their data. Therefore they need to be informed about all the possible choices and the Zelen method is ethically and methodologically problematic. The latter is the case because neither the participant nor the researcher are blinded as to the intervention which can lead to selection and observation bias. The former because participants in the control arm are left in the dark as to the treatment possibilities.

“Randomised plays the winner” is another form that was and still is used to randomise participants on an ongoing basis.⁸⁰ It functions in a way that if one treatment is more effective than the other, then more participants will receive the “effective” treatment and this can result in false-negative or false-positive overall results of the trial, because the patients are not divided equally anymore between the arms and confounders and other deviations in the participant population can severely change the results.

Howick points out the problem of “pseudo-randomisation” which occurs when a randomisation procedure is too easy to decipher. An example would be if every other patient, who would be eligible for a trial, would be put in the control arm. If this type of randomisation would be used, blinding of the different parties could be easily subverted.⁸¹ Howick makes the argument that knowledge about the allocation does not need to undermine the validity of the trial, if the randomisation is not tampered with.⁸² Further down the argument, he however admits that “allocation bias and self-selection bias can become worrisome again.”⁸³ Howick is convinced that proper blinding is simply not possible and should therefore not have as much methodological value as Worrall and others are allowing it. However, blinding seems to be the only method to successfully rule out self-selection and allocation bias. I will come back to this problem later in the chapter.

80 Elbourne, D. Field, D. Mugford, M. (2002). “Extracorporeal membrane oxygenation for severe respiratory failure in newborn infants (Review)”. The Cochrane Collaboration. Wiley.

81 John Worrall. (2003). “What Evidence in Evidence-Based Medicine.” in *Causality: Metaphysics and Methods*. CPNSS: 14.

82 Jeremy Howick. (2011): 44.

83 Jeremy Howick. (2011): 44.

2.3.2 *Confounding factors during trials:*

If, despite all efforts to the contrary, confounders do appear during the trial which are influencing the results, the trial needs to be changed in some way.

An obvious solution would be re-randomisation or stratification, taking the new known confounder into account. Re-randomisation however is hardly ever discussed since the option of choice is to match the results post trial or to abandon the trial altogether. Since re-randomisation sounds like the most common-sense approach, I will briefly discuss the reasons why it is not a viable solution to the problem. In re-randomisation it could happen that the previously gathered data cannot be used anymore. The whole trial would have to be organised again, preferably with new participants and a new protocol. If the same participants would be re-randomised, a wash-out period would be needed, since almost all drugs stay in the body for a certain amount of time. Additionally, the treatment or drug would have already shown benefits, that for example in a placebo-controlled trial only appeared in the treatment arm. Blinding would be nearly impossible to maintain. To add new participants to the ‘old’ participants would be equally complicated, even if a wash-out period is maintained, because the reactions to the new treatment might be different and again blinding would not be maintained. The participant base is therefore too contaminated to successfully extrapolate useful research results.

Another problem of re-randomising is that it is time-consuming and therefore costly. RCTs however are extremely costly to begin with and the number of participants, regardless of how many there are, is usually too small anyway. Research hospitals and clinics have provided facilities and staff for a certain amount of time to run a trial and in the case of re-randomisation all that would be needed again, again at substantial costs to the research facility. Since in many RCTs it is unsure if the trial leads to positive results, and since there is always a risk that it does not, re-randomisation is simply too expensive to consider, even if it would be scientifically the best way to go about the problem of controlling mid-trial for unknown confounders. In actual clinical practice the least time- and money consuming solution to the problem is to factor the now-known confounders in post-trial. The method most often used to account for confounders after a trial is the ‘analysis of covariance method’ also called ANCOVA.⁸⁴ However, ANCOVA only works reliably if the groups are homogenous. The method however does not work in every case. Sometimes it can only demonstrate the problem. And the same is true for

84 Medcalc. Easy to use statistical software. Software provider for statistical methods in medical research. <https://www.medcalc.org/manual/analysis-of-covariance.php>. Last accessed on January 23rd, 2020.

almost all statistical models in use today.⁸⁵ Therefore it is important that randomisation works in the first place without having to result to statistical means after the trial.

It is a simple mathematical truth however that very large and diverse trials are the ones which control the best for most possible confounders, since there are simply enough participants to equal out the groups.⁸⁶ In small trials, and many trials are small, which means too small for randomisation to do its entire magic, confounders are, almost necessarily, unequally distributed, even if some sort of stratification has taken place. This is also the main argument John Worrall uses against randomisation. However, as Jeremy Howick rightly points out, the supporters of EBM and randomisation do not make that strong a claim that randomisation is the only control. Randomisation is an important tool in the set-up of trials, but not the only one to maintain internal validity. And many critics, including Worrall, do not provide a sufficient alternative to randomisation, save then saying that confounders can be controlled for statistically once a trial is finished.⁸⁷

2.3.3 *Blinding — its problems and its virtues*

Blinding is used to control for many forms of bias, many of which can severely alter the results of every trial. Different possible biases are:

1. selection bias: the allocation of participants is done depending on their possible receptiveness to the treatment. Allocation concealment is therefore necessary and can be achieved through blinding.
2. performance bias: participants and personnel can both influence the results of the treatment if they know what treatment they are receiving. At least those two groups need to be blinded to avoid performance bias.
3. detection bias: Results can be interpreted in certain ways and made to look a certain way if and when the ‘outcome assessor’ is unblinded and knows if the novel treatment or the standard is to be assessed.
4. reporting bias: cannot be controlled for through blinding, but only through rigorous enforcement of the rule that all research findings need to be published, the positive as well as the negative ones. Preferably even the raw data is published, or at least easily accessible for researchers.

85 K. P. Suresh. (2011): 8-11.

86 Jeremy Howick. (2011): 50. Howick and many others have already pointed that out. It is almost impossible to create a trial that is sufficiently large enough to control for all confounders. That is why in the set-up for most trials, in-trial check-ups are mandatory to see if the distribution is still equal.

87 John Worrall. (2003):

Many authors have defined more biases, but these four are the most important ones because they are the ones that have to the most influence on the actual trial results.

Randomisation without blinding can only control for the initial selection-bias if and when the researcher is prevented through randomisation to decide to which treatment arm the participant is allocated.⁸⁸ All further forms of selection-bias and especially the selection-bias on part of the participant are only avoided if at least the researcher and the participant are blinded. “Six groups involved in a trial are sometimes blinded, namely participants, caregivers, data collectors, outcome evaluators, statisticians, and manuscript authors.”⁸⁹ However it is often not defined or explained in research protocols and post-trial write-ups which of the six possible groups was actually blinded.⁹⁰ All groups can be filled in by different people or one person can fill in almost all of the roles. In the write-up of an RCT it should be made clear not only which groups were blinded, but also how many different people or groups there were to blind, and if the blinding could be maintained for the duration of the trial.

Howick for example groups together the researchers and those who are dispensing the treatment into “caregivers” but separately names “data collectors” and “outcome evaluators.” It seems to be an arbitrary decision and one that every author and researcher can make for themselves. The lack of a reporting standard and the lack of clear definitions of terms can make the results of RCTs less robust and therefore vulnerable to manipulation.

The most often used method of blinding is double-blinding. This means that at least participants and researcher are blinded. If Howick’s terminology is assumed, researcher means caregiver. ‘Caregiver’ can include the acting physician and possibly the nurse who might dispense the treatment. In order to successfully maintain the blinding however it should be at least triple-blinding, including the participants, those who dispense the medication and the researcher. The latter should be blinded regardless if he or she is having an active part in the trial, or is more on the sidelines, organising the trial and assessing the results.

After a blinded randomisation, blinding has to be maintained during the trial. The randomisation codes should not be revealed and all possible treatments, be it the new, the standard, or placebo has to be sufficiently alike to not be distinguishable from each other. There are different methods available to achieve this, depending on the type of the treatment. Oral treatments, as in pills or liquids, can be manufactured in a way that the control treatment looks, smells and tastes exactly like the treatment under test. This is true for both placebo-controlled trials and

88 John Worall. (2004): 2.

89 Jeremy Howick. (2011): 65.

90 Isabelle Boutron and David Sackett, et.al. (2006). “Methods of Blinding in Reports of Randomized Controlled Trials Assessing Pharmacologic Treatments: A Systematic Review.” *PLoS Medicine*, Volume3, Issue 10: 1923.

active controlled trials, where the novel treatment is tested against the accepted standard. Taste and smell can be masked by strong flavours such as peppermint or simply sugar. Food colouring works as well, or the use of gelatine capsules which all look the same from the outside and have the same texture.⁹¹ There are many methods available to make the different treatments, standard, novel and placebo to look and taste alike and the whole scope of possibilities is usually used to successfully blind all those who are involved in the trial. Treatments which have to be administered intravenously or intramuscular or any other way rather than oral, can also be manipulated in such a way as to mimic the treatment under test.

It can be hard to maintain the blinding, especially in cases where the intervention under test yields dramatic results, either positive or negative, very early on. For some philosophers of science dealing with EBM such as John Worrall or Jeremy Howick, the fact that the blinding can be hard to maintain and easy to subvert is a real problem. However, I believe that especially in those cases where there are dramatic results, the occurring unblinding can be easily dealt with. And sometimes it is even necessary to quickly and effectively treat the participants, especially when the effects are dramatic and not in a positive, but in a negative way, as for example in the TGN1412 trial of 2006.⁹² In the TGN1412 phase I trial, eight healthy male participants were randomised to receive either TGN1412 or a placebo. Six received the active treatment and deteriorated very quickly due to a cytokine storm, a condition in which the entire immune system and consequently the organs are shutting down very quickly.⁹³ Accordingly it was immediately obvious which two participants had received the placebo. All six participants who had received the active treatment were subsequently treated in intensive care and survived, but they did so with lasting repercussions to their health and well-being. The most dangerous ‘flaw’ that happened during the trial was that the initial dose was too high and that the intervals in which the drug was administered were too short. The waiting period between the single administrations of the drug in each individual participant and the intervals between administering the drug to the next participants should have been significantly longer.⁹⁴

However, it is important to realise that especially phase I trials have the potential to be dangerous for the participants and are therefore very tightly controlled. They are only approved if a maximum of safety can be guaranteed. The unblinding in the example occurred quickly and without the possibility of avoiding it, much to the safety and security of the participants who could be treated fairly quickly.

91 Isabelle Boutron and David Sackett. (2006): 1935.

92 Michael Goodyear. (2006). “Learning from the TGN1412 trial.” in *BMJ*: 1-2.

93 E. William St. Clair. (2008). “The calm after the cytokine storm: lessons from the TGN1412 trial” in *Journal of Clinical Investigation*. 118(4): 1344–1347.

94 E. William St. Clair. (2008): 1344-1347.

A further problem which Howick mentions in connection to blinding is the Philip's paradox which claims that many dramatically successful interventions are not supported by best evidence, if one follows the claim that best evidence can only be attained through successfully blinded and randomised RCTs.⁹⁵ The usual examples for such interventions are the Heimlich manoeuvre to unblock a closed airway of a choking person or the removal of the appendix in a patient with acute appendicitis. However, it seems as if the Philip's paradox is not a real problem for medical research. Nobody would propose a non-surgical option for acute appendicitis and similar surgeries. It sometimes seems as if EBM proponents claim that research always starts at the bottom and has to work its way up. Instead, 76 - 96% of treatments used today are actually proven to be effective and can be used as comparison, or base, for novel treatments.⁹⁶ And since it is less ethically questionable to compare a novel treatment against the standard treatment, instead of against a placebo, it is actually of great value that the available 'standard' treatments are that effective.

2.3.4 *Blinding in surgical trials*

Surgical trials are a special area of medical testing, because some important tools that make RCTs so internally valid are difficult to maintain during a surgery. Blinding would be one of these tools. Among all the groups which can possibly be blinded, at least the performing surgeon needs to know which surgery he or she is performing.⁹⁷ There are two surgical trial options available however in which at least the patients and most other groups can be blinded. Either the novel surgery is compared against the standard form of surgery, or it is compared against a form of 'placebo' surgery. The latter is a so-called 'sham' surgery. 'Sham' surgeries mean that the patient is prepped as for a real surgery. A small incision is made and subsequently looks like a real wound, just that it is not as deep and the resulting scar might be less dramatic. Since the patient still has to receive anaesthesia, there is a certain risk of complications involved without any benefit at all, especially when 'general anaesthesia' is used. And even small incisions can lead to scars and sensitive tissue around the surgical area and hence, in the worst case, to longterm problems. So there is, even when sham surgeries are used, no entirely safe way to perform surgical trials. These trials are often performed with the overall goal to make a surgery as minimally invasive as possible or to reduce the time and amount

95 Jeremy Howick. (2011): 64.

96 Jeremy Howick. (2009). "Questioning the Methodological Superiority of 'Placebo' over 'Active' Controlled Trials." in *The American Journal of Bioethics*, 9(9): 33-48.

97 Paul Karanicolas, Forough Farrokhyar, and Mohit Bhandari. (2010). "Blinding: Who, what, when, why, how?" in *Canadian Journal for Surgery*, 53(5): 345-348.

of anaesthesia. Anaesthesia itself is very risky and can even be life-threatening. Recently researchers are focusing more on the connections between anaesthesia and amnesia and even dementia.⁹⁸ Anaesthesia for major surgery can, and often does, lead to some type of temporary memory loss. Most patients experience that memory loss but do not suffer any further consequences because it is such a short-term phenomenon. In rare cases when the patient wakes up during surgery, the intrinsic memories of waking up, irrelevant if the patient actively remembers or not, can lead to post traumatic stress disorder and therefore to longterm harm for the patient.⁹⁹ Current research also investigates the risk of dementia caused by anaesthesia for older, not necessarily only elderly, patients. So far there are not enough research results to back-up the hypothesis of a connection, but that does not mean that there is no risk.¹⁰⁰

The use of ‘sham’ surgeries is also regulated by the Declaration of Helsinki, since ‘sham’ surgeries are equal to the use of placebo, albeit having a higher risk. The Declaration of Helsinki states that the use of placebo is allowable “where for compelling and scientifically sound methodological reasons the use of any intervention less effective than the best proven one, the use of placebo, or no intervention is necessary to determine the efficacy or safety of an intervention and the patients who receive any intervention less effective than the best proven one, placebo, or no intervention *will not be subjected to additional risks of serious or irreversible harm as a result of not receiving the best proven intervention. Extreme care must be taken to avoid abuse of this option.*”¹⁰¹ (My emphasis). Following the Declaration of Helsinki, it is advisable to compare new surgical methods against the standard. That the blinding here cannot be fully maintained has to be accepted.

2.4 Placebo and the placebo response

Placebo means in Latin “I will please”. Placebos are pills or injections that look and smell like a treatment with an active ingredient but are actually made either from sucrose or lactose or come in the form of a saline solution as an injection.

98 V. Fodale and L.B. Santamaria, et.al. (2010). “Anaesthetics and postoperative cognitive dysfunction: a pathological mechanism mimicking Alzheimer’s disease.” in *Anaesthesia*, 65: 388–395.

99 Walter Glannon. (2014). “Anaesthesia, amnesia and harm.” in *Journal of Medical Ethics*. 0:1-7.

100 Alzheimer’s Research UK. “General anaesthesia linked to increased dementia risk.” Published online May 31st, 2013. <http://www.alzheimersresearchuk.org/general-anaesthesia-linked-to-increased-dementia-risk/>. Last accessed October 15th, 2017.

101 World Medical Association, The. Declaration of Helsinki. Adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964. Lastly editorially revised by the 64th WMA General Assembly, Fortaleza, Brazil, October 2013. Article 33, Use of Placebo.

Placebo pills can be manufactured and manipulated in many ways to mimic an actual treatment, but usually do not contain any active ingredients. Placebos can however trigger a special effect in the patient, despite their lack of active ingredients. “The placebo effect arises out of the patient’s confidence in the treatment (of the physician).”¹⁰² It can be triggered by the physician, if he or she can convince the patient that the offered treatment will be successful. These instances of the use of the placebo-effect are linked to the phenomenon of ‘classical conditioning’ and refer back to the physiologist Ivan Pavlov who discovered that his laboratory dogs would salivate if he, or one of his assistants, would enter the room to feed the dogs. After a while they would salivate even if one or the other would arrive without food. The dogs were conditioned to react in that way, regardless of the presence or absence of food. In the same way a human being can be conditioned to believe in an effect, even if there is no actual effect to be measured, as in the use of placebo.¹⁰³

The American anaesthetist Henry Beecher used saline injections on soldiers during World War I when morphine was running low. He discovered that they worked, especially when he told the soldiers that they received a powerful pain medication.¹⁰⁴ Again, the placebo-response here refers back to some type of conditioning which can be very strong. In the instance of the soldiers receiving the injection, the use of placebo was ethically correct because it had, due to the conditioning, a positive effect on the patient and there was no active treatment available. However, successful placebo effects can lure the patient into a false sense of security, since only the symptoms are addressed, not the underlying disease.

Testing active treatments against placebo factors the placebo effect for both into the results. Every active medication contains both, the active effect and the placebo effect.¹⁰⁵ However the placebo effect is not present in every patient and it is certainly not the same for every patient.¹⁰⁶ Therefore it would be a logical fallacy to attribute a 100% occurrence to the placebo effect. Still, it needs to be taken into account. When a new active ingredient is tested against the standard treatment, the placebo effect is also factored in, because it is potentially present in both cases.

Since the placebo effect is unspecific and can vary from patient to patient, it is difficult to filter it out for correct assessment. Some participants do not react at all to the placebo, some might react overly strong. And there is always the possi-

102 Edzard Ernst and Simon Singh. (2008):57.

103 Ivan P. Pavlov. (2003). *Conditioned Reflexes*. New York: Dover Publications. [Book is a compilation of his writings over time.]

104 Henry K. Beecher. (1955). “The Powerful Placebo.” in JAMA: 1602-1606.

105 Edzard Ernst and Simon Singh. (2008): 65.

106 Natalie Grams. (2015). *Homöopathie neu gedacht. Was Patienten wirklich hilft*. Berlin, Heidelberg: Springer Verlag: 133.

bility that the actual placebo effect is confused with other naturally occurring effects of the disease. Regression to the mean or fluctuation of symptoms can be mistaken for, and reported as, a placebo-effect, without the possibility of clarifying which is which.¹⁰⁷

2.4.1 *Passive versus active placebo*

To make matters even more complex, placebos can come in two different forms, active placebo and inactive placebo. Inactive or pure placebos are those that mimic the active treatment in looks, smell and taste but contain no active ingredients. Active placebos are supposed to mimic all of the above and additionally contain active ingredients to bring only the side-effects about.¹⁰⁸ If the side-effects are known or suspected beforehand and can be included in the placebo, then in the case of positive side-effects the inclusion might be at best beneficial and at worst without any consequences at all, since not all possible side-effects have to appear in every patient. The case is different with regard to negative side-effects. In some treatments, the negative side-effects are well known and accepted because the overall outcome of the treatment is beneficial and the negative side-effects are just temporary. Chemotherapy is an example in which harmful side-effects are taken for granted, but are accepted because the benefits outweigh the harms done by the treatment. The same is true for every surgery, in which bodily harm is inflicted for the purpose of healing the patient.

To produce a placebo with harmful side-effects, but without any positive treatment effects, can potentially pose a real risk to the health of the participant that is not outweighed by some positive outcome. These placebos are also called nocebos and they can either be manufactured or the nocebo-effect appears in the patient, although an inactive or an active placebo has been given.¹⁰⁹ “Nocebo” again is latin for ‘I will harm.’ Since risks should be minimised in trial settings, active placebos are highly problematic and their use is ethically challengeable. In cases where an active placebo would be needed, because otherwise the blinding is impossible to maintain due to the lack of side-effects in the control arm, it would be advisable to opt for the standard treatment as a possible control instead. Only if there is no standard treatment available should an active placebo be used.

107 Damien G Finnis, Ted J Kaptchuk, Franklin Miller, Fabrizio Benedetti. (2010). “Biological, clinical, and ethical advances of placebo effects.” in *The Lancet*; 357: 686-95

108 Isabelle Boutron and David Sackett. (2006): 1938.

109 Herbert Benson. (1997). “The Nocebo Effect: History and Physiology.” in *Preventive Medicine*, 26 (5): 612-615.

However, even inactive placebos are not without problems. They might suffer from a misnomer inasmuch as that there is no such thing as an inactive pill. Placebos contain something and they trigger a response, and even if it is just the bodies response to a small dose of sugar. Can something that does something ever be called inactive?¹¹⁰

Some proponents of placebo-controlled trials are adamant that the side-effects should be prevalent in the placebo, since it would be impossible to keep the blinding if the treatment arm noted side-effects and the control arm does not.¹¹¹ Again, due to the placebo effect however, participants in the placebo arm might report side-effects. It is not as clear-cut a phenomenon as some of the placebo-trial proponents like it to be. The knowledge and subsequent sharing of any side-effects can lead to contamination and can therefore invalidate the results of the trial. The question remains if the risk of contaminating the results is really higher than the risks inherent to the actual side-effects of an active placebo? The reason why I estimate the risk of the active placebo higher is that it is a direct risk to the participant and one which cannot be avoided if active placebos are used. The risk of contamination can be avoided by separating the participants of the two treatment arms whenever it is possible. In other cases, when participants are bound to meet in the clinic or the doctor's office, what I would call patient to patient contamination, could be avoided by putting a clause in the initial agreement that forbids to talk about any symptoms of the medication. Allegedly that approach does not help against possible selection bias on the part of the researcher when the obvious absence of known side-effects jeopardises or even violates the blinding. However, a trial can still yield successful results, even if the blinding is terminated prematurely, as in fact happens in many trials.¹¹²

2.5 Stopping clauses

The 'ideal' running of a trial would entail that it runs until set endpoints are reached. The timing is often pre-calculated, since trials are expensive and cannot run indefinitely, due to lack of funds and willing participants. As we have seen before, trials are run to either test for safety or efficacy or both. Every trial however can be stopped prematurely, if and when necessary, as was for example the TGN1412 trial. In order to do so, many trials have stopping clauses implemented in their protocol. Stopping clauses define when a trial can be stopped for either safety or efficacy, before it actually has run its pre-ordained course. In order to

110 Jeremy Howick. (2011): 72.

111 Jeremy Howick. (2009): 37. And Howard Brody and Franklin Miller. (2002). "What makes placebo-controlled trials unethical?" in *American Journal of Bioethics*. 2(2): 3-9.

112 Jeremy Howick. (2011):

successfully use stopping clauses however, terms like ‘harm’, ‘safety’ and ‘efficacy’ must be scrutinised closer.

Safety and efficacy are both natural endpoints for trials; as are mortality and morbidity. The TGN1412 trial is one example where the trial was stopped prematurely because of harm to the participants. Harm is divided into toxicity and death. In the literature the terms ‘harm’, ‘safety’ and ‘toxicity’ are often used interchangeably. “...morbidity and/or mortality outcomes may reflect both risk and benefit.¹¹³ As counterintuitive as this sounds, there are cases where obviously contradicting endpoints were established and one was deemed to be the actual endpoint, while the other was supportive, or surrogate, endpoint. This distinction means that negative results can be measured in interim analyses and do not lead to the trial being stopped, because they are considered to be a normal part of the trial. But one which is only limited and still supports an overall positive outcome. The idea behind this utilitarian sounding reasoning is that it cannot be quite clear if a treatment is harmful or successful until all the data is accrued. Surrogate endpoints should never be used as a reason to stop a trial prematurely.¹¹⁴

There are four specific reasons that are mentioned in most of the literature for stopping a trial early:¹¹⁵

1. adverse effects leading to unacceptable safety hazards
2. futility, that means that the sample size is too small to achieve significant values
3. new external information that either answers the question or raises safety issues
4. apparent benefit

Following Bassler's et al. reasoning here, it would be best if interim analysis are done by an independent committee. These committees are data monitoring committees (DMC) or data safety and monitoring boards (DSMB). They should be independent from those organising the trial and are therefore not blinded. If a committee is part of a trial set up, it can be blinded, because it would be sufficient to have it see the raw data. That is another reason why these committees are used more and more frequently. They can act as outside observers who do not have any interest in the outcome of the trial and who can therefore more easily determine if there is a problem in the setup or the conducting of the trial. “Nowadays, some

113 Yanli Zhao, Patricia Grambsch and James D. Neaton. (2007). “A decision rule for sequential monitoring of clinical trials with a primary and supportive outcome.” in *Clinical Trials*, (4): 140-153.

114 Jan Sprenger and Jacob Stegenga. (2017). “Three Arguments for Absolute Outcome Measures.” in *PSA2016: The 25th Biennial Meeting of the Philosophy of Science Association*.

115 Dirk Bassler, Victor M. Montori et. al. (2008). “Early stopping of randomised clinical trials for overt efficacy is problematic.” in *Journal of Clinical Epidemiology* 61: 241-246.

funding, ethics, and regulatory bodies consider an independent DMC essential for major RCT's. For instance, the Food and Drug Administration has published a draft guidance for clinical trials sponsors on the establishment and operation of clinical trials DMCs.^{116, 117} The important note on the FDA guidance is that it is non-binding. That means that other guidelines can be used, if and when the FDA ones do not fit to the actual setup of a trial. On the one hand, this approach gives room for multiple ways of setting up of trials and of establishing stopping clauses and DMC's, on the other hand, non-binding guidelines mean that they do not have to be used at all, and that trials can be set up without these safeguards.

