Getting Better II:
What education in less economically developed countries can learn about evidence from medicine

Caroline Fiennes
Dr Saurabh Gupta

Giving Evidence
Enabling giving based on sound evidence
On his experience as a medical doctor in a prisoner of war ‘hospital’ in Germany in 1944:

“I had cared for my patients as well as I could have in the circumstances. I knew my patients so well that I was miserable when they died. I found ‘caring’ intellectually unsatisfactory. I could not continue making decisions about intervening (for example, pneumothorax and thoracoplasty) when I had no idea whether I was doing more harm than good. I remember reading a pamphlet extolling the advantages of the freedom of doctors to do whatever they thought best. I found it ridiculous. I would willingly have sacrificed all my medical freedom for some hard evidence telling me when to do a pneumothorax.”

– Archie Cochrane

“The story of medicine is the story of how we deal with the incompleteness of our knowledge and the fallibility of our skills… When we think about the transformation of human health during the last century, we tend to focus on the discoveries - on the new drugs and the new procedures. There was another major contribution during this past century: the emergence of complex, specialised institutions for delivering those discoveries to people - hospitals, clinics, professionals who could staff them.”

– Atul Gawande

About Giving Evidence

Giving Evidence is a consultancy and campaign, promoting charitable giving based on sound evidence. Our interest in medicine arises because it is more organised around evidence than virtually any other discipline.

Through consultancy, Giving Evidence helps donors and charities to understand their impact and to increase it. Through campaigning and thought-leadership, we show what evidence is available, what is needed, what charities and funders should gather, what isn’t worth gathering, and work to improve the quality, clarity and findability of evidence.

Giving Evidence was founded by Caroline Fiennes, a former award-winning charity CEO, and author of It Ain’t What You Give. Caroline speaks and writes extensively about these issues, e.g., at the Skoll World Forum, Arab Foundations Forum, in the Stanford Social Innovation Review, Forbes, Freakonomics, and the Daily Mail. She is on boards of The Cochrane Collaboration, Charity Navigator (the world’s largest charity ratings agency) and the US Center for Effective Philanthropy.

Dr Saurabh Gupta is an orthopaedic surgeon by background, trained in the UK in public health. He was a Consultant in Public Health in the UK National Health Service (NHS), and has 18 years’ experience across various health and management settings in India and the UK.

We are grateful to the funders who enabled this work.

Caroline Fiennes
Director, Giving Evidence, +44 7803 954512, caroline.fiennes@giving-evidence.com

www.giving-evidence.com
# Contents

1 Background to the Getting Better project .................................................. 6

2 Introduction
   2.1 Contents and structure of this document ........................................... 10

3 Disparity between spending on medical research and education research ...... 11
   Our thoughts about why this mismatch arises ........................................ 12

4 Evidence system around public health in LEDCs ....................................... 13
   4.1 Production: Limited amount of primary research in LEDCs .................. 13
   4.2 Synthesis: Currently available systematic reviews do not reflect developing world priorities ......................................................... 13
   4.3 Dissemination: LEDC policy-makers and health care professionals have limited access to subscription-based information .................. 15
   4.4 Use: Many interventions reviewed cannot be initially implemented in resource-poor situations, but with help, can be implemented .......... 16
   4.5 Use: the extreme difficulty of implementation ................................... 18