If a trial is set up with stopping rules, there are initially two ways in which those can be used. One way is an 'or' stopping rule, which means that the trial is stopped "for either a safety *or* efficacy outcome."¹¹⁸ The other way is an 'and' stopping rule, meaning that the trial is stopped for safety *and* efficacy, rendering the results potentially more robust.¹¹⁹ In phase 0 and I trials, 'safety' is the main concern of the trial and therefore the natural endpoint. In phase II, III and IV trials, when a base-safety is established, 'efficacy' is of higher concern and can be the more natural endpoint of a trial, without neglecting safety. If real safety concerns do appear in later trials, those are definitive indicators to potentially stop the trial prematurely.

There are multiple ways, and multiple reasons, to terminate a trial before it would actually end. Stopping rules should be decided before the start of a trial and specified in the trial protocol.

Judged by the amount of literature about stopping rules in the case of benefit (a lot),¹²⁰ compared to the case of harm (very little)¹²¹ it would seem as if the decision to stop a trial in case of harm or toxicity is fairly easy. However, if looked at it closely, even in the case of harm, it is not an easy decision to stop a trial. A multitude of concerns can play a role here. Ethically it sounds obvious to stop a trial that is dangerous for the participants involved. However, some 'harms', like some side-effects might be only temporary. "The decision to stop early for harm is potentially more complex than benefit or futility because it may involve a trade-

116 Yanli Zhao, et.al. (2007): 142.

117 Food and Drug Administration (FDA) guidelines: "Guidelines for clinical trial sponsors." <https://www.fda.gov/downloads/regulatoryinformation/guidances/ucm127073.pdf>. The guidelines are fairly specific but nonbinding. It is assumed that when a different method is more appropriate for the test on hand then that method will be used. Last accessed on January 23rd, 2020.

118 Yanli Zhao, et. al. (2007): 140.

119 Yanli Zhao, et.al. (2007): 141.

120 Unrefined PubMed search with the search terms "stopping rules due to benefit" yielded 67 results. <https://www.ncbi.nlm.nih.gov/pubmed/?term=Stopping+rules+due+to+benefit>. Last accessed on January 23rd, 2020.

121 Unrefined PubMed search with the search terms "stopping rules due to harm" yielded 8 results. <https://www.ncbi.nlm.nih.gov/pubmed/?term=Stopping+rules+due+to+harm>. Last accessed on January 23rd, 2020.

off between potential — but as yet undemonstrated — benefits and apparent (but possibly spurious) adverse effects.”¹²² Mill et.al. argue further on how important these DMCs in the possible case of stopping a trial for harm are, because they might be able to distinguish real harm from side-effects and can react quickly to avoid greater harm for the participants.

In the case of stopping a trial for benefit, the question can only be if the benefit is a real benefit and not a spurious event at that particular point in the trial which makes the treatment look to be beneficial, but hides harms that would have come to light later on, for example through ‘regression to the mean’ of the disease in question. “The social value of the research is severely compromised when overly sanguine estimates of treatment effect result in misleading risk - benefit ratios, misguided practice recommendations, and suboptimal clinical practice....”¹²³

Stopping clauses are important, because they can act as a safe-guard to prevent harm for the participant and in very rare cases they can speed up a drug-approval process, if, and only if, benefit is established beyond a doubt. The latter case can be important if and when no standard treatment is available and the treatment under test is the only one to prevent harm for the patients in the actual target population. This is most often the case in drug-trials concerning orphan drugs.¹²⁴ Orphan drugs are those drugs that are treating a very small patient population with a very rare disease. Often orphan drugs are poorly researched because they bring little to no profit. However, if they are researched, most patients are quite desperate for the trial results because no other treatments are available. If early on a benefit is detected it might be possible to stop a trial early and launch a phase IV trial after market approval. Another solution however would be to enter in a compassionate use program, so that patients outside the trial can receive the treatment and the actual trial can run its pre-considered course, without the recourse to any stopping clauses.¹²⁵

2.6 External validity

As we have seen, when RCTs are performed correctly, they are internally valid and therefore their results should be robust and presentable. Questions however remain: presentable for what, and for whom? Can the treatment be used in the

122 Edward Mills, Cooper Curtis, et.al. (2006). “Randomized Trials Stopped Early for Harm in HIV/AIDS: A Systematic Survey.” in *HIV Clinical Trials* 7(1): 24-33.

123 Dirk Bassler and Victor M. Montori. (2008): 244.

124 Aaron S. Kesselheim, Jessica A. Myers and Jerry Avorn. (2011). “Characteristics of Clinical Trials to Support Approval of Orphan vs Nonorphan Drugs for Cancer.” in *JAMA*, 305 (22): 2320-2326.

125 Hanna I. Hyry and Jeremy Manuel et.al. (2015). “Compassionate use of orphan drugs.” in *Orphanet Journal of Rare Diseases*. 10 (100).

target population as it is used in the treatment population? Is the treatment even feasible in the target population? How can the problem of external validity or extrapolation, as it is also called, be solved?

The results of a RCT with a positive outcome show us that the treatment can be used for the population in which the treatment has been tested. However, treatments are not developed for some small trial population, they are supposed to work in the wider population who suffers from this specific disease. But are the treatment and the target population sufficiently alike so that the treatment could be successfully used in the actual target population? How can this necessary ‘alike-ness’ be established? The problem of transporting research results from trial conditions to the actual target population is called the problem of external validity or of extrapolation. Some authors do use the term generalisability, but for the sake of clarity, I will stick with ‘external validity’, as the opposite to ‘internal validity.’ Howick prefers to use the term extrapolation, because sometimes it can be necessary to extrapolate results within a trial to make it applicable to a participating subgroup. As an example Howick uses a trial in which a beneficial result was only pertaining to a very small subpopulation, within the larger trial population, with a very specific condition.¹²⁶ Even though the results have been extrapolated to a specific target population, the trial over all is not at that point deemed to be externally valid.

It is often said that the higher the internal validity of a trial, the bigger the problem of external validity. So the focus here is on assessing what contributes to the difficulty in extrapolating results in the set-up of, and recruitment for, RCTs and how trials results can be made usable for evidence-based practice.

2.6.1 External validity and trial design

The problem of external validity already becomes prevalent in the set-up of a trial. The claim: ‘the more internally valid a trial is, the less externally valid the results seem to be’ is already precluding that the ‘ideal’ set-up of a trial might not be the ‘best’ set-up.¹²⁷

The randomisation process in RCTs is used to prevent selection bias and to render control-and treatment arms as alike as possible, in order to achieve a common base line. Participants are chosen in such a way as to accomplish comparability. A certain similarity is therefore helpful. Most trial designs include a run-in

126 Jeremy Howick, et.al. (2013). “Problems with using mechanisms to solve the problem of extrapolation.” in *Theoretical Medicine of Bioethics*; 34: 277.

127 Nancy Cartwright. (2007). “Are RCTs the Gold Standard?” in *BioSocieties*. (2): 11-20.

period as a so-called ‘enrichment strategy’.¹²⁸ One type of run-in period is when already included participants have to submit to a wash-out period before data is gathered. That means that they have to stop taking other drugs or stop certain habits, like smoking, so that no traces are left in the body, before they can participate in the trial. A different form of wash-out period is when participants, also before data gathering, are given placebos and those participants that respond to the placebo are not included in the actual trial.¹²⁹ Reaction to placebo however can again be very varied and might be hard to detect, since any change in the patient receiving the placebo can have reasons unrelated to the non-working treatment. It is questionable therefore if a run-in period based on the placebo criteria is helpful.

Another type of run-in period is the randomised withdrawal design. Here participants are excluded from the trial when they are either non-compliant or if they show toxic effects because of the treatment under test. All types of enrichment strategies are compromising the external validity of RCTs. But they are still fairly prevalent, because they guarantee a more favourable result overall, and the more favourable the result for a new treatment, the earlier it will be approved and released into the market, thereby making profit. Hence, any form of enrichment strategy of a trial should be avoided if the trial results are supposed to be external valid overall and valid for the individual patient as well.

When all possible participants are included in the trial, with or without a run-in period, then the researchers have achieved nearly perfect laboratory conditions. But they have also narrowed the playing field considerably. The ideal participants for a trial seem to be male, in their mid-thirties to mid forties and without any co- or multi-morbidities.¹³⁰ And even if gender and age is more mixed, the lack of co- and multi-morbidities plays a significant role, because they can severely influence the outcome of any trial. Either because necessary medication that cannot be flushed out, interacts with the novel treatment or the disease under trial is overshadowed by these other diseases. Female patients are asked less often to participate in a trial, even if the treatment in question is treating a disease that is not dependent on gender.¹³¹

Obviously most trial participants in real life are far removed from the above ideal. The chief reason for this is that almost no one, regardless of gender, is free

128 Robert Temple. (2012). “Enrichment Strategies.” in Guidelines of the U.S. Food and Drug Administration. <https://www.fda.gov/ucm/groups/fdagov-public/@fdagov-afda-orgs/documents/document/ucm303485.pdf>. Last accessed on September 15th, 2017.

129 Ariel Pablos-Méndez, R. Graham Barr, and Steven Shea. (1998). “Run-in Periods in Randomized Trials Implications for the Application of Results in Clinical Practice.” in *JAMA* 279(3): 222-225.

130 Ben Goldacre. (2012): 159.

131 Wendy Rogers. (2004). “Evidence-Based Medicine and Women: Do the Principles and Practice of EBM further Women’s health?” in *Bioethics*. (81)1: 61.

of co- or multi-morbidities. Simple examples for this can be asthma, skin or joint problems, etc. Those minor ailments which the patient does not even perceive as morbidities but which nonetheless can alter the outcome of a trial can and are used to exclude possible participants from trials. Hence, the wish to either exclude as many of these predicaments or to distribute them equally to both groups is very understandable in theory, but hard to maintain in practice. Equal distribution is a valid goal in trial designs, but the exclusion of possible participants based on their co- or multi-morbidities is not advisable, since the results that would be garnered by such an idealised trial would not be generalisable to the target population. Often age does play a role. Children and the elderly are often excluded, again explained as a precaution. “Up to 90% of potentially eligible participants are sometimes excluded from trials according to often poorly reported and even haphazard criteria. For example, the most effective antidepressants in adults have doubtful effects in children.”¹³²

2.6.2 External validity and recruitment

Since the ideal participant, with the qualities outlined above, seems to be non-existent, the goal the recruiter has to aim for is ‘eligible’. The patient has to be eligible to become a participant. Therefore, recruitment of a group of eligible patients is of the utmost importance for a trial. But who is eligible? First of all, the patient/participant has to have the illness in question to even be considered for a trial from phase II onward. Phase I is usually conducted with healthy volunteers.¹³³ Secondly, the patient has to be at the right place at the right time. Some trials are advertised openly, but recruitment often happens on a walk-in basis, meaning that the clinic conducting the trial asks those patients who are walking in with the disease in question, to participate. Naturally, that excludes many eligible patients who are treated in a different clinic or by a GP who is not informed about the trial. Open advertising does not reach that many possible participants because as always one has to be attuned to look out for it. If one is not aware about the possibilities of a trial and is not made aware by their health-care provider, the clinics miss out on many eligible volunteers.

132 Jeremy Howick, et. al. (2013): 277.

133 There are, as always, exceptions to this rule. Some phase I trials are conducted in participants with the illness in question, because the treatment is so special that it should not or could not be trialled in healthy volunteers. A currently running trial at the UKE in Hamburg comes to mind in which children with NCL2, a severe form of childhood dementia, are receiving a very experimental treatment that would not be ethical to give to healthy volunteers. <https://clinicaltrials.gov/ct2/results?cond=NCL2&term=&entry=&state=&city=&dist=>. Last accessed on January 23rd, 2020.

Even if the patient is eligible and is asked to participate, he or she might have good reasons to refuse. Many patients are afraid of being treated as human guinea pigs and rather opt for the standard treatment. Or they are afraid of just receiving the placebo. However, if they agree, it can be for many different reasons which subsequently can influence the actual outcome of the trial.

There are different types of patients which are participating in trials, but almost none do so for a greater societal good. Either there is a financial component involved, or there is a real sheer need on the part of the patient, because the disease is so rare or so little studied, that even the chance of receiving some treatment, 50/50 in placebo controlled trials, is better than receiving none. A financial component can be that the patients are paid, as is very common in phase I trials where healthy participants are recruited. Again the TGN1412 trial comes to mind. And even in later trials, some sort of compensation is often offered.¹³⁴ Another case can be that patients within a trial receive care and medication for free and do not have to pay for it or have to claim payment from the insurers. Often trial medication can be very expensive. Additionally, in a trial setting, clinicians are a lot more thorough and the overall care is often better and the patients are more closely monitored. So, for many patients, being recruited to participate in a trial can be a win/win situation.

There is an ongoing reluctance to recruit female patients into medical trials, especially pregnant or lactating women. And to underscore the validity of my point here, one only has to look at the package insert of most established treatments. It is almost always stated that the medication has not been tested on pregnant and lactating women and should therefore not be taken by that individual group of patients. "Until 1993, the FDA excluded women of childbearing age from participating in early (Phase I and II) drug treatment trials, and this reduced their enrolment in Phase III trials (usually randomised controlled trials)."¹³⁵ It is easy to follow the actual motive behind that exclusion. Pregnant women are especially vulnerable and to not include them in any trials is often only done to safeguard them and the unborn child from harm. This became one of the cornerstones of trial recruitment after the thalidomide scandal. But as valid as this thinking is on a first glance, it is unfortunately not entirely feasible, even for pregnant women. An unfortunate consequence of this exclusion is that pregnant, and most of all, birthing women are subjected to all sorts of non-best-evidenced treatments. Among those are birthing positions, anaesthesia, and pain medication, to name the most common ones. These treatments and procedures have hardly changed over time. And some of them might be unnecessary or even harmful, but nobody challenges the practice

134 Christine Grady. (2005). "Payment of clinical research subjects." in *Journal of Clinical Investigation*. 115(7): 1681–1687.

135 Miriam Solomon. (2015): 142.

since there is no good evidence to do so. Pregnant and lactating women are therefore a huge group of patients which is still to a large extent excluded from evidence-based research. However, these women still might need treatment for certain conditions or are even dependent on it, especially if they have chronic diseases, like heart diseases or certain forms of rheumatoid arthritis. In both cases, as in many others, in order to manage the disease next to the pregnancy, the woman has to continue the medication, even though the evidence to do so is only based on experience and not on the best evidence possible.¹³⁶

It was generally assumed that research results based on research conducted with male patients could easily be extrapolated and used in a female target population. “The gender bias amongst participants in clinical trials is well known. Women have been excluded from research for many years, for a variety of reasons including the alleged need for homogenous populations, the fear of harms to pregnant women, the cost of including women, and the purported difficulty of recruiting women.”¹³⁷ It must be obvious to even those who have a limited medical knowledge that results cannot be easily transferred from one gender to the next. Although the core functions might be the same, even symptoms of the same disease, i.e. stroke, might display differently depending on the gender of the patient.¹³⁸

Minorities are another group of patients who is consistently neglected in the research of EBM. In the United States, African Americans and Latinos are consistently underrepresented in medical research. Most often because they are afraid to be used as guinea pigs.¹³⁹ It is often assumed that the experiences of Tuskegee, where only black males were recruited, without being told about their participation in a trial, are still prevalent in society and this being the reason why those minorities refuse to participate in research.¹⁴⁰ Wendy Rogers argues against this, claiming that because of their specific health needs or set-up, they are simply not asked to participate and would do so, if there would be a possibility. Many minorities in the United States do not have health insurance or see a GP on a regular basis. Research might help them, but is as equally unavailable to these groups as is regular health care. “Given higher rates of morbidity and mortality among ethnic minorities in comparison with majority populations, this lack of representation in

136 Margaret A. Honein, Suzanne M. Gilboa, and Cheryl S. Broussard. (2013). “The Need for Safer Medication Use in Pregnancy” in *Expert Review of Clinical Pharmacology*. 6(5): 453-455.

137 Wendy Rogers. (2004): 11.

138 A.H.E.M. Maas and Y.E.A. Appelman. (2010). “Gender differences in coronary heart disease.” in *Netherlands Heart Journal*. 18(12): 598–602.

139 B.R. Kennedy, C.C. Mathis and A.K. Woods. (2007). “African Americans and their distrust of the health care system: healthcare for diverse populations.” in *Journal of Cultural Diversity*. 14(2):56-60.

140 The Tuskegee experiment is explained at length in the chapter about informed consent.

research exacerbates the existing vulnerability of minorities to poor health outcomes.”¹⁴¹ And it makes it difficult to extrapolate the results of a trial conducted with Caucasian patients to African American patients, precisely because of the reasons mentioned in the quote.

The same holds true for patients with disabilities or psychiatric problems. Both groups of patients are considered minorities and are often not recruited. Either because they are not able to give informed consent or because they are also considered vulnerable. Again the reasoning is somewhat flawed, since research results garnered with ideal participants cannot be used as such for patients that have many co-morbidities and are already taking medication that has to interact with any new treatment. Especially when patients have to take psychotropic drugs on a continuous basis the interactions with other treatments have to be taken into account and carefully monitored. These interactions can and should be tested in advance, namely before the medication reaches market approval.

Another factor that should be taken into account when recruiting patients for a trial is the fact that people do change over time. They might develop some other disease and they do simply age. Worrall gives an example in which the trial population and the target population differed in age and how that affected the outcome.

“One example is the drug benoxaprofen (trade name: Opren), a nonsteroidal anti-inflammatory treatment for arthritis and musculo-skeletal pain. This passed RCTs (explicitly restricted to 18 to 65 year olds) with flying colours. It is however a fact that musculo-skeletal pain predominantly afflicts the elderly. It turned out that, when the (on average older) ‘target population’ were given Opren, there were a significant number of deaths from hepato-renal failure and the drug was withdrawn.”¹⁴²

The example not only shows that it can be dangerous to restrict the participants in an essential way. It also shows that if the existing clinical expertise would have been taken into account, the drug would from the beginning have been tested in an elderly population. The fact that “musculo-skeletal pain predominantly afflicts the elderly”¹⁴³ was well known before the trial. The trial could have been both internally and externally valid if the right participants would have been recruited into it from the very beginning.

141 Wendy Rogers and M.M. Lange. (2013) “Rethinking the vulnerability of Minority Populations in Research.” in *American Journal of Public Health*. 103(2): 2141-2146.

142 John Worrall. (2007). “Evidence in medicine and evidence-based medicine.” in *Philosophy Compass* 2(6): 995.

143 John Worrall. (2007): 996.

2.6.3 External validity and n-of-1 trials

To quote Jeremy Howick again “one type of randomised trial, namely n-of-1 trials have arguably the highest degree of external validity of any comparative clinical study.”¹⁴⁴ N-of-1 trials consist of one patient who receives either the treatment under test or a placebo or standard. In most cases the participant receives the placebo and the treatment on alternate weeks or month. Hence, the trial population and the target population is equal to each other.

N-of-1 trials do sound like the perfect alternative to standard RCTs because of the guarantee of external validity, if internal validity is given. However, n-of-1 trials are less reliable than could be assumed on a first glance. Since they only involve one patient it can be impossible to ascertain if that patient has improved because of the treatment, or because of “spontaneous remission” or because of the placebo-effect.¹⁴⁵ And it is impossible to infer how close this one patient resembles other patients with the same disease. N-of-1 trials are really only applicable for patients with chronic but otherwise stable diseases. Psoriasis and atopic eczema are examples for those, since a patient can test different skin treatments and see over time which one works the best. However, for most diseases, especially those which are unstable and are quickly changing, n-of-1 trials are not feasible, because their results can not be extrapolated at all.

“For example, it is impossible to know whether aspirin will prevent a patient’s stroke until it is too late. This is a problem with most cases of preventive medicine, and also with treatments for many acute conditions, such as meningitis, pneumonia or snake bite, where we don’t have the opportunity to test it in each individual patient and see. So we then have to rely on whether and how to apply the evidence from the experience of studying others.”¹⁴⁶

N-of-1 trials are therefore no solution to the overall problem of external validity but only provide a solution in very exceptional cases.

2.6.4 How can external validity be achieved?

External validity is hard to achieve and is lacking as main goal in many trial designs. Nonetheless, external validity is what is needed to make the trial results applicable to the target population and preferably to the individual patient. One solution described above, but only in very special cases, can be N-of-1 trials, but they are seldom feasible. The most obvious solution must be proper recruitment

144 Jeremy Howick. (2011): 55.

145 Jeremy Howick. (2011): 55.

146 Jeremy Howick. (2011): 152.

for trials. Since most people suffer from more than one condition, co-morbidities should not be discounted in the recruitment process but understood as known confounders and taken into account in the trial design. Women and children need to be included, in the case of children apparently only after careful consideration, into trial designs so that all age and gender groups are represented, depending on the disease and drug or treatment in question. That trials can carry a certain risk for the participant is acknowledged, but should not exclude the recruitment of all possible participants. In the next chapter I will discuss informed consent and the role it can possibly play in safeguarding participants as much as possible and how patients can be involved in the decision to participate.

Another way to make trial results more externally valid are phase IV trials. These happen after the market approval of a drug and can be either randomised trials or longitudinal observational studies. Goldacre proposes that, especially in cases where there are competing treatments for the same condition, large randomised phase IV trials should be conducted via a patient database to which every GP and clinician has access, at least in the UK. Germany has a fairly similar system. These databases are anonymising the data of the individual patient, but would provide the researcher with the overall number of patients having received a certain treatment and with the overall characteristics of these patients and observed adverse events or side-effects. Goldacre argues that when there is general uncertainty about which treatment is superior for a certain disease, the GP should use the regular prescription system, but instead of entering the patient data and printing a prescription, he would enter the patient into a randomised trial and either treatment A or B is assigned to the patient. The GP consequently reports in the follow-ups about the performance of the treatment and, over time, a patient population which benefits from either the one or the other treatment would be established.¹⁴⁷ The follow-up, at least in the UK, would be very easy and would not entail any more work for the GP, since all patient-data is recorded by a computer system anyway. Apparently the only methodological virtue missing in this scenario is blinding. But Goldacre argues that since existing treatments are compared, the methodological role of blinding is not as significant, since patients often do not have a preference towards either treatment. And the negative effect of non-blinding is calculated against the overall long-time effect of such a study, which can in theory run indefinitely. In the proposed circumstances the number of patients/participants is definitely big enough to make a statistical difference, and the treatments are not tested in an idealised trial population, but in the actual target population.¹⁴⁸ The problem of external validity is thereby solved. For a lot of diseases today there exists a standard treatment. New treatments are often either a variation of the standard or

147 Ben Goldacre. (2012): 227.

148 Ben Goldacre. (2012): 228.

at least comparable to it. A phase IV trial as described above would however not work for placebo controlled trials. But with general uncertainty about the efficacy of a treatment and the possibility to compare ‘new’ against ‘standard’ in the normal patient population, these kind of phase IV trials would be highly ethical.

2.6.5 *Publication bias*

Publication bias, one of the biases that does not play a role for the internal validity of the trial, should also be eliminated to make the trial results externally valid. Only if all trial information and preferably even the raw data is available, can the results be assessed for the actual patient in an hospital or at the GP. Every missing piece of information, let alone unpublished trials that had negative outcomes, severely distort the evidence-base.¹⁴⁹ All trials should be published, regardless of their outcome. This is what the AllTRials campaign is diligently working for across the globe.¹⁵⁰ And the data should not be tampered with to make results look better. This kind of tampering is called ‘spin’ and severely distorts the data so that it looks favourable when in actuality it is not.¹⁵¹

A lot of trials today are sponsored by pharmaceutical companies. These companies are naturally interested in positive results for the drugs they are trying to bring to market approval. Negative results do not bring profit. So if a trial is negative, either the data is manhandled to a degree that it looks good, even if it would only be the case for a tiny sub-group of patients, or the data is not made public at all. As if the trial never existed. The most compelling reason why this approach to data-handling is unethical is that participants who have given their time and their health for research are not rewarded and their engagement was doubly in vain because another company, or researcher, might come up with the same, or a very similar, idea and conduct a trial, absolutely unaware that it already has been done and proved unsuccessful. As Imogen Evans and colleagues are saying: “Unnecessary research is a waste of time, effort, money, and other resources; it is also unethical and potentially harmful to patients.”¹⁵²

149 Imogen Evans, Hazel Thornton, Iain Chalmers and Paul Glasziou. (2011): 163.

150 AllTRials Campaign. <http://www.alltrials.net/>. Last accessed on January 23rd, 2020.

151 Carl Heneghan, Ben Goldacre and Kamal R. Mahtani. (2017). “Why clinical trial outcomes fail to translate into benefits for patients” in *Trials*. Bio Med Central: 1-7.

152 Imogen Evans, Hazel Thornton, Iain Chalmers and Paul Glasziou. (2011): 129.

2.6.6 *Surrogate outcomes versus primary endpoints*

There are multiple ways in which data can be changed during and after the running of a trial. One method is to change the overall intended outcome in the middle of a trial. Outcomes need to be specified in the trial protocol. Changing them mid-trial or after the trial, compromises the entire trial results, because the trial initially was not designed to look for these outcomes. Outcomes are often also called endpoints and are divided into surrogate endpoints and primary endpoints. A primary endpoint of a trial can be death. An often used example are trials for cardiac conditions due to high cholesterol. An obvious primary endpoint would be cardiac arrest and subsequent death of the patient. So a question like “how many patients need to be treated with a cholesterol lowering drug, in order to save one patient from pre-mature cardiac arrest and death?” might be a valid research question. However, specific causes of deaths are hard to come by in patients with a pre-existing cardiac condition. Death might have occurred despite the successful lowering of the cholesterol. Or no deaths do occur in a given period, or they do not occur during the duration of the trial. Therefore it might be easier to opt for a surrogate, or soft, endpoint. In the example the measure of cholesterol in the blood. The argument would then be that lower cholesterol-levels overall lead to fewer overall cardiac arrests and therefore it is deemed sufficient to measure cholesterol levels. If cholesterol-levels are lower with the novel treatment than with the standard or placebo then the novel drug is deemed to be superior in preventing premature deaths due to heart attack. However, cholesterol levels, albeit playing a significant role in overall cardiac health, are not in themselves the cause of heart attacks, but are merely a contributing factor.¹⁵³ And cholesterol levels can change during the day. They are to a certain degree susceptible to diet and exercise and are certainly not constant in any given patient over time. So a trial that chooses the lowering of cholesterol as a surrogate outcome for the efficacy of an overall heart medication does not represent any really valuable results for the actual patient. A good example often used in this context is the prescription of statins to patients with high cholesterol. Statins were considered to be *the* drug to lower cholesterol levels and to prevent heart attacks with a minimum of side-effects.¹⁵⁴ Treatment with statins is a long-term treatment and patients were advised that if they needed statins, they would need them for the rest of their lives. New findings however have shown that statins *do* lower cholesterol, but *do not* lower the overall risk of a hear attack, which was the reason that they were prescribed in the first place. Since low or high cholesterol levels in the blood do not influence the overall well-being

153 JJ Kastelein, A. Wiegman, and E. de Groot. (2003). “Surrogate markers of atherosclerosis.” in *Atherosclerosis Supplements*. 4(1):31-6.

154 B. Ziaian and G.C. Fonarow. (2017). “Statins and the Prevention of Heart Disease.” in *JAMA Cardiology*;2(4):464.

of the patient and are not detectable by the patient, statins do nothing for the everyday quality of life. If, on top of things, they do not prevent heart attacks, it is questionable if they are useful as a drug, especially on a long-term basis. The surrogate endpoint of lower cholesterol versus the primary endpoint of death due to cardiac arrest has not shown to be sufficient for the efficacy of the class of drugs that are called statins.

2.6.7 *Equipoise*

Clinical, and even personal, equipoise has been heralded as an ethical tool to decide if a RCT can be performed, and is neither superfluous nor posing a risk to possible participants. The principle of clinical equipoise simply means that there has to be genuine uncertainty on the part of the clinicians who are conducting a trial, whether or not a new treatment is really more beneficial than a conventional one for the same medical problem, before a clinical trial is started. Freedman, the one who coined the term ‘equipoise’ in the 1980’s, argues that clinical equipoise is satisfied “if there is genuine uncertainty within the expert medical community — not necessarily on the part of the individual investigator — about the preferred treatment.”¹⁵⁵ Personal equipoise would mean that the individual investigator is in a state of uncertainty about which treatment is superior. However, personal equipoise is hard to maintain and even to establish in the first place. Therefore the focus of the debate is about clinical equipoise.

Worrall explains that the ethical dilemma is easily avoided, if one follows the rule that only RCTs are producing valuable evidential results, since before a RCT is conducted, there can be no certainty about superiority either way. The “telling” evidence about the novel treatment is missing.¹⁵⁶ That presumes however that the only ‘telling’ evidence comes from RCTs. As has become obvious, and hopefully will become more so, not all ‘good’ evidence is automatically generated by RCTs. ‘Good’ evidence can come from experience and observational studies as well. The ECMO case will be used as an example in which there was good prior evidence that the treatment would be successful but where RCTs were performed because they were deemed necessary. Although clinical equipoise was already, and some even argue, severely compromised.

And what about clinical equipoise in placebo-controlled RCTs? Some argue that clinical equipoise is violated in placebo-controlled trials, since there cannot be a general uncertainty about which treatment is superior. The clinician therefore

155 Benjamin Freedman. (1987). “Equipoise and the Ethics of Clinical Research.” in *New England Journal of Medicine*. (317): 141-145.

156 John Worrall. (2008): 8.

would break the trust of the patient who believes to receive the best possible treatment. The distinction between research and clinical practice is important to answer that question. The patient who is participating in a trial has changed into a participant and the clinician has changed into a researcher. The participant has consented to being part of a trial and the relationship of trust has consequently shifted. Clinical equipoise here cannot be used as an ethical safe-guard for the patient.¹⁵⁷ Nor is it a safe-guard in all other types of clinical trials. Again, clinical trials need to be performed before a novel treatment can achieve market approval. That is in most cases not only a technical but also a legal prerequisite. Therefore the question if it is legitimate to perform a RCT is raised far less often than Freedman and other proponents of clinical equipoise would like. And clinical equipoise is as easily subverted as is the blinding. If one treatment does better than the competitor, clinical equipoise is almost impossible to maintain. And again, patients agree to be part of a trial and become participants in research. The mutual trust between patient/clinician has shifted to participant/researcher and the expectation to receive the best possible treatment has changed to being part of a trial and not being subjected to unduly harm.¹⁵⁸ Equipoise therefore does not seem to help in assessing the overall ethical questions of RCTs.

2.7 Unnecessary trials

Some trials are just unnecessary. For example trials which inadvertently are replicating trials which have already been performed but the results were, due to publication bias, never published. And one more ethical problem can be that RCTs are not necessary because the novel intervention is so convincing that a RCT would not have been required. Either the mechanistic evidence or quite simply the experience with the new treatment is so convincing that a RCT would do more harm than good, because participants really would receive the ‘lesser’ care when receiving the standard. Admittedly this later case is quite rare, and can only happen where non-drug treatments are concerned where there is no legal requirement to perform a RCT. The following example of such a case can be used to illustrate quite convincingly why it is not always in the patients interest, and overall better, to perform RCTs just for the sake of a statistical outcome. In the example the legal requirement to perform a RCT is also not relevant, since a procedure was put under test, rather than a drug.

157 Mario Castro. (2007). “Placebo versus Best-Available-Therapy Control Group in Clinical Trials for Pharmacologic Therapies. Which Is Better?” in *Proceedings of the American Thoracic Society*, 4(7): 570–573.

158 Anna Floyd and Anne Moyer. (2009). “Equipoise may be in the Eye of the Beholder.” in *The American Journal of Bioethics*. 9(2): 21-22.

2.7.1 ECMO — or an example why not all RCTs are equal

ECMO stands for ‘extracorporeal membranous oxygenation.’ The particular form of ECMO I want to discuss here was established to circumvent the lungs in neonates born with persistent pulmonary hypertension, a malfunction of the lungs, in which the lungs are not yet capable of oxygenating the blood sufficiently. “The idea of treatment is very simple: venous blood is taken from the baby, pumped round a circuit which includes a membrane where the blood is oxygenated, reheated to body temperature and passed back into one of the baby’s carotid arteries — ¹⁵⁹thus bypassing the baby’s lungs, the immaturity of which is implicated in the persistent hypertension.”¹⁶⁰

The mortality rate before ECMO of neonates born with persistent pulmonary hypertension was 80%. ECMO produced a 80% survival rate.¹⁶¹ The treatment was immediately tremendously successful, as was predicted by the physicians who had designed it. Despite this huge immediate successes, it was deemed necessary to perform an RCT in order for the treatment to be implemented in other hospitals as well, and to convince the medical community that ECMO really was better than the current standard treatment. It is important to note here that ECMO was very successful but not without risks, one of these and the most severe being intercranial bleeding which can lead either to death or to a severe brain malfunction. The consequences of which can be a severe handicap, a markedly reduced life-span and a reduced quality of life. Since *not* treating these newborns would mean unavoidable death, the possibility of detrimental side-effects of the treatment versus the chance of a full recovery were certainly recognised but the severity of the possible side-effects was deemed crucial enough to establish clinical equipoise with regard to ECMO. At least when following the argument of Robyn Bluhm. It is questionable in the case of ECMO if clinical equipoise was ever present. The inventors of ECMO believed in their new system and the survival rates proved them right. Other clinics could have been convinced by mechanistic reasoning through the method of ECMO and what it was supposed to treat. Robert Truog argued along those lines by claiming that it would have been enough to conduct a long-term observational study which would have shown ECMO to be superior to the standard treatment.¹⁶² Only a very stringent view as described above by Worrall, namely

159 http://commons.wikimedia.org/wiki/File:ECMO_schema-1-de.png License: This file is licensed under the Creative Commons Attribution-Share Alike 2.0 Germany license. Uploaded on November 14th, 2019.

160 John Worrall. (2008) *Evidence and Ethics in Medicine*. in Perspectives in Biology and Medicine. 51 (3): 419.

161 R. H. Bartlett, A. B. Gazzaniga. et.al. (1986). “Extracorporeal membrane oxygenation (ECMO) in neonatal respiratory failure. 100 cases.” in *Annals of Surgery*. 204(3): 236–245.

162 Robert D Truog. (1999). “Informed consent and research design in critical care medicine.” in *Critical Care*. 3(3):R29–R33.

that there is no evidence without RCTs, did lead to the ECMO trials. And as has become clear, this stringent view is not necessarily the one that provides external validity of a treatment and the best possible outcome for the individual patient.

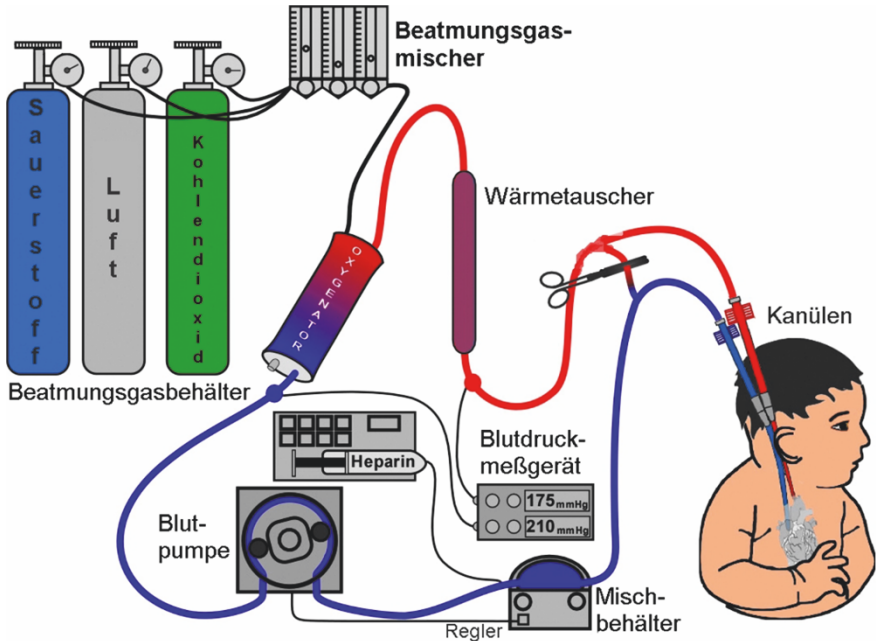


Figure 1: ECMO schema, to provide neonatal extracorporeal oxygenation

Source: https://upload.wikimedia.org/wikipedia/commons/thumb/f/fb/Ecmo_schema-1-de.png/1024px-Ecmo_schema-1-de.png. License: This file is licensed under the Creative Commons Attribution-Share Alike 2.0 Germany license. Last accessed on November 14th, 2019.

Three RCTs were performed in a fairly short amount of time, of which the first two were deemed to be flawed. The first RCT used a method of randomisation called ‘randomised plays the winner,’ a method of randomisation that happens on an ongoing basis while the trial is already in progress.¹⁶³ Most often an urn or some such device is used, even if just computer generated, that contains an equal number of, for example red and white balls. Red in this example stands for ECMO, white for the standard treatment. The urn is re-stacked with balls depending on survival. If a child survives on ECMO, then a red ball is added, if it dies on the standard, again a red ball is added. This method produced a biased urn very quickly, because

163 Catherine Cornu et al. (2013). “Experimental designs for small randomised clinical trials: an algorithm for choice.” in *Orphanet Journal of Rare Diseases*. 8(48): 4.

the first child was assigned to ECMO and survived while the second was assigned to the standard and died. Since with such a heavily biased urn, a statistically sound assessment of the validity of the trial was not possible anymore, a second trial was deemed necessary.

The second trial used a different form of adaptive randomisation to begin with and then it used the Zelen method. For those reasons, and an incorrect use of stopping clauses, or so it was claimed, the trial was stopped early.¹⁶⁴

A third trial was initiated. However, since some form of comparison was deemed necessary, the methods used in both arms of the trial were already outdated by the methods used to treat pulmonary hypertension in neonates in the every day clinical setting. The standard had become better and more successful over time, as had ECMO.¹⁶⁵ Still, the outdated versions were used, solely for the sake of science and certainly not for the sake of the actual patients. It can be said without a doubt that any RCT conducted under these circumstances is not ethically correct and should not be permitted to be performed. Clinical equipoise cannot have been present anymore. The technique was already successfully used in many clinics and had therefore proven to be externally valid. The consent given by the participants of the last trial can not have been that informed, since there were better options available and to make matters even worse, the last trial did change nothing in the overall acceptance of ECMO in every day clinical practice.¹⁶⁶

2.8 Conclusion

Medical research is important and can save lives. If done correctly, medical research is as safe as humanly possible and there are many solutions to translate the results of medical research into usable results for the individual patient. As we have seen with ECMO, medical research, and especially RCTs, should not be conducted just because the methodology is deemed superior. If other forms of evidence are available, like mechanistic reasoning, or the evidence from other forms of studies, and even experience, it should be taken into account and used wherever appropriate. This approach to evidence not only saves time and money but also patients from being subjected the superfluous trials or dangerous treatments. And patients and participants need their own voice in the medical process, not only in

164 Robyn Bluhm. (2010). "The epistemology and ethics of chronic disease research: Further lessons from ECMO." in *Theory of Medical Bioethics*. 31:107–122.

165 Robert D Truog. (1999):R29–R33.

166 Valerie Mike, Alfred N Krauss, and Gail S Ross. (1993). "Neonatal extracorporeal membrane oxygenation (ECMO): clinical trials and the ethics of evidence." in *Journal of medical ethics*; 19: 212-218.

medical practice but also and especially in medical research, since without participants there would be no research. The tool of informed consent is what can provide this voice.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder..





3 Informed Consent and shared decision making in EBM

3.1 'Informed consent' in regular medical practice

'Consent' as it is understood in the medical context has to be asked from the patient and is the explicit agreement to waive a right to certain rules and norms which are normally expected in the treatment of other people and of ourselves as patients. Every surgical procedure would, without consent from the patient, be legally understood as assault and battery and the physician could be prosecuted for performing it. 'Informed consent' therefore in its most simple form means that the patient has received a good explanation about a medical procedure, understands what is happening to him or her and then can make an informed choice to accept or refuse, in the latter case the so called 'informed refusal'.¹⁶⁷ In order to give 'informed consent' the patient has to be capable of understanding the information given by the physician. He or she must be competent to decide and to give consent voluntarily without being coerced by any means into giving consent.¹⁶⁸ 'Autonomy' of the patient, hereby equated with 'person' plays the overarching role in 'informed consent.' A competent person who exercises autonomy will have the final say about their own life. 'Autonomy' itself is a contested term in the philosophy of science and interpretations therefore vary. According to Dworkin, "Liberty (positive or negative) ... dignity, integrity, individuality, independence, responsibility and self knowledge ... self assertion ... critical reflection ... freedom from obligation ... absence of external causation ... and knowledge of one's own interests."¹⁶⁹ all fall under the definition of autonomy. For my argument I will stick with the definition of autonomy as being "aware of ones own interest" and the limitations that are set to the subsequent decisions by circumstances and societal norms. Decisions made by such an autonomous patient should be respected and adhered to. However "no theory of autonomy is acceptable if it presents an ideal beyond the reach of normal agents and choosers."¹⁷⁰ What Beauchamp and Childress mean by this is that no

167 Tom Beauchamp and James Childress. (2009). *Principles of Biomedical Ethics*. Oxford: Oxford University Press: 120.

168 Thomas Schramme. (2002). *Bioethik*. Frankfurt: Campus Verlag: 31.

169 Gerald Dworkin. (1988). *The theory and practice of autonomy*. Cambridge: Cambridge University Press: 6.

170 Tom Beauchamp and James Childress. (2009): 101.

person is ever fully autonomous, since all persons are part of society, influenced by society and acting in accordance to it. All persons are dependent on something and especially in the medical context, patients are dependent on the information provided to them and on their own prior knowledge in order to make autonomous decisions as in refusing or consenting to a treatment on the basis of being thoroughly informed about it.

Again “informed consent has a role only where an activity is already subject to ethical, legal or other requirements.”¹⁷¹ In today’s medical practice, it is a usual procedure for the patient to give written consent before a surgery is performed. The surgeon explains the procedure, what is done, what is supposed to be the result and what are possible risks and side effects, and hands out written information that the patient normally can contemplate. Additionally to the surgeon, the anaesthetist and administrative staff also require informed consent signatures, since the patient and the clinic enter into a treatment contract.¹⁷² Unfortunately, the more people are involved, the more convoluted the information gets and the more confused some patients do become. Very often a large percentage of the information provided to the patient is written in medical language and does not help to enlighten a lay person. Therefore, paperwork concerning informed consent should be written in ‘normal’ language and the clinician should take the time to answer questions. If all informed consent papers are signed, the hospital administration will provide a contract between the hospital and the patient that allows for the procedure to take place.

3.1.1 *Implied or simple consent*

Patients can also give implied or simple consent, by allowing the physician to perform certain medical acts, such as drawing blood or giving a local anaesthesia (full anaesthesia needs written consent). Dentistry is a good example which often works without prior written consent.¹⁷³ The sitting down and opening of the mouth is deemed to be implied consent for the procedure to be performed. This does not waive the necessity on the part of the dentist to inform the patient about the procedure ahead, possible risks and hopeful benefits. And the patient can revoke the consent at any time during the procedure.

171 Neil C. Manson and Onora O’Neill. (2007). *Rethinking Informed Consent in Bioethics*. Cambridge: Cambridge University Press: 72.

172 Andrew Lloyd, Paul Hayes, et.al. (2001). “The Role of Risk and Benefit Perception in Informed Consent for Surgery.” in *Medical Decision Making*, 21(2): 141-149.

173 Kevin I. Reid. (2017). “Informed Consent in Dentistry.” in *The Journal of Law, Medicine and Ethics*. 45(1): 77-94.

3.1.2 *'Informed consent' and trust*

By giving 'informed consent,' a form of trust between the physician and the patient is established in which the patient allows for some kind of bodily invasion and trusts the physician that the intervention is in the patients' best interest. Informed consent in the medical setting is target-oriented. It is not a general waiver for the physician to do whatever he wants to do, but gives a limited ok for a certain procedure. Everything above and beyond that needs a new waiver, even if a patient is already under anaesthesia and needs to be woken up again to agree to the next procedure. And informed consent can never be understood as a fail-safe since it can be revoked at any time. So the physician has to be aware that even if the patient has consented to a procedure, it is the patients right to walk out of it. In reality, especially where surgeries are concerned, this case will hardly happen, but in medical research the participant can refuse to longer be a part of a trial and revoke the consent without the researcher having any recourse on the situation.

3.1.3 *Informed refusal*

A conflict between patient and physician can arise when a patient opts for 'informed refusal' and, for example, denies a life-saving intervention. The physician is required to protect life and to save it when in danger. The patient is formally allowed to not have certain procedures performed.¹⁷⁴ In those cases where the patient is fully capable to make the decision to refuse treatment, the physician has to accept the patient's decision. These cases are most often prevalent in cancer treatments when a patient decides to stop the actual treatment and most often then opts for palliative care instead. The patient still has to sign a waiver stating that he will forgo active treatment, so that the physician is in the clear and cannot be legally prosecuted for not saving a life.

Many of these cases are uncontroversial. In many end-of-life decisions most clinicians agree to a change in treatment. The focus shifts from life-prolonging measures to palliative care where the 'quality of life' is judged more important. The main goal of a 'quality of life' care approach is to alleviate pain, fear and other symptoms, but to let nature run its course and to let the patient die with dignity.¹⁷⁵ In palliative cases the patients wishes are paramount and it is acknowledged that the best possible evidence to treat the disease in question is not viable anymore for the patient in end-of-life care. However, there are cases where it is less clear why a patient is refusing treatment and where there might be even a conflict of interest

174 Thomas Schramme. (2002): 36.

175 Peter A. Singer, Douglas K.Martin, and Merrijoy Kelner. (1999). "Quality end-of-life care: patients' perspectives." in JAMA 13;281 (2): 163-168.

between the patient and the physician. When a patient denies treatment out of religious reasons, most physicians will accept this, but if a patient denies lifesaving treatment because he or she is afraid of the procedure or of possible side effects then the physician has to aim at a better understanding of the situation on both sides. This understanding on the part of the physician is part of the ‘informing process’ to obtain ‘informed consent’ and to arrive at the ultimate goal of a shared decision. Conflicts of interest do arise when patients are refusing a life saving treatment without good reasons to do so. The physician has the obligation to save and protect life and has to react if and when the patient is in immediate danger. In some of these special cases the physician can and will question the ability and competence of the patient to refuse treatment.¹⁷⁶

3.1.4 Rare limitations to ‘informed consent’

There are very rare situations in medicine where the case is reversed. One such case has garnered enormous traction in the UK press over the summer of 2017. The terminally ill 11 month old baby boy, Charlie Gard, being treated at Great Ormond Street Hospital for mitochondrial depletion syndrome, was the focus of media attention for months.¹⁷⁷ The disease in question is a cell-disease in which the mitochondria cease to produce the energy which the cells need to function.¹⁷⁸ The hospital wanted to move the boy to palliative care and to let him die with dignity, while the parents wanted to take the boy to the USA to try a highly experimental treatment with a 10% chance of some change in the disease progression. The medication, nucleoside bypass therapy, is still in its very experimental phase and has not yet passed the laboratory stage. It has never been used on a patient with Charlie’s specific strain and the US physician admitted that it would be highly unlikely that Charlie would actually benefit from the treatment.¹⁷⁹ The parents and the hospital found themselves to be locked in a legal battle about who held the guardianship over the boy and who subsequently had the final say in treatment decisions. Even though these cases are fairly rare they do happen. In the UK, if such a case arises where the hospital and parents or carers are not agreeing on the treatment options, the courts have to make the last decision, not the parents and not the hospital. The parents in Charlie’s case could not revoke the informed consent they had initially given to allow their son to be treated in the hospital. In the end and stressing here that the boy was terminally ill, the court decided against the

176 Thomas Schramme. (2002): 36.

177 Robert D. Truog. (2017). “The United Kingdom Sets Limits on Experimental Treatments: The Case of Charlie Gard.” in JAMA. 318(11).

178 Josef Finsterer and Salma Wakil. (2015). “Mitochondrial Depletion Syndromes.” in eLS. 1–9.

179 Robert D. Truog. (2017): 1001.

parents wishes and the boy died in a hospice. In most other countries, it would have been possible for the parents to move their son to a different clinic, even to a different country. However, it is more than questionable if such a move would have changed the outcome in Charlie's case. The question in this case and in cases similar to these is not only about treatment options but also about the quality of life and how very important that aspect is in clinical medicine. Much more important than some critics of conventional medicine are giving the system credit for.

Above it is stated that 'informed consent' is target oriented and that a surgeon is not allowed to extend the 'informed consent' to perform an entirely different surgery. However, there are exceptions to this rule, when it can be deemed in the best interest of the patient to perform a lifesaving surgery, even though a different surgery was agreed upon. Such a special case falls under the category of emergency medicine. A vivid example of such a case was a young woman who was diagnosed with a tubal pregnancy which needed to be removed immediately to prevent further harm. During the surgery, the surgeon realised that the symptoms were actually not stemming from an ectopic pregnancy but from an acute appendicitis and that the pregnancy was actually in utero and intact. The surgeon decided to remove the appendix, thereby saving the patients life and the life of the unborn child.¹⁸⁰ In this case the surgeon was legally 'allowed' to perform the altered surgery without obtaining prior informed consent from the patient. Acute appendicitis can lead to death and therefore it was an emergency situation. The risk to wake the patient up first was far greater because of possible complications due the acute appendicitis and due to the need to perform a second general anaesthesia in a very short time period, than to remove the appendix then and there.

In most emergency situations the overall goal is to save the patients life and at least to stabilise the patient until a hospital is reached. Emergency physicians often do not have the time to obtain 'informed consent', especially in case of accidents where multiple patients are involved and need to be treated. In many cases there is simply not the time nor the right situation to ask for the patients consent. In most cases it is assumed that the patient wants to be saved. It is a little different however when an emergency unit is called to a dying patient. Often families or carers all of sudden become afraid when a person is about to die and dial the emergency department in panic. In most of these cases it is hard to establish for the arriving team what the actual wish of the dying patient was to begin with and subsequently patients end up in hospital, on life support, without ever having consented to this kind of treatment.¹⁸¹

180 Kurt Hartman and Bryan Lang. (1999). "Exceptions to Informed Consent in Emergency Situations." in *Hospital Physician*: 54.

181 Roberto Forero, Geoff McDonnell, Blanca Gallego, et al. (2012). "A Literature Review on Care at the End-of-Life in the Emergency Department." in *Emergency Medicine International*.

3.2 The shift from paternalism to informed consent

Before it became usual practice to ask for written consent by the patient, most physicians acted paternalistically. They decided what was best for the patient and acted accordingly, without explaining to the patient why it would be the best course of action and without pointing out alternatives or explaining that the patient does not have to agree to the intervention at all.¹⁸² As long as the physicians really had the best interest of the patient at heart, a paternalistic approach might not have been to the detriment of the patient, but the patient in this scenario is certainly lacking autonomy.