5 Useful tools developed in medicine ............................................................ 19
   5.1 Standardised units used in medicine .................................................. 19
   5.2 Quality-Adjusted Life Years (QALYs)
      5.2.1 Definition .................................................................................... 19
      5.2.2 Development of QALYs ............................................................... 19
      5.2.3 Use of QALYs ............................................................................ 19
      5.2.4 Role of institutions in promoting QALYs ..................................... 21
   5.3 Disability-Adjusted Life Years (DALYs)
      5.3.1 Why were DALYs developed? ...................................................... 22
      5.3.2 How DALYs have been used ....................................................... 23
      5.3.3 The Global Burden of Disease .................................................... 23
      5.3.4 Summary measures of population health ..................................... 23
      5.3.5 Health Adjusted Life Years ......................................................... 24
   5.4 Randomised controlled trials (RCTs)
      5.4.1 Origins of randomised trials – in psychology and agriculture ........ 25
      5.4.2 Introduction of RCTs in medicine ............................................... 25
      5.4.3 Principles and limitations of RCTs .............................................. 25
      5.4.4 RCTs as ‘gold standard’ for a trial .............................................. 26
5.5 Checklists for reporting trials and experiments
   5.5.1 Evolution of Introduction, Methods, Results, And Discussion (IMRAD) structure for reporting experiments and results 28
   5.5.2 Development of standardised format for reporting clinical trials 28
   5.5.3 Developing standardised content for trial reports: the origins of CONSORT 29
   5.5.4 Endorsement and enforcement 29
   5.5.5 The effect of CONSORT 29
   5.5.6 Funding and staffing of CONSORT 32
   5.5.7 Reporting systematic reviews: QUOROM statement and its evolution into PRISMA 32
   5.5.8 Development of PRISMA 32
   5.5.9 Advancing the work of CONSORT to other areas of medical research - EQUATOR 33

6 How did medicine come to be so evidence-based?
   Excerpts from the history of Evidence-Based Medicine (EBM) 34
   6.1 Use of Evidence in Medicine prior to 1960s 36
   6.2 Use of Evidence in Medicine after 1960s 38
      6.2.1 Introduction of key concepts 39
   6.3 Emergence of the field of clinical epidemiology 39
   6.4 Role of Institutions in promoting EBM 40
      6.4.1 McMaster University (Canada) 42
      6.4.2 Centre for EBM (UK) 42
      6.4.3 Agency for Health Care Policy and Research (USA) 44
      6.4.4 Cochrane Collaboration (worldwide) 45
      6.4.5 National Institute for Health and Care Excellence (NICE) (UK) 45
   6.5 Repositories and indexing of medical literature 46
      6.5.1 Role of National Library of Medicine in EBM 46
      6.5.2 Milestones in the creation of repositories and indexing of medical research 47
         Table 6: Milestones in evidence generation and research: 47
      6.5.3 National Network of Libraries of Medicine (NN / LM) 48
   6.6 Backlash against EBM 49
   6.7 Making it easy to access evidence 50
   6.8 Ongoing problems in EBM 51

7 Case study of discovery and deployment of developments: Oral rehydration therapy 53
   7.1 Cholera epidemics in the 19th century 53
   7.2 Causes of Cholera 53
      7.2.1 Miasma theory 53
      7.2.2 Discovery of Cholera organism 54
      7.2.3 The epidemics subside 54
      7.2.4 Ignoring evidence about ORS 55
      7.2.5 Role of institutions in developing ORS 55
      7.2.6 Adoption of ORS 55
      7.2.7 Global campaign on ORS 57
      7.2.8 Global adoption of ORS 58
      7.2.9 Challenges to adoption of ORS 58
8 Lessons for education in LEDCs

8.1 Short-cuts

8.2 Initiatives to prioritise and try
   8.2.1 ‘Show your working’
   8.2.2 Getting more evidence – about both prevalence and effectiveness
   8.2.3 Monitoring the quality of evidence produced by and about education providers
   8.2.4 Involving teachers and practitioners in prioritising and producing research
   8.2.5 Experiments with disseminating evidence and research uptake

8.3 Things to avoid

8.4 Things to investigate

8.5 Red herrings

8.6 Conclusion

9 Appendices

9.1 Appendix 1: CONSORT 2010 checklist of information to include when reporting a randomised trial, and flowchart

9.2 Appendix 2: Flow diagram of the progress through the phases of a parallel randomized trial of two groups (that is, enrolment, intervention allocation, follow-up, and data analysis)

9.3 Appendix 3: Checklist of items to include when reporting a systematic review or meta-analysis, and flowchart

9.4 Appendix 4: Flow of information through the different phases of a systematic review

10 References
1 Background to the Getting Better project

The evidence-based practice in medicine contrasts starkly with the norm in education in less economically developed countries (LEDCs). Given the marked performance improvements in many areas of health in recent decades, there are perhaps lessons for how education could improve.