3.2.1 *From the Hippocratic Oath to the Declaration of Helsinki*

The Hippocratic Oath that most medical practitioners did, and in some form still do, swear since antiquity, and which was considered binding for the medical community, does forbid to harm the patient and/or to use the patient for the physician's gain. During the Second World War there were unfortunately many German physicians who disregarded the oath. They used concentration camp and prison inmates as guinea pigs, taking their deaths for granted and claiming they were doing so for the greater good of the German nation. During that time medical experimentation was not regulated, and even if it would have been regulated internationally, the NAZIS would have disregarded such regulations. There were no restrictions in place for the 'experimenters' to do whatever they deemed necessary. After the camps were liberated, and during the Nuremberg Trials, many of these atrocities came to light and triggered a movement to prevent such blatant disregard for the integrity and well-being of human beings in the future.¹⁸³ Informed consent, given in a way as it is described above, plays a major role in this prevention program. Patients today can decide for themselves if, and often even when, they want to be subjected to medical procedures. They can choose the physician and they can refuse treatment, even though it might not be in their best interest. Nir Eyal describes informed consent as a tool to save trust in the medical profession.¹⁸⁴ However, that trust can only be established if the tool of informed consent is used correctly. Written explanations, often in somewhat medical language, have to be supplemented with an active discussion between patient and physician in which questions can be

182 A discussion of the different forms of 'paternalism' would supersede the scope of this paper. For further reading: Thomas Schramme, Ed. (2015). *New Perspectives on Paternalism and Health Care*. Heidelberg: Springer Verlag.

183 Horst H. Freyhofer. (2004). *The Nuremberg Medical Trial: The Holocaust and the Origin of the Nuremberg Medical Code (Studies in Modern History)*. New York: Peter Lang Publishing.

184 Nir Eyal. (2012). "Using Informed Consent to save trust." in JME Online.

asked and treatment decisions can be made together. The physician has to recognise if the patient needs more time to make a decision, or if an initial discussion was sufficient to both parties to be comfortable with the treatment plan. The patient has to be aware that at any given point in time he or she has the possibility to stop the treatment and to either clarify procedures or to abandon a certain treatment altogether.

An even more important historical tool to achieve wide-spread trust in the medical profession was the Declaration of Geneva which is still in use today and followed the Nuremberg Code of 1947. The Nuremberg Code, conceived during and after the Nuremberg Trials, specified consent as ‘voluntary consent’ and demands that the subject giving consent cannot be coerced into doing so and knows what he is giving consent to.¹⁸⁵ Manson and O’Neill point out that the Nuremberg Code does not exactly specify informed consent. The concept of patient autonomy or of consent with regard to later use of already acquired patient data or tissue is never mentioned or explained in detail in the Nuremberg Code.¹⁸⁶ In order to clarify the points made in the Nuremberg Code, the World Medical Association¹⁸⁷ came together in Geneva, Switzerland, and formulated eleven points which were hoped to represent a binding agreement for everyone in the medical profession. And thus was formulated for the first time the Declaration of Geneva.¹⁸⁸

The Declaration of Geneva was first drafted in 1948 and was continuously amended over time, for the last time on October 14th, 2017. The Declaration of Geneva was and is understood, and used, as a modernised version of the Hippocratic Oath. But even the Declaration of Geneva only specifies the relationship between patient and physician, again without specifying informed consent and without taking medical research into account. But as has become already obvious, medical research transgresses the boundaries of medical practice. The patient-physician relationship collapses to a certain degree and transforms into a participant-researcher relationship. There will always be a conflict between the physician in the role as caregiver to the patient and the physician as researcher who needs participants for his research. The researcher cannot guarantee that his actions are in the patient’s best interest. They might be beneficial for the actual patient, but they as well might not be. They certainly are beneficial for future patients, if all research

185 For the exact wording of the Nuremberg Code, please refer to <https://history.nih.gov/research/downloads/nuremberg.pdf>. Last accessed on January 23rd, 2020.

186 Neil C. Manson and Onora O’Neill. (2007): 3.

187 The World Medical Association speaks for health care practitioners globally and continually strives to achieve and maintain better standards of medical care. To learn more about the World Medical Association, please refer to www.wma.net.

188 World Medical Association, The. “Declaration of Geneva.” Adopted by the 2nd General Assembly of the World Medical Association, Geneva, Switzerland, September 1948. Lastly editorially revised by the WMA General Assembly, Chicago, United States, October 2017.

is published, because either the new treatment is successful or it was abandoned due to harm and future patients are not exposed to it.

In order to codify and integrate medical research, the World Medical Association yet again came together and formulated the Declaration of Helsinki in 1964. The correct title of the Declaration is “Declaration of Helsinki – ethical principles for Medical Research involving Human Subjects.”¹⁸⁹ Helsinki acknowledges the difficulty of the physician who is researcher at the same time. The first general principle sites the first point of the Declaration of Geneva. “The Declaration of Geneva of the WMA binds the physician with the words, ‘The health and well-being of my patient will be my first consideration,’¹⁹⁰ and the International Code of Medical Ethics¹⁹¹ declares that, ‘A physician shall act in the patient’s best interest when providing medical care’”.¹⁹²

Helsinki is supposed to set the standard for medical research ethics and is as such often used in national regulations or laws. However, the Declaration is not legally binding in international law. The latest version of Helsinki from 2013 includes 37 points and specifies and explains informed consent in medical research settings. In comparison to the version of 2008, the 2013 version has a whole paragraph titled “informed consent”. The earlier version uses the term “competent human subjects” which is changed in the version of 2013 into “individuals capable of giving informed consent.” The main aim of the paragraph about informed consent is not only to specify that it is important that consent is obtained, but also how encompassing the information needs to be and that the methods with which information is given should be fitting to the patient and not just a folder filled with medical terminology.

“In medical research involving human subjects capable of giving informed consent, each potential subject must be adequately informed of the aims, methods, sources of funding, any possible conflicts of interest, institutional affiliations of the researcher, the anticipated benefits and potential risks of the study and the discomfort it may entail, post-study provisions and any other relevant aspects of the study. The potential subject must be informed of the right to refuse to participate in the study or to withdraw consent to participate at any time without reprisal.

189 World Medical Association, The. WMA. “Declaration of Helsinki.” Adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964. Lastly editorially revised by the 64th WMA General Assembly, Fortaleza, Brazil, October 2013.

190 World Medical Association. (2017). “Declaration of Geneva.” 2nd paragraph.

191 World Medical Association, The. International Code of Medical Ethics from 1949. Adopted by the 3rd General Assembly of the World Medical Association, London, England, October 1949 and lastly amended by the 57th WMA General Assembly, Pilanesberg, South Africa, October 2006.

192 World Medical Association. (2013). “Declaration of Helsinki.” Paragraph 3. <https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/>. Last accessed on January 23rd, 2020.

Special attention should be given to the specific information needs of individual potential subjects as well as the methods used to deliver the information.”¹⁹³ The Declaration of Helsinki herewith acknowledges the distinction between medical practice and medical research and the shift in the patient/physician relationship.

3.2.2 *‘Informed consent’ and ‘assent’ for children, teens and physically or mentally incapacitated patients*

The whole chapter about ‘informed consent’ in the Declaration of Helsinki consists of eight paragraphs of which the one above is only the second and most elaborate. The following paragraphs specify special cases, such as when the patient is not capable of giving consent, as in children, teens or physically or mentally incapable persons. In these cases consent can be given by the parents or a “legally authorized representative.” Being deemed incapable of giving informed consent however does not necessarily mean that the patients cannot voice their will at all. If the patient is able to give to understand the given information and form an opinion about it, he or she is deemed able to give ‘assent,’ for example teenager or mentally incapacitated patients. The investigator should in these cases get the patients assent or dissent in conjunction with the consent by a legal guardian. And the researcher should feel to be beholden to it the patients views. Consent, especially of teenagers, is often obtained without their knowledge, because it is deemed as being sufficient if and when the guardians are giving consent.¹⁹⁴ Only when the patient is incompetent to give consent, either because he or she is in a coma or the handicap is too severe, then the investigator can forgo ‘assent’ but still has to obtain ‘informed consent’ by a legal guardian. The research has to “entail minimal risk and minimal burden”¹⁹⁵ and the reason why these patients need to be involved has “been stated in the research protocol and the study has been approved by a research ethics committee”.¹⁹⁶ Again, that point is extremely important, because during the Second World War, mentally incapacitated patients were deemed ‘unworthy to live’ [unwertes Leben] and were experimented on and killed without any ethical compunction by the researchers.

193 World Medical Association. (2013). “Declaration of Helsinki.” Paragraph 26.

194 World Medical Association. (2013). “Declaration of Helsinki.” Paragraph 29.

195 World Medical Association. (2013). “Declaration of Helsinki.” Paragraph 28.

196 World Medical Association. (2013). “Declaration of Helsinki.” Paragraph 30.

3.3 The Tuskegee Experiment

One rather spectacular case that transgressed the Declaration of Geneva and the Declaration of Helsinki is the Tuskegee Experiment.¹⁹⁷

The Tuskegee Experiment started in 1932, hence before the Declaration of Geneva and well before the first Declaration of Helsinki, but in the time of segregation between black and white people in the South of the United States. It lasted until 1972. So instead of the initially proposed six months, the experiment lasted for 40 years. The experiment was initiated in and around Macon County, Alabama and conducted at the Tuskegee University. It was designed by the Public Health Service¹⁹⁸ to observe the health effects of untreated syphilis in “black” men and was officially named “Tuskegee Study of Untreated Syphilis in the Negro Male.” 399 men with syphilis and 201 without the disease were admitted to the study. The men were not informed of their health status but were only told that they were treated for ‘bad blood’, a term used to describe many, otherwise unspecified, illnesses. They were given some type of treatment to alleviate the symptoms and they were subjected to regular medical check-ups. Even when penicillin was recognised and became available as a cure against syphilis in 1947, these men did not receive the new treatment.¹⁹⁹ The ethical and procedural problems of the Tuskegee experiment are manifold. First of all, the men randomised into the experiment were not informed of their health status and therefore could in no way give even some sort of informed consent to participate in the study. The benefits that were promised, like regular health check-ups, free meals and a burial insurance, coerced most of the men into participating since these were benefits which were otherwise unavailable for the participants. The area around Tuskegee was poor and the black population was even more so. The experiment was not designed to test a new drug to treat and/or cure syphilis but purely to show the overall effect of the disease if and when left untreated. In 1972 the experiment caused an outrage fuelled by different papers and was subsequently stopped a year later. An action - class law suit was filed and eventually settled out of court with a payment of 10 million Dollars. Survivors and their families received special benefits given to them by the government.²⁰⁰

197 Centres for disease control and prevention. “The Tuskegee Timeline.” <http://www.cdc.gov/tuskegee/timeline.htm> Last accessed on January 23rd, 2020.

198 The Public Health Service (PHS) is a primary division of the United States Department of Health and Human Services. <https://www.usphs.gov/>. Last accessed on January 23rd, 2020.

199 Tuskegee University. National Centre for Bioethics in Research and Health Care. <http://tuskegeebioethics.org/> The usphs syphilis study can be found on this website, together with the historical apology and the information how the University of Tuskegee is coping with the legacy of the study. Last accessed on January 23rd, 2020.

200 Allan Brandt. (1978). “Racism and Research: The Case of the Tuskegee Syphilis Study.” in *Hastings Center Report*, 8: 21–29.

Those physicians and medical officials participating in the experiment violated at least five of the eleven points of the Declaration of Geneva and almost all points of the Declaration of Helsinki. The two most important violations in both are “the health of my patient will be my first consideration” and, only in the Declaration of Geneva, “I will not permit considerations of age, disease or disability, creed, ethnic origin, gender, nationality, political affiliation, race, sexual orientation, social standing or any other factor to intervene between my duty and my patient.”²⁰¹ The Tuskegee Experiment included only African American men in lower class circumstances without much education. The health of the patients was not the overall concern of the researchers. They were merely interested in the disease and its progression. Most of the men with syphilis were chosen because they were in the second and latent stage of the disease. Some historians seem to think that this fact makes the experiment a little less questionable, because these men might not have benefitted from penicillin at all and they could not pass on the disease anymore. Realistically, syphilis in its second or latent stage is symptom free and, after about one year, cannot be passed on anymore. However, following the NHS syphilis guide, syphilis can successfully be treated in the latent stage with antibiotic medication, such as penicillin.²⁰² The excuse given by some historians and those researchers who proposed to continue the experiment therefore was not valid. Even though the disease in its second stage is tolerable to live with, if it moves into the third stage it is most often, if not always, deadly.

Basically these men were treated as a form of lab animal. Some were not yet in the latent stage or in the early latent stage, and did pass on the disease. All participants were discouraged to seek medical help outside the study. When the patients and the wider public were informed of the obvious maltreatment, the previously established trust between the researchers and the patients was irreversibly broken. And unfortunately that breaching of trust has its repercussion in the African American Community until today.

3.3.1 *Consequences of Tuskegee*

Most studies, especially in the USA where minority groups are so prevalent, try to have as wide a cross-section through the population as possible and therefore try to involve as many members of minority groups as possible. Although advertisements of trials are often especially designed as to reach these communities, participation is minimal. In almost every paper concerning itself with this problem, the

201 World Medical Association. (2017). “Declaration of Geneva.” Paragraph 5.

202 National Health Service, UK. NHS Choices: your health, your choices. <http://www.nhs.uk/Conditions/Syphilis/Pages/Treatmentpg.aspx> Last accessed on January 23rd, 2020.

Tuskegee Experiment is cited as the negative example for a study that is still prevalent in the minds of the people.²⁰³ Most minority groups in the USA seem to not know the difference between a clinical experiment and a clinical study and believe that they will be treated as guinea pigs for the white people.²⁰⁴ It might sound like semantics, but the difference between ‘experiment’ and ‘study’ is quite important. Experiments should be strictly confined to the laboratory setting and not involve human beings on the receiving end. Studies are conducted outside the laboratory and are so-called first-in-man studies.

3.4 ‘Informed consent’ in clinical research

One of the pledges of both Declarations that was so irretrievably broken in the Tuskegee Experiment, is “The health of my patient will be my first consideration.” In ‘normal’ clinical practice, this pledge seems to be easy enough to maintain. In clinical research, the health of ‘the patient as such’ is also the first consideration. But, since new procedures or medications are developed, researchers cannot be sure if the actual patient will benefit from their research. Even though they might believe their procedure to be more effective, it might not be and worse it might even be dangerous. The danger of adverse effects is one of the main risks patients need to be informed about. Today, clinical research is conducted in a way that differs widely from earlier ‘experiments’. It is under strict regulations and a trial involving human beings is only allowed after rigorous testing in the laboratory has happened. “Medical research involving human subjects must conform to generally accepted scientific principles, be based on a thorough knowledge of the scientific literature, other relevant sources of information, and adequate laboratory and, as appropriate, animal experimentation. The welfare of animals used for research must be respected.”²⁰⁵ Danger to the participants is supposed to be minimised as much as possible. And the participant has to understand the difference between clinical practice and clinical research and be aware that in a research setting his status has changed from a regular ‘patient’ to a ‘participant.’ Accordingly, the physician changes into a researcher, and both parties, participant and researcher should be aware of that and adjust their expectations accordingly.²⁰⁶ The researcher does not automatically act in the patients’ best interest anymore. He or she is well aware that there might be unknown dangers in the proposed procedure

203 Vicki S. Freimuth and Sandra Crousse Quinn, et.al. (2001). “African Americans’ views on research and the Tuskegee Syphilis Study.” In *Social Science and Medicine* 52: 2.

204 Vicky S. Freimuth and Sandra Crousse Quinn, et.al. (2001):

205 World Medical Association. “Declaration of Helsinki.” (2013). Paragraph 21.

206 Robert D. Truog. (1999). “Informed Consent and research design in critical care medicine.” in *Critical Care*. 3(3). and Manson and O’Neill.

or drug. Therefore it seems to be important that researcher and participant understand each other as two parts of a research team which is operating outside the usual clinical practice. If there is an actual benefit for the participant involved, then this benefit should be understood as a form of positive side-effect, but not as an expected and predictable outcome. Although clinical research by its very nature is aiming to produce positive effects, there can be no guarantee that it does so. For the actual research, negative effects are equally valid and informative, especially in preventing harm for future patients.

3.4.1 Use of already established data in and for research

In some cases, trials or observational studies have been performed but the data has not been used when it was amassed. The question which needs to be asked then is if the available data can be used for further research without the consent of the patients? And what to do with the data if consent is not possible to obtain or if the patients refuse to give it? This question can also pertain to health data that is amassed during regular practice. Is it allowable to use regular patient data, without the patients consent, for research? Just to make it clear, data here is never personal data which is in almost all cases strictly confidential. The data we are talking about here is anonymised data without disclosure about the actual patient. Following Goldacres' idea, discussed in a previous chapter, of a large randomised trial for different statins for example, it would be absolutely enough for the researcher to have basic, anonymised data available containing gender, age, and weight of the patient, the type of statins prescribed and the reason for the prescription.²⁰⁷ The report about positive and negative effects of the treatment and if other medication had been taken at the same time and potentially interacted with the statin would also be included. This information would be impossible to trace back to a single individual patient. The question would be if it would be necessary in this example to obtain informed consent? There is no agreed upon standard available in medicine to ultimately decide this question. Some argue that it in every case in which any kind of data is used, informed consent has to be acquired. Some argue that if and when the data is sufficiently anonymised it is allowable to use it for the greater good of the community and as part of the social welfare system. Especially were the adoption of health policies is concerned.²⁰⁸

207 Goldacre, Ben. (2012): 225.

208 Onora O'Neill. (2003). "Some limits of informed consent." in *Journal of Medical Ethics*. 29: 4-7.

3.4.2 *The issue of trust revisited*

Using data without the informed consent of the patient can lead to an overall mistrust of the population into the wider medical community. Manson and O'Neill argue that we "need to build and maintain trust as reason to demand informed consent for clinical care."²⁰⁹ One reason for this is, and here I am following Eyal, that the "value of mutual trust is a contributor to population health."²¹⁰ Her argument is that if there is an overall trust in the medical services within the population, then people are more willing to consult the medical community or to participate in medical trials. If there is an overall mistrust, as there is in the African American community, Eyal also mentions Tuskegee in her paper, then the overall willingness to seek treatment is lower and the overall population health is worse in comparison to more trusting communities. In order therefore to have patients as well informed and willing participants in a medical trial, there has to be an overall trust in the medical community. Eyal's argument sounds compelling, however, it does not provide a solution for establishing trust, other than saying that it needs to be won through information and explanations.

In western countries the medical community however seems to be losing the trust of the population, at least in parts. Many patients are dissatisfied with the overall medical treatment, suffer from the time constraints that most physicians are under and do not care about the medical information available. A case in point here is the overall popularity of homeopathic treatments and the refusal to vaccinate children.²¹¹ Especially parents with a higher education and a higher income increasingly refuse to follow the advice of the WHO for vaccinations. Reasons for this refusal are manifold. Some claim that the vaccine against mumps, measles and rubella causes autism. A claim initially made by Andrew Wakefield, MD, which has been debunked by the medical community and which led to Dr. Wakefield to not longer be allowed to practise medicine in the UK.²¹² Another reason which is often used is that these children's diseases are needed for the healthy build-up of the immune system and that vaccines are more dangerous than possible complications of these diseases. Again a myth which has been debunked by the medical community but which is widely upheld in lay circles.²¹³ To reestablish the trust in the medical community for these parents would first and foremost mean to actually

209 Neil C Manson and Onora O'Neill. (2007):

210 Nir Eyal. (2012): 1.

211 I have devoted an entire chapter to the problems and pitfalls of homeopathy.

212 Fiona Godlee, Jane Smith, and Harvey Marcovitch. (2011). "Wakefield's article linking MMR vaccine and autism was fraudulent." in *BMJ* 342.

213 Daniel A. Salmon, Matthew Z. Dudley, Jason M. Glanz, and Saad B. Omer. (2015). "Vaccine Hesitancy Causes, Consequences, and a Call to Action" in *American Journal of Preventive Medicine and Elsevier*.

reach them and to provide convincing information. But since medicine and health today has sort of reached the status of a religion, it is not on a scientific basis that those who do not believe in actual medicine can be reached. The problem of informed consent and trust in population health is therefore divided into two distinct problems. The first problem is to maintain the trust in those who use mainstream, or conventional medicine. Here information, based on evidence, is the key. The information must be complete, yet again that is why all data is needed, and needed to be published, to make a decision, and the physician has to aim at explanations of the available evidence in terms that are understandable to the lay person. The second problem is to build up trust in those who are sceptical of conventional medicine. Here evidential scientific information is not enough. It is still the basis which is needed to convince people but it seems as if the convincing has to happen on an emotional level. However, I fail to see a solution how those sceptics can be reached at all.

3.5 How to transmit medical information successfully

An important question that seems to me to be often overlooked in the literature about informed consent is how to actually achieve the ‘informed’ part. The guidelines are rather vague. The researcher is supposed to explain the procedure or medication, its benefits, side effects and risks and in case of a placebo-controlled trial, the researcher should also make clear that those receiving the placebo are actually not receiving any medication at all. Those who are deemed to be able to give consent are deemed to be able to fully understand that information and to make their decision accordingly. To reach the ‘informed’ part therefore, some type of exchange between the researcher and the participant has to have taken place.²¹⁴ This presupposes however, that both parties are speaking the same language and use and understand the same terminology. The researcher has to phrase the medical information in such a way that the lay person can make educated deductions and understand what he is told. The researcher also has to take the social background of the patient into account. If the information were couched in purely medical terms, then the lay person probably will gain little knowledge from the exchange and although it can be stated that he has been informed, it is doubtful if the consent that is subsequently given can be understood as ‘informed’ consent. The patient might agree out of trust, even though that trust is not based on information and therefore also not on the act of informing, but on a personal feeling. And personal feelings can also play a huge part in refusing a treatment, as in the case of those refusing vaccinations.

214 Neil Manson and Onora O’Neill.(2007): 42.

Manson and O'Neill argue that since the act of informing is based on communication, information can also be conveyed by non-verbal communication.²¹⁵ Non-verbal communication can lead to the establishing of trust without the necessity that the verbal communication is understood. This trust can lead to a form of selection bias that subsequently can influence the outcome of the trial. Informed consent, trust and selection bias on the part of the participant are therefore intertwined. Selection bias can lead the participant to the 'false' conclusion that the care he is receiving is superior to the standard care. This positive feeling can provoke a feeling of well-being in the patient, even though he might receive 'only' a placebo. However, selection bias can be prevented in trials through sufficient blinding of all participants. Since a blinded physician can only convey what he or she knows about the trial and not in which arm the participant is randomised into, even though the participant might be trusting, he or she cannot base the trust on any information that is not known.

The available information needs to be transmitted in a way that the patient or participant can understand it, can question it, and is subsequently able to base a decision on it. For that purpose the information should not only be written but also transmitted verbally so that the physician is able to make amendments to the information based on the patients/participants questions and reactions.

3.6 Conclusion

Informed consent is an important tool for both medical research and medical practice. It can and should be used as a patient/participant safe-guard and as a needed process in shared decision making. The information should be provided in written and verbal form so that the patient/participant has, in the former case, the time to read the information and to formulate questions, and in the latter case, to ask the questions and to seek clarifications.

However, the limits of informed consent should be acknowledged. There are cases, as we have seen, where informed consent would be deemed necessary but cannot be freely given because of limitations on the part of the patient. However, a lack of informed consent does not automatically prevent medical interference to, for example, save a patient in an emergency situation. A lack of informed consent should also not prevent the use of data for the establishing of health policies. The emphasis in this case should be on the use of *data*, not the use of humans as research objects. The latter case always and under any circumstances needs informed consent by the person. Informed consent therefore is an important tool in medical research which leads to greater understanding for both, patients and researchers.

215 Neil Manson and Onora O'Neill. (2007): 55.

In this way, informed consent is useful because it can promote knowledge, since the need to explain a procedure inevitably leads to greater understanding. And greater understanding can lead to the advance of medical knowledge. As we have seen however, medical research does not inevitably lead to greater medical knowledge. RCTs do not lead to knowledge, they lead to evidence and evidence can sometimes be severely flawed. Therefore, the next chapter will deal with the question why it is called evidence-based medicine and not knowledge-based medicine and where the differences are.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





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.

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, where 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 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 than 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 statement²⁴⁸ and the GRADE Working Group²⁴⁹ 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*. 1:2.

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 interest 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.

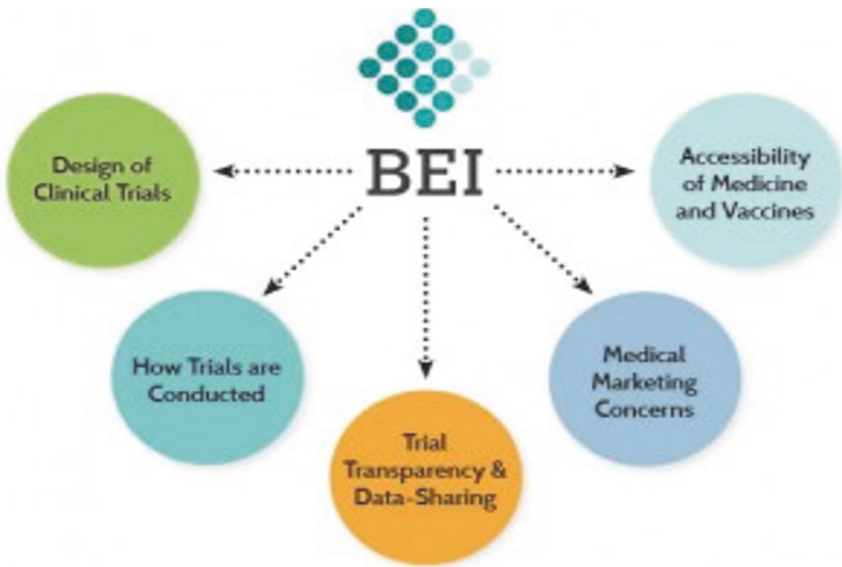


Figure 3: What the Good Pharma Scorecard wants to achieve and their points of access

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

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





5 Homeopathy — a case in point why EBM is so important — or, “the plural of anecdote is not data.”³⁰³

5.1 Alternatives to medicine?

EBM is by now *the* established way to conduct medical research and practice, sometimes more, sometimes less successful and always dependent on the user, both the physician and the patient. And although a lot of the criticism levelled at EBM can be refuted or EBM changed in a way that it is still true to its principles and still patient friendly, many patients are looking for alternatives to EBM and medicine in general. They are either dissatisfied by the way they are treated in hospitals and by conventional GPs, or they are slightly afraid of the treatments which can have side effects and whose ingredients lists contain long and hard-to-understand words. These dissatisfied patients look for other means to treat and cure their ailments and a whole market of ‘alternative’ medicines has sprung up to cater to these patients wants and needs. Foremost of these alternative treatments is homeopathy, closely followed by acupuncture. The treatments claim to be more ‘natural’ in that they do not use any harsh chemicals and that they are more gentle, taking the whole person into account. This chapter will mainly deal with homeopathy and its methodological problems as a case in point why EBM is so important and why we need modern science to heal and cure. If these ‘alternative’ treatments are effective, they belong in the realm of EBM, and if they are not, they are no alternative to EBM, but should be abandoned. Either something is a medicine or it is not. There is no such thing as an ‘alternative’.³⁰⁴

5.2 Historical context of homeopathy

The methodology of homeopathy was developed by Samuel Hahnemann (1755-1843) and is based on two principles. ‘Like cures like’ latinised by Hahnemann into ‘*similia similibus curantur*’, and the assumption that water retains memory of

303 Edzard Ernst and Simon Singh. (2008). *Trick or Treatment? The undeniable facts about alternative medicine*. London: Transworld Books.

304 Edzard Ernst. (2015). *A Scientist in Wonderland: A memoir of searching for truth and finding trouble*. Exeter: Imprint Academic.

the molecules it once contained.³⁰⁵ Hahnemann even considered the first principle, that ‘like cures like’ to be a law of nature.³⁰⁶ At the time Hahnemann came up with his homeopathic principles and remedies, regular medicine was not very successful in treating many conditions. Blood letting was still a very much accepted treatment and many patients died because of it.³⁰⁷ Many remedies commonly prescribed by doctors were poisonous, like lead or arsenic. Cocaine for example was considered a usable and successful treatment for pain and anxiety.³⁰⁸ Since homeopathic treatments usually do not contain any active ingredients, they could do no active harm compared to the many harmful treatments that were used. ‘Usually’ because some homeopathic treatments in very low potencies do contain active ingredients, but low potencies were and are seldom, if ever, prescribed. Patients that were treated by Hahnemann were treated compassionately and with time. They received good care and the body had time to heal itself. Naturally homeopathy looked very successful in comparison. And the comparison became even more favourable in cases of epidemics such as the cholera outbreak in London in 1854. At this point in time London already had a homeopathic hospital and the survival rate of cholera patients there was higher than in the regular hospitals. The homeopaths naturally argued that it was because of their treatments. However, the so called ‘heroic medicine’ used in the conventional hospitals, including treatments like blood-letting, was actively harming the patients. And the standards of hygiene and overall good food and cleanliness was higher in the homeopathic hospital as well.³⁰⁹ Doing nothing, in a clean environment, was in many cases much preferable than being subjected to some of the quackery of the time.

It was very important to Hahnemann that homeopathy was not equalled to herbal medicine, which he deemed dangerous because of the use of poisonous herbs and plants. Herbal medicine does contain active ingredients and is strictly plant based.³¹⁰ Hahnemann also insisted that his system of diagnosis and treatment was not allowed to be altered. Followers had to strictly adhere to his rules. He himself had come up with homeopathy by ingesting cinchona bark, a bark containing quinine used to cure, or at least treat, malaria. Since he was healthy at the time, but soon after ingesting the bark developed symptoms which he figured where like the symptoms of malaria, he surmised that what can trigger the symptoms in a healthy person can cure the same symptoms in a patient afflicted with

305 Edzard Ernst. (2016). *Homeopathy - The Undiluted Facts*. Switzerland: Springer: 9

306 Edzard Ernst. (2016): 23.

307 Ben Goldacre. (2009): 29.

308 L. Grinspoon, JB Bakalar. (1981). “Coca and cocaine as medicines: an historical review.” in *Journal of Ethnopharmacology*: 149-159.

309 Edzard Ernst. (2016): 29.

310 Edzard Ernst. (2005). “The efficacy of herbal medicine – an overview.” in *Fundamental & Clinical Pharmacology*, 19: 405–409.

the disease. So was born the ‘like cures like’ principle. He most probably only had an idiosyncratic adverse reaction, but believed the sensation to be genuine.³¹¹ His second assumption was that the more the treatment was diluted, the more potent it was. During Hahnemann's time, the knowledge about the atom or the molecule as the smallest possible unit of a chemical substance was only just being recognised. So Hahnemann could quite rightfully claim that very high dilutions of a substance, to the point where there were no molecules of the treatment in the original substance left, were possible because it was not known otherwise. However, to make the solution really potent in Hahnemann's view, it needed to be shaken vigorously. He called the process succussion and claimed that the memory of water is ‘triggered’ specifically by that method. Succussion means the banging of the bottle with the prepared tincture on a hard but yielding surface. Hahnemann ‘invented’ these surfaces by creating wooden boards covered with leather which were stuffed with horse hair, making them yielding enough to not break any glass vials.³¹² He came up with the idea while riding in a horse drawn carriage. “He believed that the vigorous shaking of the vehicle had further increased the so-called *potency* of his homeopathic remedies...”³¹³ However, so far no difference has ever been detected in the tincture before or after succussion. The alleged water memory cannot be shown, even though it has been repeatedly put under test. “The process of dilution and succussion is termed ‘dynamization’ or ‘potenization’ by homeopaths. In industrial manufacture this may be done by machine.”

Hahnemann believed that homeopathy was a true ‘alternative’ to regular health care, going so far as to claim that patients were not allowed to be treated by a homeopath and a regular doctor at the same time and even claiming that homeopaths who did not adhere to his rules were ‘traitors.’³¹⁴

Today there exist a number of different schools of homeopathy, each having a slightly different focus. The actual way that remedies are produced however and the focus on the patient as an individual who needs individualised treatment is largely the same for all schools. Therefore it is reasonable to look at an overall methodology of homeopathy, especially in comparison to the methodology of EBM.

5.3 The methodology of homeopathy

One big criticism towards EBM is that it is not holistic enough, that it loses, sometimes simply because of time constraints, the patient out of sight. Practicing homeopaths, regardless of school or inclination, claim to fill this gap. Homeopaths

311 Ben Goldacre. (2009): 30.

312 Ben Goldacre. (2009): 33.

313 Edzard Ernst and Simon Singh. (2008): 96.

314 Edzard Ernst. (2016): 12.

look at the entire patient, also taking the psychological well-being into account and design individualised treatments for each patient. A first visit to a homeopath can often last over an hour or even longer, in which the homeopath is exclusively listening to the patient and encourages the patient to keep on talking about symptoms, but also about their lives and problems.³¹⁵

Homeopathic remedies are supposed to be individualised to a degree that the same treatment might not treat the same symptoms in two different patients. And two different homeopaths might choose very different remedies for the same patient presenting with the same symptoms. All the treatments are listed in the ‘repertories.’ These consist of long lists of the different symptoms that are caused by the different remedies and which these remedies then are supposed to cure. Repertories have only been altered over time by the addition of new remedies. They have neither been questioned nor altered according to science.³¹⁶ They are available today for downloading, so every lay person can access them and since homeopathic remedies are available over the counter, can devise their own treatment, without ever having to see a homeopath.³¹⁷ The only reason why that approach is not in itself dangerous is precisely because there are no active ingredients in homeopathic remedies and hence consuming them, even the ‘wrong’ remedy or too much or too little, cannot lead to adverse effects. The ‘only’, but significant danger is that patients are delaying or foregoing life-saving treatment because they rely on homeopathy, and consequently harm themselves or those in their care.

‘Like cures like’, Hahnemann's first principle, means that the substance that can bring symptoms of a disease about in a healthy person can cure the ill person of that disease. For example, homeopathic red onion is used for curing watering eyes in a cold and Apis, made from bee venom is used against pain and swelling from bee stings.³¹⁸ The remedies should bring about an “‘artificial disease’ which would stimulate the patient’s *vital force*, which would in turn defeat the patient’s real disease.”³¹⁹

This approach however is not to be confused with vaccines which use the germs that cause a disease to immunise against that disease. These germs however

315 Natalie Grams. (2015). *Homöopathie neu gedacht: Was Patienten wirklich hilft*. Berlin, Heidelberg: Springer Verlag: 32.

316 Edzard Ernst. (2016): 133.

317 Kent Homeopathic repertory for Iphone. http://download.cnet.com/Kent-Homeopathic-Repertory/3000-2129_4-75978808.html. Last accessed on October 6th, 2017. Or: Complete Dynamics — Professional homeopathy. http://download.cnet.com/Kent-Homeopathic-Repertory/3000-2129_4-75978808.html. Last accessed on January 23rd, 2020. These two are only a demonstration of the possibilities to access homeopathic repertories on the internet.

318 Janko von Ribbeck. (2012). *Schnelle Hilfe für Kinder. Notfallmedizin für Eltern*. München: Kosel Verlag: 240.

319 Edzard Ernst. (2016): 10.

proponents [of homeopathy] claim that even when the last molecule is gone, a 'memory' of the original substance is retained."³²³

Avogadro's number is 6.023×10^{23} and is the number of atoms or molecules in a mole of any substance. A mole is the molecular weight of the substance in grams. Meaning that when anything is diluted beyond Avogadro's number, there are no more molecules of the original ingredient found in the substance. In a 30C solution, the number of molecules that one would need to consume to actually consume a molecule of the active ingredient would span the oceans of the globe. And again, so far all the research has not shown that water contains any lasting memory of any substance it ever touched. "Physicists have studied the structure of water very intensively for many decades, and while it is true that water molecules will form structures round a molecule dissolved in them at room temperature, the everyday random motion of water molecules means that these structures are very short-lived, with lifetimes measured in picoseconds, or even less. This is a very restrictive shelf life."³²⁴

Additionally, since the earth does have a closed water circuit, water would need to have a huge memory of all the molecules it ever touched. Homeopaths allege that the water memory is triggered by vigorously shaking the vial that contains the solution. They call it succussion, following Hahnemann. Yet again, it has never been scientifically proven that the shaking makes any difference. The homeopathic substance is either water or alcohol, nothing more.³²⁵ Most homeopathic remedies however are not given as water or alcohol, i.e. in their diluted form, but as little pills made from lactose, a milk sugar. These lactose pills are moistened with the diluted tincture and left to dry before they are given to the patient. Many placebo pills are made from lactose, since it is an easy to obtain and well-to digest compound. Only those patients being lactose-intolerant might have a reaction, but since the pills are so small, not even that is a real concern. So basically 90% of homeopathic remedies are just sugar with about 10% being water or alcohol.

Another important part of the methodology of homeopathy is the so-called 'proving'. Remedies are given to healthy individuals and they have to record their symptoms for a number of days or weeks. These symptoms are then compiled and if they appear in multiple people, the remedy is believed to treat these symptoms in sick patients. Provings are done with between 2 and 200 people and the symptoms are not checked for different causes. Homeopaths claim that these provings

323 Stephen Barrett. (2002) "Homeopathy's 'Law of Infinitesimals'". in HomeoWatch Home Page. <https://www.homeowatch.org/basic/infinitesimals.html>. Last accessed on January 23rd, 2020.

324 Ben Goldacre (2009): 35.

325 L. R. Milgrom. (2007). "Conspicuous by its absence: the Memory of Water, macro-entanglement, and the possibility of homeopathy." in Homeopathy 96(3): 209-219.

are sufficient to show that the remedy is working and that standardised medical trials, like RCTs are insufficient in the case of homeopathy, because homeopathic remedies are individualised to the patient and cannot be tested outside their specific context. The homeopath together with the patient finds the appropriate remedy. There are very limited remedies specific to one symptom which can be given as a fail-safe. Arnica is one of the remedies that is given frequently to treat cuts and bumps from falls, especially in children, and to reduce pain and swelling. A useful remedy can be obtained from the plant arnica which does help against muscle pain and eases bumps from falls when administered as a cream. However, when arnica is used in herbal medicine, the remedy actually contains arnica as an active ingredient. Therefore there is a difference between arnica in homeopathy and arnica in herbal remedy. The former is not part of medicine, the latter however is. Since arnica is often the first homeopathic remedy patients get acquainted with, and since the effects of arnica are generally well-known, the placebo-effect easily kicks in and people believe that the homeopathic arnica remedy has the same effect as the actual herbal remedy.

Some homeopaths are also providing homeopathic ‘vaccinations’ and are claiming that they are sufficient safeguard against disease. These so-called ‘nosodes’ are made from viruses and bacteria causing the disease in question. But instead of them being weakened, but able to trigger the immune system, like a usual vaccine, these ‘nosodes’ again contain the material in such a diluted form that they do not trigger any response in the body and therefore cannot prevent any diseases.³²⁶

5.4 Homeopathy today

Homeopathy was significant at the time of its founder, since it did herald better results than what was in retrospect called ‘heroic medicine.’ Due to the advances in sciences, in not small thanks in Europe to Hahnemann, homeopathy fell into decline and only appeared again around the 1970’s, the time of the flower-power and back to nature movement.³²⁷

Only very recently, in July 2017, the NHS in the UK has stopped funding and payment for homeopathy on the grounds that it has no curative potential.³²⁸ The NHS is continuously short of money and in order to save resources, it has finally

326 Edzard Ernst. (2016): 110.

327 Edzard Ernst. (2016):17.

328 NHS England. “NHS England launches action plan to drive out wasteful and ineffective drug prescriptions, saving NHS over 190 million pound a year.” <https://www.england.nhs.uk/2017/07/medicine-consultation/>. Last accessed on January 23rd, 2020.

decided to not longer pay for homeopathic and some other over-the-counter remedies. In Germany and Austria, some insurances pay for homeopathic treatments but are under scrutiny that they are doing so. Their main argument is that it is the wish of the patients to be treated by homeopaths and to take homeopathic remedies and that it is therefore for the benefit of all if the insurances cover the cost.³²⁹ Since my main topic is not the cost-efficiency of insurance companies, but the scientific value of medical interventions and their effectiveness, I will not pursue this angle of inquiry, but only remark on it.

Homeopaths often argue, as described above, that regular medical trials can just not be performed on homeopathic treatments, because the remedies are individualised, and differ from homeopath to homeopath and from patient to patient. However, many eminent homeopaths, among them Peter Fisher, have claimed that they have conducted, and that there are many trials performed, according to EBM standards, that show the efficacy of homeopathy, and that the trial data is freely available.³³⁰ Homeopaths here actively contradict themselves, often depending on which school they belong to or if they are lay homeopaths or medical doctors as well. The truth is that many studies have been performed and that only a very few show the required rigorousness to be EBM worthy, and have positive results. Most homeopaths “cherry pick” these to further their argument. However, they neglect that a much larger number of trials, equally rigorous, have shown no benefit of homeopathy other than placebo.³³¹ The rest of the available studies do not conform to the rigorous requirements of EBM. A few examples of flaws are that the patient base is much too small, the patients are not randomised, the homeopathic remedy is not tested against placebo or the evidence all together is just anecdotal. Anecdotal evidence is the evidence that is most cited with regard to homeopathy. However, when meta-analysis of these trials have been conducted, for example by Cochrane, none of these trials could, beyond a doubt, show the efficacy and effectiveness of homeopathy. Many of the trials were not rigorous enough, as described above and those that were rigorously conducted showed homeopathy to be not better than placebo. A large study was performed in 2015 in Australia, by the National Health and Medical Research Council (NHMRC) headed by Paul Glasziou, one of the eminent figures promoting EBM. He and his colleagues found that homeopathy does not have a discernible positive effect on any illnesses or diseases beyond the placebo effect.

329 Krankenkassen Deutschland. “Integrierte Versorgung mit klassischer Homöopathie.” <https://www.krankenkassen.de/gesetzliche-krankenkassen/leistungen-gesetzliche-krankenkassen/wahl-tarife-besondere-versorgung/integrierte-versorgung-homoeopathie/>. Last accessed on January 23rd, 2020.

330 PubMed search. Search words: Peter Fisher homeopathy trials. 21 hits. Last accessed on January 23rd, 2020.

331 Edzard Ernst. (2016): 56.

“Based on the assessment of the evidence of effectiveness of homeopathy, NHMRC concludes that there are no health conditions for which there is reliable evidence that homeopathy is effective.”³³² In Germany the “Informationsnetzwerk Homöopathie”³³³ is an internet-based information resource about homeopathy and is headed mostly by physicians who were homeopaths, but are now trying to educate the public about homeopathy and its pitfalls.

One argument that is often used in favour of homeopathy is that very small children and animals are reacting positively to it and that they can have no understanding of the placebo effect. However, it can be easily shown that this phenomenon is a simple case of observation bias.³³⁴ Because the observer, i.e. the parent or carer, is aware of the treatment, he or she influences the participant and since in the case of children and animals, the parents or owners normally report on the symptoms, their opinion is heavily influenced by their beliefs.³³⁵ Parents and owners would need to be blinded to the treatment to really form an educated opinion. Expectation and hope are the two important words here. Many parents are very afraid of conventional medicine to treat their children and rely on homeopathy as the more gentle way of treating illnesses. As long as only minor illnesses are treated with homeopathy and therefore the body essentially is left in peace to get well on its own, the use of homeopathy is not dangerous. However, it quickly becomes dangerous when ‘active’ treatments for infections are denied or significantly delayed leading to bodily harm.

Since the bodies capability of ‘regression to the means’, meaning to heal itself from many illnesses, it means that most of the time we approach a cure when the symptoms are at their very worst and then quickly get better on their own. Therefore we attribute whichever cure is given at the peak to have provided the cure. It is called the post-hoc-ergo-propter-hoc fallacy, and many patients using homeopathy and homeopaths are falling for it, because they confuse the concept of ‘cause and effect’ with ‘spurious correlations.’³³⁶

332 National Health and Medical Research Council, Australia. NHMRC Statement (2015). “Statement on homeopathy.” <https://www.nhmrc.gov.au/sites/default/files/images/nhmrc-statement-on-homeopathy.pdf>. Last accessed on January 23rd, 2020.

333 Informationsnetzwerk Homöopathie. <https://www.netzwerk-homoeopathie.eu/>. Last accessed on January 23rd, 2020.

334 Katja Weimer, Marco D. Gulewitsch, Angelika A. Schlarb, Juliane Schwille-Kiuntke, Sibylle Klosterhalfen & Paul Enck. (2013). “Placebo effects in children: a review.” in *Pediatric Research*, 74: 96-102.

335 David Ramey. (2008). “Is There a Placebo Effect for Animals.” in *Science Based Medicine: Exploring issues and controversies in the relationship between science and medicine*. <https://sciencebasedmedicine.org/is-there-a-placebo-effect-for-animals/>. Last accessed on January 23rd, 2020.

336 Steve E. Hartman. (2009). “Why do ineffective treatments seem helpful? A brief review.” in *Chiropractic & Osteopathy*.

However, even if the symptoms are not getting better right away when a homeopathic remedy has been taken, homeopaths have an explanation for this. They claim that one of the effects of homeopathic remedies is that they might, initially, make the symptoms worse and that that is a ‘normal’ effect of the remedies. So even if the illness or disease gets worse under the treatment, for homeopaths that does not seem to be a reason to take recourse to regular medicine.

As we have seen, homeopathic remedies do not contain any measurable active ingredient. Nonetheless scientists have conducted trials to establish if these remedies go beyond the placebo effect. However, all the trials done by scientists have not shown a discernible effect of homeopathic remedies above and beyond the placebo effect. The foremost research unit in the UK was at Exeter University, lead by Professor Edzard Ernst who, as a medical doctor had learned homeopathy in Germany as an established part of medicine and who was very interested in providing the evidence that homeopathy and other alternative medicines could be a part of medicine. However, during the course of his research, Ernst and his large team of scientists had to conclude over and over again that most ‘alternative’ treatments are no ‘alternative’ to evidence-based medicine but can only be understood as parts of pseudoscience.³³⁷ Since the ‘alternative’ community, with their figure-head the Prince of Wales was not happy with Ernst’s findings, the funding for his research unit was dried up and the unit ultimately closed down. Only in 2017 does the NHS seriously consider to drop homeopathy from its agenda as being a useless treatment regimen that only costs money without discernible benefits.

5.4.1 A defence of homeopathy — or is it?

The defence of homeopathy that is taken into account here is a defence formulated by practising medical doctors, who also are dealing with or researching homeopathy. Many scientists and authors who are dealing with homeopathy cannot, and will not, refute the amount of patients who feel better with homeopathic treatment and who believe in its effectiveness. Therefore some of them try to argue that homeopathy actually fulfils a need for patients. The main patient-friendly aspect of a consultation with a homeopath is ‘time’. A homeopathic first-time consultation can last up until three hours and homeopaths are trained in the art of active listening and encouraging the patient to talk about all parts of their lives, not just the symptoms that are bothering them.³³⁸ Active listening is something that the regular GP can hardly ever do because of the number of patients he has to see in a single day and because of the amount of paperwork that needs to be dealt with for each

337 Edzard Ernst. (2015): Chapter 7:Kindle Version: Off with his head.

338 Natalie Grams. (2015): 34.

patient. However, through active listening, a vital part also of psychotherapy incidentally, the patient feels that all his feelings are being taken seriously and that all experiences are valuable. He feels understood and often can phrase the problem at the end of a consultation.³³⁹ Being able to talk about the problems and having somebody who listens and is sympathetic is already an overall acknowledged part of the treatment-, and often, healing process. That is true for all medicine, but the homeopaths have a lot more time on their hands to really be there for the patient. Some homeopaths who have turned critical and acknowledge that there is nothing in the actual remedies, are still saying that the process of talking and interacting with the patient alone might be sufficient to cling to homeopathy as an area of medicine.³⁴⁰ However, if it wants to be a proper part of medicine, homeopathy has to forgo its main principles that ‘like cures like’ and that water has a memory and has to acknowledge that it cannot be treatment option based on actual medication, but only based on the positive homeopath/patient relationship. It would be worthwhile, according to Grams, to look closely at that route and to maybe establish homeopathy as an option of diagnostics, situated between EBM and psychotherapy and being able to send the patient in either direction, once it is established if there is a problem beyond bodily symptoms.³⁴¹

Another argument often formulated in favour of homeopathy is that even if the remedies and diagnostic sessions are just triggering a placebo response, it would still be worthwhile. It is even possible to measure a placebo response and see changes in the brain or measure how pain lessens, because of the patients expectations. The placebo response is often used to advantage in combination with an actual treatment. The expectation of the patient that a pain reliever will work is part of the process of administering such a treatment. The patient knows what he will receive and the body starts responding accordingly. For Ernst this is the ideal way to administer treatments, by utilising both, the real and the expected results.³⁴²

For homeopathy to fit into the realm of medicine, it has to not only follow the scientific rules that are prevalent in the natural sciences, it also has to maintain the same ethical standards as medicine, for example to ‘do no harm.’ The authors of a 2013 paper,³⁴³ writing about such an ethical defence of homeopathy suggest that “homeopathy is ethical as it fulfils the needs and expectations of many patients; may be practiced safely and prudentially; values care and the virtues of the thera-

339 Natalie Grams. (2015): 7.

340 Natalie Grams, et.al. Informationsnetzwerk Homöopathie. <https://netzwerk-homoeopathie.info>. Last accessed on January 23rd, 2020.

341 Natalie Grams. (2015):167.

342 Edzard Ernst. (2016): 49.

343 David Levy, Ben Gadd et.al. (2013) “A Gentle Ethical Defence of Homeopathy.” in Bioethical Inquiry.

peutic relationship; and provides important benefit for patients.” The authors however are not interested in the epistemological or scientific aspect of homeopathy, but only in its ethical aspect. Those who condemn homeopathy as ‘quackery’ are making, in their view, an error in ethical judgement, because even if it is quackery, it does help the people.

The authors claim, and as we have seen, rightfully so, that not all conventional therapy is based on definitive evidence as ‘prescribed’ in EBM. Even though that assessment is correct, there is a huge difference between treatments that are based on less rigorous evidence than RCTs and treatments that are not based on anything at all and cannot be measured because of the lack of active ingredients. Homeopathic treatments firmly belong in the latter category.

The authors go so far as to claim that “There is no ethical requirement for definitive explanations of mechanisms, knowledge of molecular effects, or epidemiological “proof” from large-volume RCTs for consent to any health care intervention to be valid, and the notion that the absence of these things makes homeopathy by definition deceptive, coercive, or unethical is morally and clinically incoherent.”³⁴⁴ The quote, in its essence, is correct. Often we do not know the mechanism of a treatment or its molecular effects, but due to rigorous tests we do know that these interventions either work or that they do not work. There is the possibility to detect an effect beyond the placebo-effect, be it positive or negative. In homeopathy there is no effect detectable in the treatments beyond the placebo-effect. Claiming however that homeopathic remedies work because they are ‘energised’ by succussion is not testable and therefore a form of deceit and arguably unethical. Since homeopaths themselves believe in the validity of their claims, it might be possible to not accuse them of unethical ‘behaviour,’ because they act according to the mistaken believe to really help the patient. However, that does not exonerate homeopathy as such.