This observation led Giving Evidence to start the Getting Better project in 2013 to help improve education by learning from healthcare. Several current debates in education about performance and measurement are similar to those in medicine as recently as the 1970s: teachers and educationalists often object to measurement on the grounds of the individuality of the people they serve, on the basis that they know what’s best, because the true value of their work would elude measurement, and because findings from one situation couldn’t possibly be used elsewhere. All these objections were made by medical practitioners about evidence-based medicine (EBM) – the formal practice of identifying the best treatment of patients based on the systematic and detailed appraisal of research evidence. They were fiercely debated as EBM began and has integrated into medical education and practice over the last three decades. Medicine has seen a paradigm shift so that today evidence has become an essential part of clinical treatment.

We aim to learn from the way that evidence is used in health care and medicine, not necessarily to emulate it, since clearly education and medicine differ in many respects. The first Getting Better document outlined the ‘evidence systems’ (see below) in both medicine and education in LEDCs. This current paper builds upon and should be read in conjunction with that publication.

In this document, we seek to understand (i) how evidence took root in medicine, which was not always so evidence-based, and the lessons for funders, policy-makers and others in education as to how its development can be accelerated in education [the first Getting Better document contained a brief section on the history of EBM, and we go deeper here]; and (ii) factors which have enabled public health to make such significant progress in developing countries. Our aim is to inform funders, policy-makers and practitioners to improve education by gaining an improved understanding of how medicine and public health have evolved in past decades.

Perhaps the most striking difference between medicine and education is that medical interventions can make things worse. Archie Cochrane concludes the reflections of treatment in Prisoner Of War camps, quoted on page 2 here, by saying “I feared I shortened some lives by doing it [pneumothorax] on the wrong cases.” This fear and suspicion of harming seems to have prompted the early thirst for evidence. By contrast, educational interventions rarely harm: the child rarely ends up knowing less than if the teacher had not been there. Hence the danger of bad interventions is less stark and less compelling in education than in health. The effect of bad educational interventions is that children develop less, or less fast, than good interventions would have enabled. For sure, there is an opportunity cost, but this is much harder to demonstrate and energise people around than is avoiding outright harm.

This is our challenge.

1 We use ‘health’, ‘healthcare’ and ‘medicine’ interchangeably, and use them to refer to healthcare in more economically developed countries (MEDCs), unless stated otherwise. By ‘education’, we mean school education in LEDCs, unless otherwise stated.
2 Introduction

Professor Sir Richard Peto FRS has been at the forefront of empirical medicine for decades, most famous for his work establishing the link between smoking and lung cancer. He founded the acclaimed Clinical Trial Service Unit at Oxford University and has run dozens of trials including the biggest drug trial ever seen. He made a keynote speech in Oxford in April 2015, beginning by saying that “most things in medicine either don’t work, or don’t work very well. The purpose of a trial is to work out which.” He then explained that the impressive drop in mortality from breast cancer (illustrating it with the graph in Figure 1) isn’t due to some wonder drug but rather to ‘dozens and dozens of boring, tiny tweaks’.

Figure 1: Breast cancer mortality rate

It is easy to be impressed by the gains achieved in healthcare and by medicine – e.g., the giant rises in life expectancy in the decades since the scientific approach really took root (shown below in Figure 2) – and wonder how and whether they could be replicated elsewhere, such as in education. Perhaps they can, and there certainly are many things which can be learnt. But we need to be careful: it’s easy to be beguiled by a few ‘silver bullets’ such as vaccines which really do look miraculous (see Figure 3), but it is important to remember that to deliver the silver bullet (and get a success like small pox or polio eradication), a robust health delivery system is required to be developed and maintained, which needs commitment, hard work and resources. Peto’s lesson is that most battles are like that: they involve painstaking work to find modest improvements, and then implement them correctly and consistently, every time.

Neither should health’s impressive gains delude us into thinking that medicine has figured it all out. Any number of senior doctors report that ‘medicine is broken’; 90% of the world’s health spend goes on 90% of the disease burden; a widely-cited leader in the British Medical Journal (BMJ) in 2014 posted that Evidence-Based Medicine (EBM) is ‘a movement in crisis’; and it’s thought that 85% of all medical research is wasted, at a cost of around $200bn every year. Half of it never even gets published.