Many medical interventions which we deem necessary today have at their very core the goal to heal the patient or to alleviate suffering by dealing with certain symptoms. However, EBM is not first and foremost a ‘feel-good’ medicine. Many remedies might not be pleasant in the short term, although they do help in the long run. Their risk/benefit ratio is in favour of the treatment. In homeopathy that is not always the case, especially if and when necessary treatment with ‘allopathic’ drugs, i.e. chemo therapy as cancer treatment, has been delayed or even forgone together. There are multiple cases of cancer patients who went to the hospital too late after having attempted homeopathic treatment first.³⁴⁵ Another risk might be the taking of the mother tincture or a low potency remedy. Since if it

344 David Levy, Ben Gadd, et.al. (2013): 5.

345 Skyler B. Johnson et.al. (2018). “Use of Alternative Medicine for Cancer and Its Impact on Survival.” in *Journal of the National Cancer Institute*. 110(1).

contains any poisonous material or actual germs, it can be dangerous and even, in the case of arsenic, deadly. The claim that homeopathy is entirely without risk is therefore not maintainable.³⁴⁶

5.5 Acupuncture

Acupuncture is an ancient treatment that involves thin needles that are inserted into the skin at certain, pre-specified, points. Acupuncture initially came from China. The needles are supposed to hit a life force, the Ch'i which flows through the body via so called meridians and these meridians can be manipulated through needling.³⁴⁷ By controlling the Ch'i, the body is supposed to go back into a kind of healthy balance. Since it was not allowed to conduct post-mortem examinations on human bodies in ancient China, the meridians were just assumed to be there and their number was established to be exactly twelve, like the number of main rivers in China.³⁴⁸ The Ch'i as a life-force was merely postulated. Since the needles were, and still are, only inserted in the outer layer of the skin, acupuncture was fairly safe, even though the needles could be contaminated, but most often were warmed before the treatment and the heat killed a lot of the bacteria.

Before the patient is 'needled' the acupuncturist will examine the patient according to five techniques "namely inspection, auscultation, olfaction, palpation and inquiring. Inspection means examining the body and face, including the colour and coating of the tongue. Auscultation and olfaction entail listening to and smelling the body, checking for symptoms such as wheezing and unusual odours. Palpation involves checking the patient's pulse: importantly, acupuncturists claim to be able to discern far more information from this process than any conventional doctor. Inquiring, as the name suggests, means simply interviewing the patient."³⁴⁹

Chairman Mao was the one to reinvent traditional Chinese medicine, because he wanted and needed affordable health care for everyone, and did not care if the medical system worked or not. Medicine needed to be home-grown instead of being expensive and influenced by the west. It is strikingly similar to the push of homeopathy in Germany in the 1930's and 1940's by the Nazis as being German medicine.³⁵⁰ If it would not have been for a political trip of President Nixon and before him Henry Kissinger and the journalist James Reston who suffered from appendicitis and was operated on, acupuncture would probably have vanished in the West. Reston received acupuncture after his operation to deal with abdominal

346 Edzard Ernst. (2016): 52.

347 Edzard Ernst and Simon Singh. (2008): 43.

348 Edzard Ernst and Simon Singh. (2008): 43.

349 Edzard Ernst and Simon Singh.(2008): 44.

350 Edzard Ernst. (2016): 95.

pain and the treatment helped him. He wrote an article about his experience in the New York Times and all of a sudden acupuncture was back on the map.³⁵¹ A number of years later, the Americans were naive enough to believe films about operations that were allegedly done without anaesthesia and solely using acupuncture as pain medication. Specialists however could and did see in the films that the patients must have received drugs beforehand to numb them and to deal with the pain. The needles were only window dressing. However, the interest in using acupuncture as pain management in surgery has resurfaced because of the overall costs of healthcare. So, today some clinics in Shanghai and mainland China are using acupuncture in addition to other pain medication but are operating without general anaesthesia.³⁵²

5.5.1 *Acupuncture under trial*

As with all medical treatments, acupuncture can be tested, and has been tested, with the EBM methodology. Since acupuncture does become increasingly more popular, even many GPs are interested in having scientific proof that the treatment works. A sham needle has been developed that looks and feels like a real acupuncture needle but that retreats back into the shaft, much like a stage dagger, when it is put on the skin. It therefore does not penetrate the skin, even though the sensation for the patient feels the same.³⁵³ These sham needles allow at least for the patient to be blinded as to which treatment he or she is receiving. The acupuncturist however is impossible to blind so that a certain amount of bias needs to be taken into account when interpreting the results.³⁵⁴ Still, unlike with homeopathy, the trial results concerning acupuncture are not universally negative. In some cases of pain, acupuncture is proven to work quite well, at least in the short term. Osteoarthritis of the knee is one of those areas where acupuncture, at least as a short term solution, shows to be effective.³⁵⁵ However, Ernst and Singh in their book claim that the more science and EBM advances, the less likely will it be that acupuncture will have positive results beyond a placebo response.³⁵⁶

The Cochrane Collaboration has produced multiple meta-analyses about acupuncture, but most of them show that either more evidence is needed or that the

351 A. White and E. Ernst. (2004). "A brief history of acupuncture." in *Rheumatology*, 43(5):662–663.

352 Zhou J, et al. (2011). "Acupuncture anaesthesia for open heart surgery in contemporary China." in *International Journal of Cardiology*. (150:1).

353 Elizabeth Tough, Adrain White, et.al. (2009). "Developing and validating a sham acupuncture needle." in *Acupunct Med*, BMJ. (27):118.

354 Edzard Ernst and Simon Singh. (2008): 81.

355 C Witt, B Brinkhaus, et.al. (2005). "Acupuncture in patients with osteoarthritis of the knee: a randomised trial." in *The Lancet*. 366: 136–43.

356 Edzard Ernst and Simon Singh. (2008): 85.

actual RCTs were not rigorous enough. Some conclude that acupuncture helps in the short term, but yet again that its long-term benefits are questionable. The overall results are inconclusive.³⁵⁷

As are the results of the WHO report on acupuncture of 2003. “Acupuncture: Review and analysis of reports on controlled clinical trials.”³⁵⁸ However, Ernst and Singh allege that the WHO report included incorrectly conducted trials and therefore lacked in quality control.³⁵⁹ They argue, quite strongly, that trials conducted in China should be excluded from the overall meta-analysis because their results were repeatedly too good to be true.³⁶⁰

5.5.2 *Acupuncture and safety*

Acupuncture has been shown, different to homeopathic treatments, to have, at least in some cases, an effect beyond placebo, at least as the evidence stands so far. It can help in cases of pain and nausea and other very minor ailments. For many unspecific pains the evidence is not yet conclusive.³⁶¹ Strict safety procedures however have to be adhered to. The needles must be sterile and only used once on each patient. The safety catch in acupuncture can be that not all practitioners use simple rules of hygiene. Unclean or non-sterile needles have already led to infections in patients and the incorrect use has also led to the puncturing of arteries and even to collapsed lungs.³⁶²

If it is done correctly and in an hygienic environment, and the expectations are curbed to its long-term effects, acupuncture does not lead to active harm, provided it is used only as an addition to other treatment. That, in effect, takes acupuncture, for the moment, out of the realm of alternative medicine and into the realm of EBM, even with the effectiveness caveats attached. However, as also Ernst and Singh are pointing out, there are many conventional treatments which help with pain and nausea which are proven beyond a doubt to actually work and which are in essence more cost effective and proven to be safe. Acupuncture sessions can cost up to 25 pounds or 30 Euros each and need to be administered repeatedly, because of their short term effect.³⁶³

357 Cochrane Evidence. <https://www.cochrane.org/search/site/acupuncture>. Last accessed on January 23rd, 2020.

358 WHO report on acupuncture. (2003). <http://www.acucentre.com.au/Clinic/WHOConditionsTx.pdf>. Last accessed on January 23rd, 2020.

359 Edzard Ernst and Simon Singh: (2008): 71.

360 Edzard Ernst and Simon Singh: (2008): 72.

361 Edzard Ernst and Simon Singh: (2008): 77.

362 Michael Stenger, Nicki Eithz Bauer and Peter B. Licht. (2013). “Is pneumothorax after acupuncture so uncommon?” in the *Journal of Thoracic Disease*. 5(4): 144–146.

363 Edzard Ernst and Simon Singh. (2008): 85.

5.6 Conclusion

The EBM methodology has shown beyond a doubt that homeopathy does not have an effect beyond the placebo response, if at all. The actual treatments, at least in the high potencies favoured by homeopaths, do not contain any active ingredients and the only measurable success is in the time given to the patient to express the ailments and to talk about their health and associated worries. The lack of active ingredients in the treatments lets homeopathy firmly fall outside the realm of medicine and is not a valid alternative to it. Some of its methodology in terms of patient interaction could fall into psychotherapy and even be adapted for conventional medicine, but again that has nothing to do with the homeopathic treatments. The rigorous methods of EBM have not been employed to discredit homeopathy, as alleged by those promoting homeopathy, but to ensure that any treatment is safe and effective in curing disease, or at least alleviating symptoms, to lead to a better quality of life for the patient. The same is true when EBM is employed to assess acupuncture, and although in the moment it looks as if acupuncture can help to alleviate certain symptoms and might be used in conjunction with conventional medicine, the actual verdict is still out if it really belongs in the realm of EBM or if it needs to be abandoned as a pure placebo treatment that is more expensive and more invasive, due to the needling, than conventional therapies for the same ailments.

EBM however could learn from patients, from their views and their opinions, and could incorporate it into its own methodology. The success of the ‘alternative treatment movement’ has little to do with the overall success of its pseudomedical treatments, but everything with the time and care that alternative practitioners offer their patients. Most practitioners are kind and caring, take time to listen and to understand. The patient ‘feels’ to be in good hands. In clinical medicine, the patient has to intellectually understand to be in good hands and to receive the best care possible. Often there is simply not enough time or energy on the part of the clinician to be as compassionate as the patient would wish for. If it would be acknowledged on both sides that the clinician cannot devote an hour to each patient every day, but that he at least listens to the immediate concerns and demonstrates to be looking for an answer, then maybe those who are trusting alternative medicine today will in future trust EBM even more.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.





6 Conclusion

Regardless of all the criticism and regardless of the multitude of propositions for a replacement of EBM, it seems certain that EBM, simply because it is medicine based on solid scientific evidence, is the best method we have to save lives and to treat the sick and infirm. Evidence-based research is what gives us credible results, and physicians who are using those, together with their experience, are the ones providing their patients with the best care possible. Mistakes can happen and EBM is no safeguard against mistakes. Humans use EBM and guide EBM and make decisions based on EBM, but they are still human. The further we can go in medicine, the bigger it seems, are the mistakes that can happen, but the greater are also the triumphs that medical research can achieve. This does not mean that EBM, and especially research is allowed to do and research whatever it chooses. Ethical questions and qualms still need to be answered and EBM has to acknowledge, the same as medicine at all times, where it is not prudent to continue with research, because it would not be for the societal good but for personal, political or economical gain. But, there are cures available now for many diseases that were deadly. And disease prevention, like vaccines, have played a huge role in eliminating many debilitating diseases. The infant mortality rate has declined significantly and the overall population, at least in the developed world, gets ever older, and many of the elderly are in good, or at least stable health, being able to enjoy their life. There are many reasons that contribute to this trend, and a functioning and successful medical system is one of those reasons.

However, EBM has to change to maintain its stronghold. The EBM community is already refocusing on the individual patient in medical practice and is acknowledging that statistical data is not in itself sufficient to lead to an individualised treatment. Therefore it is imperative that EBM is divided into 'evidence-based research' and 'evidence-based practice.' Both are based on evidence, but both acknowledge that the evidence they are using or trying to achieve might be slightly different. Research evidence is robust, quantifiable, reproducible and internally valid, if the research is done correctly. It has a statistical relevance and is usable on the population level. An important caveat here is that the research results also need to be made available. The most devastating bias for EBM is publication bias, since data that is not shared cannot be evaluated. When data is deliberately not shared it is impossible to assess trial results objectively and to even discover if any other form of bias might have influenced the results. In order for research results

to be usable, publication bias needs to be eliminated so that all data is accessible at any given time.

And not only, but partially due to publication bias, research evidence is almost never automatically externally valid. To achieve external validity, evidence for medical practice needs to be informed by more than the stringent criteria used in research. The patients wishes, values and concerns need to play a role, as well as the expertise and knowledge of the physician. The latter needs to be able to distinguish if ‘good’, as in robust evidence overall, is also the best evidence for the patient seeking treatment. Because of age or multi-and co-morbidities, adjustments in the treatment might be necessary which are not part of official guidelines or research recommendations. Results of ‘lesser’ studies, like observational studies can also play a vital role in the medical practice, even though their results are not deemed as robust as those of randomised controlled trials.

Evidence hierarchies need to be questioned and evidence must be made ‘digestible’ for the evidence user. This ‘user’ is not simply the physician, clinician, or researcher anymore, but the patient as well. For medical practice the hierarchy should be flattened out so that ‘all’ evidence is assessed for the single patient. The paternalistic medical order has changed to a ‘partnership’ between physician and patient, at least in an ideal setting, where both parties are aware of the evidence and are allowed to question if the ‘best’ evidence overall is still the best ‘evidence’ for that particular patient. Where such a partnership is not possible, usually the necessary trust between the patient and the physician breaks down and a meaningful consultation and treatment are not possible. So medicine should be based on trust and solid evidence, and the knowledge how to use it and share its results are what makes medicine successful.

To not undermine this trust, the patient and the physician have to be aware that in ‘evidence-based research’ they are changing into participant and researcher. In the research setting ‘informed consent’ plays an even bigger role than in the practice setting. The risk to the participant can be quite significant, and the participant has to be made aware of this. Here again, as in almost all areas of EBM, the tool of informed consent is not a fail-safe to avoid all flaws, faults and mistakes. And it has significant ethical drawbacks where those patients are concerned that are not able to freely consent to a procedure. But it can provide some sort of security to those participants that can freely give consent, insofar as that all information is made available and the participant can decide to be a part of the trial and he or she can decide at any point in the trial to withdraw. The same is true for evidence-based practice. The patient can decide to accept or refuse a treatment or to seek a second opinion. Informed consent is the tool, next to shared decision making, which is imperative to a functioning and successful partnership between patient and physician.

Evidence hierarchies can be useful for medical research because research questions and proposed trials can already be classified according to their ranking. And it can be assessed if a research question can be answered by randomised controlled trials or if other forms might be sufficient to answer specific questions, especially where treatments and not drugs are concerned. But as important as evidence hierarchies can be for medical research, it needs to be acknowledged that what they are producing is not medical knowledge. EBM is deliberately not called knowledge-based medicine because evidence is fluid and changeable and ‘good’ can be distinguished from ‘bad’ evidence. Evidence means ‘good reason for belief’ and does not contain a truth element as ‘knowledge’ does. Medical knowledge is what the physician, clinician and researcher should possess since it entails their expertise and their tacit knowledge as so-called soft skills which are needed to render the available evidence useful for medical practice. However, these skills are not and should not be the reason that EBM is sometimes claimed to be a new theory of epistemology in medicine. As we have seen, EBM draws on different theories of epistemology but in order to establish a new theory, such a theory would need to hold in other areas of science as well, and the methodology of EBM is not simply transferable to other sciences. The focus should therefore not be on shaping EBM into a theory of epistemology but on producing better evidence for better health care.

EBM does not have valid alternatives. EBM is not only better than person-centred medicine, narrative-based medicine or any of the other forms that try to replace it, although much of what these have to offer should and easily could be included in EBM since most proposals aim at making EBM more person-centred which is of vital necessity for evidence-based practice. But EBM is also eminently superior to ‘alternative’ medicines, such as homeopathy and acupuncture. Both have scientific and methodological flaws which EBM can easily detect but not solve, since both alternatives are not based on a sound scientific basis. The ‘evidence’ that is used is merely anecdotal and although I proclaim that in evidence-based practice many forms of available evidence should be used to help, heal and cure the patient, anecdotal evidence about unscientific treatments should not be part of the medical knowledge that is conveyed in an EBM setting. What EBM can learn from alternative treatment options is to invest more time in the patient and in the shared decision making process to regain and maintain the trust in the system.

Lastly, EBM cannot be understood as a ‘new’ Kuhnian paradigm, nor as having triggered a paradigm shift, but as the inevitable response to rapid change in medical science and understanding. EBM therefore is ready to be challenged, but not as a paradigm but as a form of medicine based on scientific principles which needs to adapt and diversify in order to put the patient back into the centre of medicine, in both research and practice.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder..



Bibliography

- Achinstein, Peter. (2010). *Evidence, Explanation, and Realism: Essays in the Philosophy of Science*. Oxford: Oxford University Press.
- Agency for healthcare research and quality. AHRQ. www.ahrq.gov. Last accessed on January 23rd, 2020.
- Agich, George, W. (2005). "What kind of doing is Clinical Ethics?" in *Theoretical Medicine*. 26.
- Alzheimer's Research UK. "General anaesthesia linked to increased dementia risk." Published online May 31st, 2013. <http://www.alzheimersresearchuk.org/general-anaesthesia-linked-to-increased-dementia-risk/>. Last accessed on October 15th, 2017.
- AllTrials campaign. <http://www.alltrials.net>. Last accessed on January 23rd, 2020.
- Ansaloni, Luca, Fausto Catena, and Ernest E Moore. (2007). "WJES and case reports/case series." in *World Journal of Emergency Surgery*. 2(11).
- Ashcroft, R.E. (2004). "Current Epistemological Problems in Evidence Based Medicine." in *Journal of Medical Ethics*. BMJ Publishing Group. 30(2): 131-135.
- Austin, Peter C. (2011). "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." in *Multivariate Behavioral Research*: 399-424.
- Barenboim, Elias., Brito, Carlos., Pearl, Judea.(2011). "Local Characterizations of Causal Bayesian Networks." Technical Report. R-384 in M. Croitoru, O. Corby, J. Howse, S. Rudolph, and N. Wilson (Eds.) *Proceedings of the Second International IJCAI Workshop on Graph Structures for Knowledge Representation and Reasoning (GKR 2011)* 1-6.
- Barrett, Stephen. (2002). "Homeopathy's Law of Infinitesimals." in *HomeoWatch Home Page*. <https://www.homeowatch.org/basic/infinitesimals.html>. Last accessed on January 23rd, 2020.
- Bartlett, R.H., A.B. Gazzaniga, et.al. (1986). "Extracorporeal membrane oxygenation (ECMO) in neonatal respiratory failure: 100 cases." in *Annals of Surgery*. 204(3): 236-245.
- Bassler, Dirk, Victor M. Montori, et. al. (2008). "Early stopping of randomized clinical trials for overt efficacy is problematic." *Journal of Clinical Epidemiology*. 61: 241-246.
- Beauchamp, Tom L., and James F. Childress.(2009). *Principles of Biomedical Ethics*. Sixth Edition. Oxford: Oxford University Press.
- Beecher, Henry K. (1995). "The Powerful Placebo." in *JAMA*: 1602-1606.
- Beller, Elaine, Paul Glasziou, Iain Chalmers. et. al. (2013). "PRISMA for Abstracts: Reporting Systematic Reviews in Journal and Conference Abstracts." in *PLOS Medicine*.
- Benedetti, Fabrizio. (2014). *Placebo Effects*. 2nd Ed. Oxford: University of Oxford Press.

- Bennett, Charlotte., Ann Johnson, David J Field, Diana Elbourne. (2001). "UK collaborative randomised trial of neonatal extracorporeal membrane oxygenation: follow-up to age 4 years." for the UK Collaborative ECMO Trial Group. *The Lancet*. 357.
- Benson, Herbert. (1997). "The Nocebo Effect: History and Physiology." in *Preventive Medicine*. 26(5): 612-615.
- BetterEvaluation: Sharing information to improve evaluation. <http://betterevaluation.org>. Last accessed on October 15th, 2017.
- Bioethics International. The Good Pharma Scorecard. <http://bioethicsinternational.org/good-pharma-scorecard-overview/>. Last accessed on January 23rd, 2020.
- Bird, Alexander. (2013). "Thomas Kuhn." in *The Stanford Encyclopaedia of Philosophy*, Ed. Edward N. Zalta. <https://plato.stanford.edu/archives/fall2013/entries/thomas-kuhn/>. Last accessed on January 23rd, 2020.
- Bluhm, Robyn. (2005). "From Hierarchy to Network: a richer view of evidence for evidence-based medicine." in *Perspectives in Biology and Medicine*. 48(4): 535-547.
- Bluhm, Robyn. (2010). "The epistemology and ethics of chronic disease research: Further lessons from ECMO." In *Theoretical Med Bioeth*. 31:107-122.
- Bluhm, Robyn 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.
- Botting, Jack. (2002). "The History of Thalidomide" in *Drug News & Perspectives*. 15(9).
- Boutron, I, Estellat, C. Guittet, L. Dechartres, A. Sackett, DL. et.al. (2006). "Methods of Blinding in Reports of Randomized Controlled Trials Assessing Pharmacologic Treatments: A Systematic Review." in *PLoS Medicine*, 3(10): 1931-1939.
- Brandberg, Yvonne. Hemming Johansson, Mia Bergenmar. (2016). "Patients' knowledge and perceived understanding —Associations with consenting to participate in cancer clinical trials." in *Contemporary Clinical Trials Communications* 2: 6-11.
- Brandt, Allan. (1978). "Racism and Research: The Case of the Tuskegee Syphilis Study." in *Hastings Centre Report*. 8: 21-29.
- Brody, Howard and Franklin Miller. (2003). "The Clinician Investigator: Unavoidable but Manageable Tension." in *Kennedy Institute of Ethics Journal*. 13(4): 329-346.
- Brody, Howard and Franklin Miller. (2009). "What makes placebo-controlled trials unethical?" in *American Journal of Bioethics*. 2(2): 3-9.
- Cartwright, Nancy. (1999). *The Dappled World: A Study of the Boundaries of Science*. Cambridge University Press.
- Cartwright, Nancy. (2007). *Hunting Causes and Using them: Approaches in Philosophy and Economics*: Cambridge University Press.
- Cartwright, Nancy. (2007). "Are RCT's the Gold Standard?" in *BioSocieties*. 2:11-20.
- Cartwright, Nancy. (2009). "Evidence-Based Policy: What's to Be Done about Relevance? For the 2008 Oberlin Philosophy Colloquium." in *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition*. 143(1): 127-136.
- Cartwright, Nancy. (2011). "A philosopher's view of the long road from RCTs to effectiveness." in *The Lancet*. 377(9775): 1400 - 1401
- Cartwright, Nancy. (2009). "Causality, Invariance, and Policy." in *The Oxford Handbook of Philosophy of Economics*. Edited by Don Ross and Harold Kincaid. Oxford: Oxford University Press.

- Cartwright, Nancy and Jacob Stegenga. (2008). "A Theory of Evidence for Evidence-Based Policy." for the Centre for the Philosophy of Natural and Social Science Contingency and Dissent in Science Technical Report 08/08. Published by the Contingency And Dissent in Science Project Centre for Philosophy of Natural and Social Science. The London School of Economics and Political Science.
- Castro, Mario. (2007). "Placebo versus Best-Available-Therapy Control Group in Clinical Trials for Pharmacologic Therapies. Which Is Better?" in Proceedings of the American Thoracic Society, 4(7): 570–573.
- CEBM, Centre for Evidence-Based Medicine, Oxford. "Definition of ARR and NNT." <http://www.cebm.net/number-needed-to-treat-nnt/>. Last accessed on January 23rd, 2020.
- Centres for disease control and prevention. "The Tuskegee Timeline." <http://www.cdc.gov/tuskegee/timeline.htm>. Last accessed on January 23rd, 2020.
- Chavez-MacGregor, Mariana and Sharon H. Giordano. (2016). "Randomized Clinical Trials and Observational Studies: Is There a Battle?" in Journal of Clinical Oncology. (34)8: 772-773.
- Chintu, Bat, Walker, et al. (2004). "Co-trimoxazole as prophylaxis against opportunistic infections in HIV-infected Zambian children (CHAP): a double blind placebo-controlled randomized trial." in The Lancet, 364.
- Clarke, Brendan, Donald Gillies, Phyllis Illari, Federica Russo and Jon Williamson. (2013). "The evidence that evidence-based medicine omits" in Preventive Medicine, 57(6): 745-747.
- Clarke, Brendan, Donald Gillies, Phyllis Illari, Federica Russo and Jon Williamson.(2014). "Mechanisms and the Evidence Hierarchy." in Topoi: 33(2): 339-360.
- Clifton-Soderstrom, Michelle. (2003). "Levinas and the Patient as Other: The Ethical Foundation of Medicine." in Journal of Medicine and Philosophy, Vol. 28:4.
- Cochrane Collaboration. <http://www.cochrane.org>. Last accessed on October 14th, 2017.
- Cochrane Collaboration. Our Logo. <http://www.cochrane.org/about-us/our-logo>. Last accessed on January 23rd, 2020.
- Cochrane Evidence. <https://www.cochrane.org/search/site/acupuncture>. Last accessed on January 25th, 2020.
- Comoretto, Nunziata.(2012). "Rethinking the Therapeutic Obligation in Clinical Research." in Chapter 6, J. Schildmann et. al. (eds.) *Human Medical Research*. Basel: Springer.
- Complete Dynamics — Professional homeopathy. http://download.cnet.com/Kent-Homeopathic-Repertory/3000-2129_4-75978808.html. Last accessed on October 6th, 2017.
- Consort Statement, The. <http://www.consort-statement.org>. Last accessed on January 23rd, 2020.
- Cook, DJ, Sackett DL, Spitzer, WO. (1995). "Methodological Guidelines For Systematic Reviews of Randomized Control Trials in Health Care from the Potsdam Consultation on Meta-Analysis." in the Journal of Clinical Epidemiology. 48(1): 167-171.
- Cornu, Catherine, et al. (2013). "Experimental designs for small randomised clinical trials: an algorithm for choice." in Orphanet Journal of Rare Diseases. 8(48):1-12.
- Darlenski, RB., et.al. (2010). "Evidence-based medicine: Facts and controversies." in Clinics in Dermatology, Elsevier.

- Djulgovic, Benjamin, Gordon Gyatt, and Richard Ashcroft. (2009). "Epistemologic Inquiries in Evidence-Based Medicine." in *Cancer Control: Journal of the Moffitt Cancer Centre*. 16(2):158-168.
- Donovan, Jenny L., Paramasivan, Sangeetha, de Salis, Isabel., Merran Toerien. (2014). "Clear obstacles and hidden challenges: understanding recruiter perspectives in six pragmatic randomised controlled trials." in *BioMed Central. Trials*. 15(5):1-12.
- Douglas, Heather. (2009). *Science, Policy, and the Value-Free Ideal*. Pittsburgh: University of Pittsburgh Press.
- Dowswell, T., G. Carulli, L. Duley, S. Gates, AM. Gülmezoglu, D. Khan-Neelofur, GGP. Piaggio (2010). "Alternative versus standard packages of antenatal care for low-risk pregnancy." *Cochrane Database of Systematic Reviews Issue 10*.
- Dworkin, Gerald. (1988). *The theory and practice of autonomy*. Cambridge: University Press.
- Dworkin, Gerald. (2010). "Paternalism", in *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta, ed. <http://plato.stanford.edu/entries/paternalism/>. Last accessed on April 30, 2014.
- Elbourne D, Field D, Mugford, M. (2002). "Extracorporeal membrane oxygenation for severe respiratory failure in newborn infants." in *Cochrane Database of Systematic Reviews Issue 1*.
- Emerson, Scott S. (1995). "Stopping a Clinical Trial Very Early Based on Unplanned Interim Analyses: A Group Sequential Approach." in *Biometrics*. 51(3): 1152-1162.
- Ernst, Edzard. (2005). "The efficacy of herbal medicine – an overview." in *Fundamental & Clinical Pharmacology*, 19: 405–409.
- Ernst, Edzard and Simon Singh. (2008). *Trick or Treatment? The undeniable facts about alternative medicine*. London: Transworld Books.
- Ernst, Edzard. (2015). *A Scientist in Wonderland: A memoir of searching for truth and finding trouble*. Kindle Version. Exeter: Imprint Academic.
- Ernst, Edzard. (2016). *Homeopathy The Undiluted Facts*. Switzerland: Springer Publishing.
- European Society for Person Centred Health Care. www.pchealthcare.org.uk. Last accessed on January 23rd, 2020.
- Evans, Imogen. Hazel Thornton. Iain Chalmers and Paul Glasziou. (2011). *Testing Treatments. Better Research for Better Healthcare*. 2nd. Edition. London: Pinter & Martin.
- Eyal, Nir. (2012). "Using informed consent to save trust." in *Journal of Medical Ethics*.
- Eyal, Nir. (2012). "Informed Consent" in *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta, ed. plato.stanford.edu/archives/fall2012/entries/informed-consent/ Last accessed on January 23rd, 2020.
- Eyding, Dirk, 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*.
- Feinstein, Alvan. Horwitz, Ralph. (1997). "Problems in the 'Evidence' of Evidence-based Medicine" in *The American Journal of Medicine*, 103: 529-535.
- Finnis, Damien G., Ted Kaptchuk, Franklin Miller, Fabrizio Benedetti. (2010). "Biological, clinical, and ethical advances of placebo effects." in *The Lancet*; 357: 686-695.

- Finsterer, Josef. and Salma Wakil. (2015). "Mitochondrial Depletion Syndromes." in eLS: 1–9.
- Floyd, Anna and Anne Moyer. (2009). "Equipose may be in the Eye of the Beholder." in *The American Journal of Bioethics*. 9(2): 21-22.
- Fodale, V. and L.B. Santamaria, et.al. (2010). "Anaesthetics and postoperative cognitive dysfunction: a pathological mechanism mimicking Alzheimer's disease." in *Anaesthesia*, 65: 388-395.
- Food and Drug Administration: "Guidelines for clinical trial sponsors." <https://www.fda.gov/downloads/regulatoryinformation/guidances/ucm127073.pdf>. Last accessed on January 23rd, 2020.
- Food and Drug Administration: Guidelines for DMC's. "The Establishment and Operation of Clinical Trial Data Monitoring Committees for Clinical Trial Sponsors." <https://www.fda.gov/RegulatoryInformation/Guidances/ucm127069.htm>. Last accessed on January 23rd, 2020.
- Forero, Roberto. Geoff McDonnell, Blanca Gallego, et.al. (2012). "A Literature Review on Care at the End-of-Life in the Emergency Department." in *Emergency Medicine International*.
- Fortin, Martin, et.al. (2006). "Randomized Controlled Trials: Do They Have External Validity for Patients With Multiple Comorbidities." in *Annals of Family Medicine*. 4(2).
- Freedman, Benjamin. (1978). "Equipose and the ethics of clinical research." in *The New England Journal of Medicine*, 317 (3): 141-145.
- Freimuth, Vicki S., Sandra Crouse Quinn, et. al. (2001). "African Americans' views on research and the Tuskegee Syphilis Study" in *Social Science and Medicine*. 52: 797-808.
- Freyhofer, Horst H. (2004). *The Nuremberg Medical Trial: The Holocaust and the Origin of the Nuremberg Medical Code (Studies in Modern History)*. New York: Peter Lang Publishing.
- Fries, James F., Krishnan, Eswar. (2004). "Equipose, design bias, and randomized controlled trials: the elusive ethics of new drug development." in *Arthritis Research & Therapy*. 6 (3): 250-255.
- Gabbay, John and Andrée LeMay. (2004). "Evidence based guidelines or collectively constructed 'mindlines'? Ethnographic study of knowledge management in primary care." in *BMJ* 329.
- Gabbay, John and Andrée LeMay. (2011). *Practice-Based Evidence for Health Care: Clinical Mindlines*. London: Routledge.
- Gadamer, Hans-Georg. (2003). *Schmerz: Einschätzungen aus medizinischer, philosophischer und therapeutischer Sicht*. Heidelberg: Universitätsverlag Winter.
- Garrison, Louis P., et.al. (2007). "Using Real-World Data for Coverage and Payment Decisions: The ISPOR Real-World Data Task Force Report" in *Value in Health*. 10 (5).
- Gettier, Edmund. (1963). "Is Justified True Belief Knowledge?" in *Analysis* 23(6): 121-123.
- Glannon, Walter. (2014). "Anaesthesia, amnesia and harm." in *Journal of Medical Ethics*: 1-7.
- Glasziou, Paul. (2006). "Why is evidence-based medicine important?" in *EBM notebook*. 11: 133-135.
- Glasziou, Paul. et.al. (2007). "When are randomised trials unnecessary? Picking signal from noise." *Analysis*. *BMJ* Vol. 334: 349-351.

- Glasziou, Paul. et.al. (2013). "Poor description of non-pharmacological interventions: analysis of consecutive sample of randomised trials." in *BMJ* Vol. 347: 1-10.
- Goldacre, Ben. (2009). *Bad Science*. London: Fourth Estate, Harper Collins Publishers.
- Goldacre, Ben. (2012). *Bad Pharma: How medicine is broken, and how we can fix it*. London: Fourth Estate, Harper Collins Publishers.
- Goldenberg, Maya J. (2006). "On evidence and evidence-based medicine: Lessons from the philosophy of science." in *Social Science & Medicine* 62: 2621-2632.
- Goldet, Gabrielle, Howick, Jeremy. (2013). "Understanding Grade: an introduction." in *Journal of Evidence-Based Medicine*: 50-54.
- Goodyear, Michael. (2006). "Learning from the TGN1412 trial." in *BMJ*: 1-2.
- GRADE working group. From evidence to recommendations — transparent and sensible. <http://www.gradeworkinggroup.org/>. Last accessed on January 23rd, 2020.
- Grady, Christine. (2005). "Payment of clinical research subjects." in *Journal of Clinical Investigation*. 115 (7): 1681-1687.
- Grams, Natalie. (2015). *Hömoöpathie neu gedacht: Was Patienten wirklich hilft*. Berlin, Heidelberg: Springer Verlag.
- Greene, Brian. (2003). *The Elegant Universe: Superstrings, Hidden Dimensions, and the Quest for the Ultimate Theory*. New York: Vintage Books.
- Greenhalgh, Trisha. (2014). *How to read a paper: The Basics of Evidence-Based Medicine*. Oxford, New York: Wiley and BMJ Books.
- Greenhalgh, Trisha. (1999). "Narrative based medicine: Narrative based medicine in an evidence based world." in *BMJ* 318.
- Greenhalgh, Trisha. (2012). "Why do we always end up here? Evidence-based medicine's conceptual cul-de-sacs and some off-road alternative routes." in *Journal of Primary Health Care*. (4): 92-97.
- Greenhalgh, Trisha, Jeremy Howick, Neal Maskrey for the Evidence Based Medicine Renaissance Group. (2014). "Evidence based medicine: a movement in crisis." in *BMJ*: 1-7.
- Grinspoon, L. and JB Bakalar. (1981). "Coca and cocaine as medicines: a historical review." in *Journal of Ethnopharmacology*: 149-159.
- Godlee, Fiona, Jane Smith, and Harvey Marcovitch. (2011). "Wakefield's article linking MMR vaccine and autism was fraudulent." in *BMJ* 342.
- Guideline Central. <http://www.guidelinecentral.com/summaries/>. Last accessed on January 23rd, 2020.
- Guyatt, Gordon. et.al. for the Evidence-Based Medicine Working Group. (1992). "Evidence Based Medicine: A New Approach to Teaching the Practice of Medicine." in *JAMA*, 268(17): 2420-2425.
- Guyatt, Gordon. et.al. for the Evidence-Based Medicine Working Group. (2000). "Users' Guide to the Medical Literature: XXV. Evidence-Based Medicine: Principles for Applying the Users' Guides to Patient Care." in *JAMA* 284(10): 1290-1296.
- Hartman, Kurt and Bryan Lang. (1999). "Exceptions to Informed Consent in Emergency Medicine." in *Hospital Physician*: 53-59.
- Hartman, Steve E. (2009). "Why do ineffective treatments seem helpful? A brief review" in *Chiropractic and Osteopathy*.

- Hay, M. Cameron, et. al. (2008). "Harnessing experience: exploring the gap between evidence-based medicine and clinical practice." in *Journal of Evaluation in Clinical Practice*.
- Heath, Iona. (2015). "Eminence or evidence-based medicine: why this question is still relevant today." Conference talk at the Evidence Live Conference, Oxford.
- Heneghan, Carl, Ben Goldacre and Kamal R. Mahtani. (2017). "Why clinical trial outcomes fail to translate into benefits for patients." in *Trials*. Bio Med Central: 1-7.
- Hideshima, Teru. Dharminder Chauhan. et.al. (2000). "Thalidomide and its analogs overcome drug resistance of human multiple myeloma cells to conventional therapy." in *Blood*. 96: 2943-2950.
- Hirschhausen, Eckhart von. (2008). *Die Leber wächst mit ihren Aufgaben: Komisches aus der Medizin*. Berlin: Rowohlt Verlag.
- Hoffmann, Tammy and Sarah Thorning. (2012). "The scatter of research: cross sectional comparison of randomised trials and systematic reviews across specialities." in *BMJ*.
- Honein, Margaret, Suzanne M. Gilboa and Cheryl S. Broussard. (2013). "The Need for Safer Medication Use in Pregnancy." in *Expert Review of Clinical Pharmacology*. 6(5): 453-455.
- Howick, Jeremy. (2009). "Questioning the Methodological Superiority of 'Placebo' Over 'Active' Controlled Trials." in *The American Journal of Bioethics*: 34-48.
- Howick, Jeremy. (2011). *The Philosophy of Evidence - Based Medicine*. Oxford: Wiley - Blackwell.
- Howick, Jeremy, Paul Glasziou and Jeffrey K. Aronson. (2009). "The evolution of evidence hierarchies: what can Bradford Hill's 'guidelines for causation' contribute?" in *Journal of the Royal Society of Medicine*. 102: 186-194.
- Howick, Jeremy. (2011). "Exposing the Vanities — and a Qualified Defense — of Mechanistic Reasoning in Health Care Decision Making." *PSA, Chicago Journals*. 78(5): 926-940.
- Howick, Jeremy, Paul Glasziou and Jeffrey K. Aronson. (2013). "Problems with using mechanisms to solve the problem of extrapolation." in *Theoretical Medicine and Bioethics*. 34: 275-291.
- Hutchinson, K., Rogers, W. (2012). "Challenging the epistemological foundations of EBM: what kind of knowledge does clinical practice require?" in *Journal of Evaluation in Clinical Practice*. 18.
- Hyry, Hanna I. and Jeremy Manuel, et.al. (2015). "Compassionate use of orphan drugs." in *Orphanet Journal of Rare Diseases*. 10(100).
- Ichikawa, Jonathan Jenkins and Steup, Matthias. (2017). "The Analysis of Knowledge", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.) <https://plato.stanford.edu/archives/spr2017/entries/knowledge-analysis>. Last accessed on January 23rd, 2020.
- Informationsnetzwerk Homöopathie. <https://www.netzwerk-homoeopathie.eu/>. Last accessed on October 8th, 2017.
- Institute for Work and Health. *At Work: Issue 55*, (2009). "Cross-sectional studies." www.iwh.on.ca. Last accessed on July, 15th 2015.

- Janice, Milos, Pat Croskerry and David Hitchcock. (2011). "Evidence and its issues in health care and research: The role of critical thinking." in *Medical Science Monitor*: 12-17.
- Jenicek, Milos, Pat Croskerry and David Hitchcock. (2011). "Evidence and its uses in health care and research: The role of critical thinking." in *Medical Science Monitor*.
- Jenkins, E., M. Yip, et. al. (2010). "Informed Consent: Cultural and Religious Issues Associated with the Use of Allogeneic and Xenogeneic Mesh Products." in Elsevier, *The American College of Surgeons*: 402-410.
- Johnson, Skyler B., et.al. (2018). "Use of Alternative Medicine for Cancer and Its Impact on Survival." in *Journal of the National Cancer Institute*, 110(1).
- Juth, Niklas. (2014). "For the Sake of Justice: Should We Prioritize Rare Diseases?" in *Health Care Annals*.
- Kastelein, JJ, A. Wiegman, and E. de Groot. (2003). "Surrogate markers of atherosclerosis." in *Atherosclerosis Supplements*. 4(1):31-6.
- Karanicolas, Paul. Forough Farrokhyar and Mohit Bhandari. (2010). "Blinding: Who, what, when, why, how?" in *Canadian Journal for Surgery*. 53(5): 345-348.
- Kelly, Thomas. (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.
- Kennedy, B.R., C.C. Mathis and A.K. Woods. (2007). "African Americans and their distrust of the health care system: healthcare for diverse populations." in *Journal of Cultural Diversity*. 14(2):56-60.
- Kent homeopathic repertory for iPhone. http://download.cnet.com/Kent-Homeopathic-Repertory/3000-2129_4-75978808.html. Last accessed on January 23rd, 2020.
- Kesselheim, Aaron S., Jessica A. Myers, and Jerry Avorn. (2011). "Characteristics of Clinical Trials to Support Approval of Orphan vs Nonorphan Drugs for Cancer." in *JAMA*; 305(22): 2320-2326.
- Klee. Ernst. (2014). *Euthanasie im Dritten Reich. Die Vernichtung lebensunwerten Lebens*. Frankfurt: Fischer Taschenbuch Verlag.
- Konrad, Peter. (2005). "The Shifting Engines of Medicalization." in *Journal of Health and Social Behavior*. 46(1): 3-14.
- Kool, Jacob L. and Robert A. Weinstein. (2005). "Risk of Person-to-Person Transmission of Pneumonic Plague." in *Clinical Infectious Diseases*. 40(8): 1166-1172.
- Krankenkassen Deutschland. "Integrierte Versorgung mit klassischer Homöopathie." <https://www.krankenkassen.de/gesetzliche-krankenkassen/leistungen-gesetzliche-krankenkassen/wahltarife-besondere-versorgung/integrierte-versorgung-homoeopathie/>. Last accessed on January 23rd, 2020.
- Kuhn, Thomas. (1996). *The Structure of Scientific Revolutions*. 3rd Ed. Chicago: The University of Chicago Press.
- Laan, Laura van der, and Marianne Boenink. (2015). "Beyond Bench and Bedside: Disentangling the Concept of Translational Research Health Care." in *Health Care Analysis*; 23: 32-49.
- La Caze, Adam. (2011). "The Role of Basic Science in Evidence-Based Medicine." in *Biology and Philosophy*, 26(1).

- Lakatos, Imre. (1978). "The Methodology of Scientific Research Programmes." in *Philosophical Papers Vol. I*. ed. John Worrall and Gregory Currie. Cambridge: Cambridge University Press.
- Lakatos, Imre. (1978). "Mathematics, science and epistemology." in *Philosophical Papers, Vol. II*. Ed. John Worrall and Gregory Currie. Cambridge: Cambridge University Press.
- Leibovici, Leonard. (2001). "Effects of remote, retroactive intercessory prayer on outcomes in patients with bloodstream infection: randomised controlled trial." in *BMJ*; 323(7327):1450–1451.
- Leil, TA. (2014). "A Bayesian Perspective on Estimation of Variability and Uncertainty in Mechanism-Based Models." in *CPT Pharmacocentrics Syst. Pharmacol.* 3, e12.
- Levy, David, Ben Gadd, et.al. (2013). "A Gentle Ethical Defence of Homeopathy." in *Bioethical Inquiry*.
- Library Guides: Subjects, Services and Resources. <http://libguides.rgu.ac.uk/c.php?g=536793&p=4389919>. Last accessed on January 23rd, 2020.
- Lloyd, Andrew, Paul Hayes. et.al. (2001). "The Role of Risk and Benefit Perception in Informed Consent for Surgery." in *Medical Decision Making*. 21(2): 141-149.
- Loughlin, Michael. (2009). "The basis of medical knowledge: judgement, objectivity and the history of ideas." in *Journal of Evaluation in Clinical Practice*. 15(6): 935-940.
- Luce, Bryan R. et.al. (2010). "EBM, HTA and CER: Clearing the Confusion." in *The Milbank Quarterly*. 88.
- Maas, A.H.E.M., and Y.E.A. Appelman. (2010). "Gender differences in coronary heart disease." in *Netherlands Heart Journal*. 18(12): 598-602.
- Machamer, Peter. (2004). "Activities and Causation: The Metaphysics and Epistemology of Mechanisms." in *International Studies in the Philosophy of Science*. 18(1).
- Machamer, Peter, Lindley Darden, Carl F. Craver. (2000). "Thinking about Mechanisms." In *Philosophy of Science*. 67(1).
- Machin, David and Peter M. Fayers. (2010). *Randomized Clinical Trials: Design, Practice and Reporting*. Oxford: Wiley-Blackwell.
- Mackie, John L. (1965). "Causes and conditionals." in *American Philosophical Quarterly* 2:245–65.
- Manson, Neil C. and Onara O'Neill. (2007). *Rethinking Informed Consent in Bioethics*. Cambridge: Cambridge University Press.
- McKibbin, K. Ann. (1998). "Evidence-based practice." in *Bulletin of the Medical Library Association*. 86(3): 369-401.
- Medcalc. Easy to use statistical software. Software provider for statistical methods in medical research. <https://www.medcalc.org/manual/analysis-of-covariance.php>. Last accessed on January 23rd, 2020.
- MEDLINE/PubMed Resources Guide. www.nlm.nih.gov. Last accessed on May 21st, 2017.
- Milgrom, L.R. (2007). "Conspicuous by its absence: the Memory of Water, macro-entanglement, and the possibility of homeopathy." in *Homeopathy*; 96(3): 209-219.
- Mills, Edward, Curtis Copper, et. al. (2006). "*Randomized Trials Stopped Early for Harm in HIV/AIDS: A Systematic Survey*." in *HIV Clinical Trials*; 7(1): 24-33.

- Mike, Valerie, Alfred N Krauss, and Gail S Ross. (1993). "Neonatal extracorporeal membrane oxygenation (ECMO): clinical trials and the ethics of evidence." in *Journal of medical ethics*; 19: 212-218.
- Moher, David, Schulz, Kenneth, Douglas Altman. (2011). "The CONSORT statement: revised recommendations for improving the quality of reports of parallel-group randomised trials." in *The Lancet*. 357:1191-1194.
- Morgan, S.G., K.L. Bassett, et.al. (2005). "'Breakthrough' drugs and growth in expenditure on prescription drugs in Canada." in *BMJ* 331: 815-816.
- Muhlhausler, Beverly and Frank Bloomfield, et.al. (2013). "Whole Animal Experiments Should Be More Like Human Randomized Controlled Trials." in *PLoS Biology* 11(2).
- Murad, M. Hassan, et.al. (2016). "New evidence pyramid." in *ebmed. Journal for evidence-based medicine*. 21(4).
- National Health and Medical Research Council, Australia. NHMRC Statement (2015). "Statement on homeopathy." <https://www.nhmrc.gov.au/sites/default/files/images/nhmrc-statement-on-homeopathy.pdf>. Last accessed on January 23rd, 2020.
- National Health Service, NHS UK. NHS Choices: your health, your choices. <http://www.nhs.uk/Conditions/Syphilis/Pages/Treatmentpg.aspx>. Last accessed on January 23rd, 2020.
- National Health Service, NHS UK."NHS England launches action plan to drive out wasteful and ineffective drug prescriptions, saving NHS over 190 million pound a year." <https://www.england.nhs.uk/2017/07/medicine-consultation/>. Last accessed on January 23rd, 2020.
- NCL2 Trial. University of Hamburg, UKE. <https://clinicaltrials.gov/ct2/results?cond=NCL2&term=&cntry1=EU%3ADE&state1=&recrs=>. Last accessed on January 23rd, 2020.
- Nordenstrom, Jorgen. (2007). *Evidence-Based Medicine in Sherlock Holmes' Footsteps*. Oxford: Blackwell Publishing.
- Nuovo, Jim, Joy Melnikow, and Denise Chang. (2002). "Reporting Number Needed to Treat and Absolute Risk Reduction in Randomized Controlled Trials." in *JAMA*. 87(21).
- Nuremberg Code, The. <https://history.nih.gov/research/downloads/nuremberg.pdf>. Last accessed on January 23rd, 2020.
- Nurses' Health Study. <http://www.nurseshealthstudy.org/>. Last accessed on January 23rd, 2020.
- O'Neill, Onora. (2003). "Some limits of informed consent." in *Journal of Medical Ethics*. 29: 4-7.
- Pablos-Méndez, Ariel, R. Graham Barr, and Steven Shea. (1998). "Run-in Periods in Randomized Trials: Implications for the Application of Results in Clinical Practice." in *JAMA*; 279(3): 222-225.
- Pavlov, Ivan. V. (2003). *Conditioned Reflexes*. New York: Dover Publication. Reprint.
- Philips, Bob. (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.
- Pocock, Stuart. (1992). "When to stop a clinical trial." *BMJ*. 3: 235-240.
- Polyani, Michael. (1966). *The Tacit Dimension*. Chicago: Chicago University Press.
- Polyani, Michael. (1974). *Personal Knowledge: Towards a Post-Critical Philosophy*. Chicago: Chicago University Press.