The clear key lessons for education and other disciplines are around detail. Doing empirical investigations into what works, using robust research methods, ensuring that investigations are published, ensuring that the detail of the research is published so that others can judge whether it is robust, ensuring that interventions are properly described such that others can replicate them, synthesising the results of multiple investigations, monitoring outcomes fastidiously, studying how to improve implementation. One of the most striking features of EBM is that they study everything: how
to teach EBM, how best to synthesise multiple studies into systematic reviews, how to automate production of systematic reviews, how to gather prevalence data in the field, the effect of different ways of randomising people within a trial, what key elements are most commonly missing from trial reports, non-publication by geography, the extent of randomness in what research gets funded, optimal configurations of committees, and so on. This level of detail and curiosity are perhaps what marks medicine out.

Neither is new. Cholera was stemmed thanks to a detailed study of the precise location of incidences, which famously implicated the water pump in Broad Street in London’s Soho in 1846; Francis Bacon proposed quantitative descriptions of diseases and family size in his *History of Life and Death* in 1623; and ‘medical arithmetick’ (essentially prevalence mapping) was born in the 18th century, described in 1788 as ‘what trigonometry, geometry and the telescope are to the arithmetician and astronomer’.

Yet other disciplines have been slow to gather this kind of detail and adopt this curiosity. Even Archie Cochrane, the great architect of EBM, having seen the benefits of controlled trials and even having run them in prisoner of war camps undeterred by enemy bullets whizzing through his own hair, couldn’t get schools to use them to determine what works in disciplining pupils. He couldn’t even get medical colleges to use them to figure out how best to teach.

Figure 2: Life expectancy at birth in countries around the world 1540-2011

![Life expectancy graph](image-url)
Figure 3: Reduction of cases and death of vaccine preventable diseases in US after introduction of vaccines\textsuperscript{20}

<table>
<thead>
<tr>
<th>Disease</th>
<th>Cases</th>
<th>Deaths</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diptheria</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Measles</td>
<td>99.99%</td>
<td>100%</td>
</tr>
<tr>
<td>Mumps</td>
<td>97.4%</td>
<td>100%</td>
</tr>
<tr>
<td>Pertussis</td>
<td>96.6%</td>
<td>99.7%</td>
</tr>
<tr>
<td>Acute Poliomyelitis</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Paralytic Poliomyelitis</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Rubella</td>
<td>99.98%</td>
<td>100%</td>
</tr>
<tr>
<td>Congenital Rubella Syndrome</td>
<td>99.6%</td>
<td>No data</td>
</tr>
<tr>
<td>Smallpox</td>
<td>100%</td>
<td>100%</td>
</tr>
<tr>
<td>Tetanus</td>
<td>96.6%</td>
<td>99.6%</td>
</tr>
<tr>
<td>Hepatitis A</td>
<td>89%</td>
<td>88.7%</td>
</tr>
<tr>
<td>Acute Hepatitis B</td>
<td>83.9%</td>
<td>83.6%</td>
</tr>
<tr>
<td>Haemophilus Influenza type b</td>
<td>99.8%</td>
<td>No data</td>
</tr>
<tr>
<td>Pneumococca Disease</td>
<td>40.5%</td>
<td>31.3%</td>
</tr>
<tr>
<td>Varicella</td>
<td>87.2%</td>
<td>84.3%</td>
</tr>
</tbody>
</table>

The purpose of this document

This document aims to show something of how evidence in medicine is produced and organised, and how medicine came to be this way. The goal is that others will be able to find lessons in here for their disciplines, including for education in less developed countries. It should be read in conjunction with (ideally, after) Getting Better\textsuperscript{21}. It addresses some questions about EBM which were asked in response to Getting Better. We again structure our thinking by using an evidence system: the way that evidence is produced, synthesised, disseminated and used.
2.1 Contents and structure of this document

We start by looking at the disparity between spending on medical research and education research, and some observations and hypotheses about why this arises. We then look in further detail at the evidence system in public health in LEDCs, to see how it compares to the evidence system in education.