- Popper, Karl. (1992). *The Logic of Scientific Discovery*. London, New York: Routledge.
- Public Health Service. A primary division of the United States Department of Health and Human Services. <https://www.usphs.gov/>. Last accessed on January 23rd, 2020.
- Ramey, David. (2008). "Is There a Placebo Effect for Animals." in *Science Based Medicine: Exploring issues and controversies in the relationship between science and medicine*. <https://sciencebasedmedicine.org/is-there-a-placebo-effect-for-animals/>. Last accessed on January 23rd, 2020.
- Reid, Kevin I. (2017). "Informed Consent in Dentistry." in *The Journal of Law, Medicine and Ethics*. 45(1): 77-94.
- Reiss, Julian and Ankeny, Rachel A. (2016). "Philosophy of Medicine" in *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.), <https://plato.stanford.edu/archives/sum2016/entries/medicine>. Last accessed on January 23rd, 2020.
- Ribbeck, Janko von. (2012). *Schnelle Hilfe für Kinder: Notfallmedizin für Eltern*. München: Kösel-Verlag.
- Rogers, Wendy. (2004). "Evidence-Based Medicine and Women: Do the Principles and Practice of EBM further Women's Health?" in *Bioethics* 8 (1): 50-71.
- Rogers, Wendy and A. Ballantyne. (2008). "Exclusion of Women From Clinical Research: Myth or Reality?" in *Mayo Clinic Proceedings*. 5: 536-542.
- Rogers, Wendy and M.M. Lange. (2013). "Rethinking the vulnerability of Minority Populations in Research." in *American Journal of Public Health*. 103(2): 2141-2146.
- Rosner, Anthony. (2012). "Evidence-based medicine: Revisiting the pyramid of priorities." in *Journal of Bodywork & Movement Therapies*. 16: 42-49.
- Rothwell, Peter, M. (2005). "External validity of randomised controlled trials: 'To whom do the results of this trial apply?'" in *The Lancet*. 365: 82-93.
- Rycroft-Malone, Jo. Kate Seers, Angie Titchen, et.al. (2004). "What counts as evidence in evidence-based practice?" in *Journal of Advanced Nursing*. 47(1): 81-90.
- Sackett DL, Rosenberg WMC, Gray JAM, Haynes RB, Richardson WS. (1996). "Evidence based medicine: what it is and what it isn't." in *BMJ* 312: 71-2.
- Sackett, David, Strauss SE, Richardson WS, et al. (2000). *Evidence-based medicine: how to practice and teach EBM*. London: Churchill-Livingstone.
- Sackett, David. L. (2000). "The Sins of Expertness and a proposal for redemption." in *BMJ*; 320.
- Sackett, David L. (2000). "Why randomised controlled trials fail but needn't: 1. Failure to gain 'coal-face' commitment and to use the uncertainty principle." in *Canadian Medical Association Journal*. 126 (9): 1311-1314.
- Sackett, David L. (2001). "Why randomised controlled trials fail but needn't: 2. Failure to employ physiological statistics, or the only formula a clinician-trialist is ever likely to need (or understand!)" in *Canadian Medical Association Journal*. 165 (9): 1226-1237.
- Salmon, Daniel A., Matthew Z. Dudley, Jason M. Glanz, and Saad B. Omer. (2015). "Vaccine Hesitancy Causes, Consequences, and a Call to Action" in *American Journal of Preventive Medicine and Elsevier*.
- Schramme, Thomas. (2002). *Bioethik*. Frankfurt/New York: Campus Verlag.
- Schramme, Thomas. (2015). *New Perspectives on Paternalism and Health Care*. Heidelberg: Springer Verlag.

- Schon, Scott R. and Donald E Stanley. (2003). "A philosophical analysis of the evidence-based medicine debate" in *BioMedCentral Health Services Research*. 3(14).
- Seidelmann, William E. (1995). "Mengele Medicus: Medicine's Nazi Heritage." in *International Journal of Health Services*. 19(4): 599-610.
- Shah, Sonia. (2008). *Am Menschen Getestet! Wie die Pharmaindustrie die Ärmsten der Welt für Medikamententests missbraucht*. München: Redline Wirtschaft.
- Shang, Aijing, Huwiler-Müntener, Karin, Linda Nartey, et.al. (2005). "Are the clinical effects of homeopathy placebo effects? Comparative study of placebo-controlled trials of homeopathy and allopathy." in *The Lancet*, 366: 726-732.
- Shao, J. and SC. Chow. (2002). "Reproducibility probability in clinical trials." in *Statistics in Medicine*. 21(12): 1727-42.
- Silja, Anja. (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.
- Silva, Suzanna A. and Peter Wyr. (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. "Cancer Control." 16, in *Journal of Evaluation in Clinical Practice*:15: 158-168.
- Singer, Peter A., Douglas K. Martin and Merrijoy Kelner. (1999). "Quality end-of-life care: patients' perspectives." in *JAMA* 13; 281 (2): 163-168.
- Singh, Sonal and Yoon K. Loke. (2012). "Drug safety assessment in clinical trials: methodological challenges and opportunities." in *TRIALS*. 13(138): 1-8.
- Smith, Gordon C.S., Pell, Jill P. (2003). "Parachute use to prevent death and major trauma related to gravitational challenge: systematic review of randomised controlled trials." in *BMJ*. 327: 1459-1461.
- Smith, Robert. (2014). "Medical research — still a scandal." in the *bmjopinion*. <http://blogs.bmj.com/bmj/2014/01/31/richard-smith-medical-research-still-a-scandal/>. Last accessed on January 23rd, 2020.
- Solomon, Miriam. (2015). *Making Medical Knowledge*. Oxford: Oxford University Press.
- Song, Jae W. and Kevin Chung. (2010). "Observational studies: Cohort and Case Control Studies." in *Plastic and Reconstructive Surgery* 126(6): 2234-2242.
- Sprenger, Jan and Jacob Stegenga. (2017). "Three Arguments for Absolute Outcome Measures." in *PSA 2016: The 25th Biennial Meeting of the Philosophy of Science Association*.
- Spruance, Spootwood L., Julia Reid, Michael Grace, and Matthew Samore. (2004). "Minireview: Hazard Ratio in Clinical Trials." in *Antimicrobial Agents and Chemotherapy*. 48(8): 2787-2789.
- Stavrakis, P. (1997). "Heroic medicine, bloodletting, and the sad fate of George Washington." in *Maryland Medical Journal*. 46(10): 539-540.
- St. Clair, E. William. (2008). "The calm of the cytokine storm: lessons from the TGN1412 trial." in *Journal of Clinical Investigation*. 118 (4): 1344-1347.
- Stanley, Kenneth. (2007). "Design of Randomized Controlled Trials." in *Circulation, Journal of the American Heart Association*. 115: 1164-1169.
- Stegenga, Jacob. (2011). "Is Meta-Analysis the Platinum Standard of Evidence?" in *Studies in History and Philosophy of Science Part C* 42 (4): 497-507.

- Stenger, Michael, Nicki Eithz Bauer and Peter B. Licht. (2013). "Is pneumothorax after acupuncture so uncommon?" in the *Journal of Thoracic Disease*. 5(4): 144-146.
- STROBE Statement: Strengthening the reporting of observational studies in epidemiology. www.strobe-statement.org/index.php?id=strobe-group0. Last accessed on January 23rd, 2020.
- Suresh, K.P. (2011). "An overview of randomisation techniques: An unbiased assessment of outcome in clinical research." in *Journal of Human Reproductive Sciences*. 4(1): 8-11.
- Suter, Katja, Matthias Briel and Judith Günther. (2015). "Number needed to treat (NNT) and Number needed to harm (NNH): weitere Abkömmlinge der Vier-Felder-Tafel." in *Medizinische Monatszeitschrift für Pharmazeuten* 38(3): 103-106.
- Tabery, James G. (2004). "Synthesizing Activities and Interactions in the Concept of a Mechanism." in *Philosophy of Science*. 71.
- Temple, Robert. (2012). "Enrichment Strategies." in Guidelines of the U.S. Food and Drug Administration. <https://www.fda.gov/ucm/groups/fdagov-public/@fdagov-afda-orgs/documents/document/ucm303485.pdf>. Last accessed on January 23rd, 2020.
- Thomas, N. (2008). "Historical Control." in *Wiley Encyclopedia of Clinical Trials*. 1-3.
- Timmermans, Stefan and Marc Berg. (2003). *The Gold Standard: The Challenge of Evidence-Based Medicine and Standardization in Health Care*. Philadelphia. Temple University Press.
- Tough, Elizabeth, Adrian White. (2009). "Developing and validating a sham acupuncture needle." in *Acupunct Med, BMJ*. 27: 118-122.
- Travers, Justin., Marsh, Suzanne., et.al. (2007). "External validity of randomised controlled trials in asthma: to whom do the results of the trial apply?" in *Thorax*. 62: 219-223.
- Truog, Robert D. (1992). "Randomized Controlled Trials: Lessons from ECMO." in *Clinical Research*. 40(3): 519-527
- Truog, Robert D. (1999). "Informed Consent and research design in critical care medicine." in *Critical Care*, 3(3).
- Truog, Robert D. (2017). "The United Kingdom Sets Limits on Experimental Treatments: The Case of Charlie Gard." in *JAMA*. 318(11).
- Tunis, Sean, Daniel b. Stryer, and Carolyn M. Clancy. (2003). "Practical Clinical Trials: Increasing the Value of Clinical Research for Decision Making in Clinical and Health Policy." in *JAMA*. 290(12).
- Turner, Stephen. (2013). *The politics of expertise*. Oxford: Routledge.
- Tuskegee University. National Centre for Bioethics in Research and Health Care. <http://tuskegeebioethics.org>. Last accessed on January 23rd, 2020.
- Upshur, Ross. (2002). "If not evidence, then what? Or does medicine really need a base?" in *Journal of Evaluation of Clinical Practice*. 8(2): 113-119.
- Vandenbroucke, Jan P, 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*.
- Vandenbroucke, Jan P. (2001). "In defence of case reports and case series." in *Annals of Internal Medicine*. 134(4): 330-334.

- Van Spall, Harriette, Toren, Andrew. et. al. (2007). "Eligibility Criteria of Randomized Controlled Trials Published in High-Impact Medical Journals. A systematic sampling review." in JAMA. 297(11): 1233-1240.
- Weimer, Katja. Marco D. Gulewitsch, Angelika A. Schlarb, Juliane Schwille-Kiuntke, Sibylle Klosterhalfen & Paul Enck. (2013). "Placebo effects in children: a review." in Pediatric Research, 74: 96-102.
- White, A. and Edzard Ernst. (2004). "A brief history of acupuncture." in Rheumatology. 43(5): 662-663.
- Williams, Hywel. (2015). "Reducing avoidable waste in eczema research." Conference talk at the EvidenceLive Conference Oxford.
- Witt, C and B Brinkhaus, et.al. (2005). "Acupuncture in patients with osteoarthritis of the knee: a randomised trial." in The Lancet. 366: 136-43.
- Woods, John and Douglas Walton. (1997). "Post Hoc, Ergo Propter Hoc." in The Review of Metaphysics. 30(4): 569-593.
- Woodward, Jim. (2002). "What Is a Mechanism? A Counterfactual Account." in Philosophy of Science, 69.
- Woolf, Steven H., Richard Grol, Allen Hutchinson, Martin Eccles, and Jeremy Grimshaw. (1999). "Clinical guidelines Potential benefits, limitations, and harms of clinical guidelines." in BMJ 318: 527- 530.
- WHO report on acupuncture. (2003). [http://www.acucentre.com.au/Clinic/WHOConditions Tx.pdf](http://www.acucentre.com.au/Clinic/WHOConditionsTx.pdf). Last accessed on January 23rd, 2020.
- World Health Organisation. WHO vaccines website. <http://www.who.int/topics/vaccines/en/>. Last accessed on January 23rd, 2020.
- World Medical Association, The. *Declaration of Geneva*. Adopted by the 2nd General Assembly of the World Medical Association, Geneva, Switzerland, September 1948. Lastly editorially revised by the 173rd WMA Council Session, Divonne-les-Baines, France, May 2006.
- World Medical Association, The. *Declaration of Helsinki*. Adopted by the 18th WMA General Assembly, Helsinki, Finland, June 1964. Lastly editorially revised by the 64th WMA General Assembly, Fortaleza, Brazil, October 2013.
- World Medical Association, The. *International Code of Medical Ethics* from 1949. Adopted by the 3rd General Assembly of the World Medical Association, London, England, October 1949 and lastly amended by the 57th WMA General Assembly, Pilanesberg, South Africa, October 2006.
- Worp, van der, H. Bart, David Howells, et.al. (2010). "Can Animal Models of Disease Reliably Inform Human Studies?" in PLoS Med. 7(3).
- Worrall, John. (2003). "What Evidence in Evidence - Based Medicine?" in Causality: Metaphysics and Methods. Technical Report 01/03 by the Centre for Philosophy of Natural and Social Science.
- Worrall, John. (2007). "Why There's no Cause to Randomize." in The British Journal for the Philosophy of Science. 58 (3): 451-488.
- Worrall, John. (2008). "Evidence and Ethics in Medicine." in Perspectives in Biology and Medicine 51: 418-31.

- Worrall, John. (2010). "Do we need some large, simple randomized trials in medicine?" in M. Suarez, M. Dorato and M. Redei (eds). *EPSA Philosophical Issues in the Sciences*. Dordrecht: Springer.
- Yang, W, A Zilov, P. Soewondo, OM Bech, F. Sekkal and PD Home. (2010). "Observational studies: going beyond the boundaries of randomized controlled trials." in *Diabetes Res Clinical Practice*; 88 Supplement 1: S3-9.
- Zelen, Marvin. (1997). "A new design for randomized clinical trials." In *New England Journal of Medicine* 300: 1242-1245.
- Zhao, Yanli, Grambsch, Patricia M. Grambsch, and James D. Neaton. (2007). "A decision rule for sequential monitoring of clinical trials with a primary and supportive outcome." in *Clinical Trials* 4: 140-153.
- Ziaecian, B. and G.C. Fonarow. (2017). "Statins and the Prevention of Heart Disease." in *JAMA Cardiology*; 2(4).
- Zhou J, et al. (2011). "Acupuncture anaesthesia for open heart surgery in contemporary China." in *International Journal of Cardiology*. 150(1).

Appendix

Kurzfassung der Ergebnisse

Die vorliegende Dissertation beschäftigt sich mit den Fragen, ob Evidenz-basierte Medizin (EbM), seit sie sich seit den 1990'er Jahren als Begriff und dann auch schnell als Methode etabliert hat, einen Paradigmenwechsel im Sinne Thomas Kuhns hervorgerufen hat? Inwieweit und ob sie zu einer neuen epistemologischen Theorie der Naturwissenschaften beigetragen hat? Und welche methodologischen und ethischen Stärken und Schwächen EbM im Umgang mit Patienten und mit der medizinischen Forschung aufweist?

Evidenz-basierte Medizin steht für eine Medizin die sich auf wissenschaftlich fundierte Begründungen zur Behandlung individueller Patienten stützt. Allerdings ist diese Definition sehr verkürzt, da EbM auch die wissenschaftliche Seite der Generierung von Evidenz beinhaltet. Um Evidenz zu generieren werden Studien durchgeführt die die Sicherheit und Wirksamkeit von Medikamenten und Behandlungen testen und bewerten. Diese Studien gibt es in verschiedenen Formen, deren interne Validität, also ob die wissenschaftliche Methodik korrekt ausgeführt worden ist, und externe Validität, also ob die Ergebnisse generalisierbar sind, unterschiedlich bewertet werden. Basierend auf dieser Bewertung werden die Studien dann hierarchisch geordnet. Am besten bewertet werden randomisierte kontrollierte Studien, RTCs, da diese die am wissenschaftlich unumstößlichsten Ergebnisse erzielen. Für diese Studien werden Teilnehmer meistens in zwei Gruppen randomisiert. Eine der Gruppen erhält das zu testende Medikament und die andere Gruppe entweder das bereits bekannte Standardmedikament oder ein Placebo. Randomisierte kontrollierte Studien sind allerdings nur dann höherwertig als andere Studien, wenn sie korrekt durchgeführt wurden und interne Validität nachweisbar ist. Da die interne Validität von randomisierten kontrollierten Studien allerdings durch viele systematische Fehler, wie zum Beispiel das bewusste Aussuchen von Patienten für Studien oder das Wissen welcher Patient welches Medikament erhält, in Mitleidenschaft gezogen werden kann, ist es zwingend notwendig diese so gut wie möglich zu eliminieren, und die gesamte Datenlage sowie das initiale Studienprotokoll öffentlich zugänglich zu machen. Nur so können Fehler entdeckt und Patienten und zukünftige Studienteilnehmer geschützt werden.

Externe Validität, also das für den Patienten entscheidende Ergebnis, nämlich ob das Medikament das Richtige für den jeweiligen Patienten ist, kann unter anderem durch ein gedankliches Umstossen von Evidenzhierarchien erfolgen. Und wenn Studienteilnehmer nach Kriterien ausgesucht werden, die mit denen der späteren Patienten die mit dem neuen Medikament therapiert werden sollen, möglichst nah übereinstimmen.

Der Arzt muss sein ganzes Wissen nutzen, um, zusammen mit dem Patienten eine Therapientscheidung zu treffen. Allerdings wird EbM dadurch nicht zu einer neuen Theorie des Wissens, sondern bedient sich lediglich einiger der Methoden von Wissenstheorien. Der Name Evidenz-basierte Medizin ist bewusst gewählt in Abgrenzung zum Wissensbegriff, da Evidenz als ‘guter Grund etwas zu glauben’ definiert werden kann, und damit das Wahrheitselement welches dem Wissensbegriff zugrunde liegt, nicht beinhaltet. Damit wird impliziert das Evidenz, im Gegensatz zu Wissen, veränderbar ist und sich stetig ändert, als auch von unterschiedlicher Qualität sein kann. Wissen kennt diese Qualitätsunterschiede zwischen ‘gut’ und ‘schlecht’ oder ‘fehlerhaft’ so nicht.

Um also zu erreichen, dass Evidenz-basierte Medizin nicht nur fehlerfreie Evidenz produziert, sondern diese auch für den Patienten und den Arzt nutzbar ist, muss in der Diskussion zwischen Evidenz-basierter Forschung und Evidenz-basierter Praxis unterschieden werden. In der Evidenz-basierten Forschung geht es also um das Erreichen von ‘guter’, also robuster, reproduzierbarer und quantifizierbarer Evidenz. Es geht in der Unterscheidung aber auch darum, klarzustellen dass sich die Rollen der an der Forschung beteiligten Parteien ändern. Der Patient wird zum ‘Teilnehmer’ und der Arzt wird zum ‘Forscher’. Damit wird das Arzt-Patientenverhältnis welches idealerweise auf Vertrauen beruht, naturgemäss unterminiert, da es in der Forschung nicht mehr möglich ist, dem Patienten die bestmögliche Versorgung zukommen zu lassen. Neue Medikamente können gefährlich sein, und ein Placebo enthält keine wirksamen Inhaltsstoffe. Der Studienteilnehmer muss sich über die assoziierten Risiken bewusst sein und sich ihnen freiwillig stellen. In diesen Fällen ist es besonders wichtig, auf das Werkzeug der ‘informierten Einwilligung’ zu bestehen, welches allerdings auch in der Praxis genutzt werden muss, um überhaupt den Patienten behandeln zu dürfen. Die ‘informierte Einwilligung’ ist kein absoluter Schutz, und scheitert für Patienten die nicht volljährig, oder anderweitig in ihrer freien Entscheidung eingeschränkt sind. Hier müssen andere Wege gefunden werden. Allerdings ist die ‘informierte Einwilligung’ dennoch ein wichtiges Instrument, speziell in der Forschung, da ohne sie medizinische Forschung am Menschen nicht durchgeführt werden dürfte.

In der Evidenz-basierten Praxis geht es darum alle Evidenz, also auch weniger robuste, zu bewerten und zu nutzen im Hinblick auf den einzelnen Patienten. Dieses Vorgehen inkludiert allerdings nicht den Rekurs auf alternative Heilmetho-

den wie Homöopathie oder Akupunktur, da diese nicht auf einer soliden wissenschaftlichen Datenlage basieren und somit keine gesicherten Resultate für den Patienten liefern können. Die Evidenz-basierte Praxis sollte sich nicht nur mit statistisch gesicherten Forschungsergebnissen befassen, sondern sie sollte sich vor allen Dingen auf die Wünsche, Werte und Vorstellungen des Patienten konzentrieren. Hierbei ist es von entscheidender Wichtigkeit, dass der Arzt sein ganzes Wissen, also auch sein implizites Wissen und seine Erfahrung nutzt um zusammen mit dem Patienten Entscheidungen treffen zu können. Dieses Wissen ist somit Teil der benötigten Evidenz für die Praxis.

Zu guter Letzt ist zu bewerten, ob EbM wirklich einen Paradigmenwechsel im Sinne Thomas Kuhns ausgelöst hat. In aller Kürze ist hier zu festzustellen, dass das nicht der Fall ist. Kuhn stützt seine These des wissenschaftlichen Paradigmenwechsels auf die 'Inkommensurabilität' des neuen und des alten Paradigmas und bezieht damit ein, dass die beiden Gruppen nicht mehr miteinander kommunizieren können, da sich die Sprache von einem zum nächsten Paradigma geändert hat und kein Verständnis mehr vorhanden ist. In der Geschichte der Medizin hat so ein Sprachenwechsel allerdings nie stattgefunden. Die medizinische Fachsprache hat sich im Laufe der Zeit und durch die Zunahme der medizinischen Evidenz verändert und ist nuancierter geworden. Allerdings ist es immer noch möglich medizinische Literatur von vor der Zeit EbMs zu lesen und zu verstehen und es ist auch noch möglich mit Medizinern die Medizin vor EbM gelernt haben, zu kommunizieren. Es gibt keine zwei Gruppen. Daher ist es also nicht nötig und nicht möglich von einem Paradigmenwechsel zu sprechen. Vielmehr ist die Entstehung der Evidenz-basierten Medizin eine notwendige Entwicklung gewesen, um die sich zwischen den 1950ern und 1990ern massiv verändernde Medizin neu zu fassen und in eine einheitliche Form zu gießen. Es ist daher nötig, EbM heraus zu fordern, allerdings nicht als Paradigma, sondern um den Patienten wieder in das Zentrum der Medizin zu rücken, und zwar sowohl in der Evidenz-basierten Forschung als auch in der Evidenz-basierten Praxis.

English Version

The dissertation is to occupy itself with the questions if evidence-based medicine (EBM), which since the 1990's is not only a term but has established an entire methodology, has led to a paradigm change in the sense of Thomas Kuhn? If it has contributed to or even established a new theory of epistemology? And which methodological and ethical strengths and weaknesses EBM possesses with regard to the individual patient and to medical research?

Evidence-based medicine is based on scientifically validated evidence in order to treat the individual patient. But it also contains the generating of such evidence. And in order to generate robust and therefore 'good' evidence, trials are conducted which test the safety, efficacy and effectiveness of novel drugs and treatments. These studies are ranked hierarchically according to their internal and external validity. Internal validity is that which proclaims the trial to be correctly done while external validity is that which makes the results usable for the actual patient population that needs to be treated by the drug or treatment under test.

On top of most evidence hierarchies are randomised controlled trials (RCTs), since these are providing the most scientifically robust results. These studies divide their participants into two groups, the treatment group receives the novel treatment and the control group receives either a standard treatment or a placebo. RCTs however are only superior to other study designs if they are both internally and externally valid. Internal validity stands for how well the procedures in the study measured what they were supposed to measure. Internal validity can be compromised through biases, such as allocation- or selection bias in which patients are deliberately chosen for a trial, or the knowledge on part of the participant and the researcher which treatment each participant receives. It is therefore imperative to control studies for these biases and to publish all accumulated data, even the raw statistical data, so that future researchers can detect possible flaws and save patients and future participants from 'bad' evidence.

External validity can be achieved by toppling the evidence hierarchies for medical practice and by choosing trial participants which most closely resemble the actual target population.

The physician has to utilise not only the available evidence, but also his tacit knowledge and his expertise to arrive, together with the patient, at a treatment decision. EBM however is not a theory of epistemology, but only utilises some of the features of epistemological theories. The actual term 'evidence-based medicine' already implies its difference from medical knowledge. Evidence is defined as 'good reason to believe' and is therefore lacking the truth-element that is inherent to the definition of 'knowledge.' Evidence is changeable and quantifiable. And it

can be classified into 'good' and 'bad' evidence. Knowledge can only ever be incomplete.

To achieve the goal for EBM to be more patient-centred, it should be divided into 'evidence-based research' and 'evidence-based practice.' Evidence-based research is focused on producing 'good' and therefore robust, quantifiable and reproducible evidence. And the patient and the physician are changing their roles in research. The patient becomes a participant and the physician becomes a researcher. The patient-physician relationship which is based on trust brakes down since in research the patient might not receive the best available treatment anymore. Novel drugs and treatments can be dangerous or the participant just receives a placebo and therefore no active treatment. The participant needs to be informed about all these inherent risks and has to participate freely and willingly. Here the tool of 'informed consent' is especially important. The patient needs to consent to be part of medical research, as well as just being treated in medical practice. Without 'informed consent' treatment would not be possible. However, informed consent is not a fail-safe and reaches its limits where underage patients or patients without the possibility of freely giving informed consent are concerned. There are possible ways to deal with these situations, such as including guardians. 'Informed consent' however is an important tool, especially for research, since without it, no research involving human beings would be legally possible to conduct.

Evidence-based practice has to take all available evidence into account in order to successfully treat the individual patient. However, alternative methods like homeopathy or acupuncture are not part of the broader evidence-base, since they are not based on scientifically proven evidence and can therefore not guarantee a benefit for the patient.

Evidence-based practice should focus on the patients wishes, values, concerns and individual needs rather than on purely statistical evidence. In medical practice the tacit knowledge and expertise of the physician is of vital importance for the individualised care of the single patient. This knowledge is therefore part of the evidence for medical practice.

Lastly it can be concluded that EBM is not a new paradigm and has not provoked a paradigm shift in the Kuhnian sense. Kuhn bases his theory of a paradigm shift heavily on his theory of 'incommensurability' which essentially means that those conducting research before the paradigm shift and those working in the new paradigm cannot meaningfully converse with each other anymore, since the language and terminology has changed in a way as to make the two groups unintelligible to each other. The methodologies and language before and after EBM however are not 'incommensurable' with each other. The medical language his changing and has become more nuanced, but it does so gradually and over time. However, it is still possible to read and understand, and even use medical literature from the times before EBM. It is therefore neither necessary nor really possible to

talk about a paradigm shift. On the contrary, EBM is the inevitable response to rapid change in medical science and understanding. EBM therefore is ready to be challenged, but not as a paradigm, but as a form medicine based on scientific principles which needs to adapt and diversify in order to put the patient back into the centre of medicine, in both research and practice.