Then we look at some tools in medicine which have been helpful, and which could be more widely used / for which analogies could be found in education:

- standardised measures, such as Quality-Adjusted Life Years (QALYs),
- we look at how experimental designs developed and particularly at how randomised controlled trials (RCTs) began and took root,
- because RCTs vary wildly in how reliable they are, it’s important that the detail of how they were done is reported such that other clinicians can see whether they pertain to their patients and whether they’re reliable. The same is true of other experimental designs and studies, such as observational studies. We look at medicine’s checklists for reporting trials and experiments of various types and how those began and have been adopted,
- since it’s important that people can find those reports, medicine has indexed repositories of trials, and we look at those,
- and we notice how little research is done, even in medicine, in LEDCs – both primary research (studies of people) and secondary research (studies of studies, including systematic reviews).

Second, we turn to the history of medicine, and the growth in use and sophistication of evidence.

- medicine was not always so evidence-based, and other disciplines, including education, need to / are trying to make that transition. So how did they do it? We examine the transition, and the historical tradition and ‘building blocks’ on which it drew,
- we look at the role of information management in evolution of EBM,
- we look at some of the problems which remain,
- we look at a case study of major improvements in public health, specifically around cholera and oral rehydration therapy (ORT). We look at the role that evidence played in it: who created it, how it reached decision-makers, how important it was in the decisions, and how data about progress and prevalence were gathered and used.

Lastly, we pull out some of the lessons from all this for people in education in LEDCs (and other disciplines where evidence is not so advanced). What short-cuts are now visible? What has medicine built which others should avoid?

And finally

The period in which we wrote this report saw the passing of David Sackett. Sackett was one of the founders of EBM, coming from McMaster University in Canada to set up the Centre for Evidence-based Medicine in Oxford and the Clinical Trial Service Unit with Peto. He taught and studied EBM and championed the cause for decades. His method is salutary: having found the medical establishment to be “negative, condescending and dismissive” to the EBM approach, he engaged with people – not by publishing ‘the right answer’ in some journal – but rather through constant face-to-face engagement. He made teaching visits to more than 200 district general hospitals in the UK and to scores in Europe. All of us who aim to increase the use of evidence should remember that, perhaps ultimately, this is a ground offensive – and a charm offensive.
3 Disparity between spending on medical research and education research

Much more is spent on research in medicine than on research in education. We here present some evidence that this is the case, and then offer some thoughts as to why it might be the case.

Example 1: the UK
In 2011 the British government spent around £679m on medical research through the Medical Research Council (MRC) and a further £901m via the Department of Health. This was relative to the NHS budget of £104bn. The proportion of the total health budget going on research is therefore 1.52%. By contrast, the British government spent around £57bn delivering on education in 2012, of which £27m was spent on research: in other words, 0.05% of education spending. (Figure 4)

Figure 4: UK spend on research by the Department for Education and its predecessors (blue) and spending on research by the Department of Health and the MRC (red)^33
Example 2: the USA

Table 1 shows the amount spent by the US government on higher education research and medical (health) research during 1994-2012. Though we could not find data for primary education research, it is clear that medical research is much larger.

Table 1: US funding for higher education research and medical research 1994-2012

<table>
<thead>
<tr>
<th>Year</th>
<th>Higher education research</th>
<th>Medical research</th>
<th>Higher education research</th>
<th>Medical research</th>
<th>Higher education research</th>
<th>Medical research</th>
</tr>
</thead>
<tbody>
<tr>
<td>1994</td>
<td>$21.0 billion</td>
<td>$59.5 billion</td>
<td>$44.8 billion</td>
<td>$109.7 billion</td>
<td>$65.7 billion</td>
<td>$116.5 billion</td>
</tr>
</tbody>
</table>

Our thoughts about why this mismatch arises

This mismatch does not appear to be widely discussed in the literature. Our hypotheses as to why it arises include that:

- Health outcomes such as disease, infirmity or death are more easily measured. This means that gains are easier to identify, and they can be used to make a compelling case for further research.

- Poor health affects adults, including those who control budgets. It is scary, and obviously involves ‘the unknown’. By contrast, most education involves young people, who are separate to the decision-makers.

- Much global effort has gone to develop objective, reproducible, widely acceptable outcome measures for health conditions.

- There are dedicated national and international institutions which monitor health outcomes.

- And finally, a significant adverse health incident (such as Ebola or H1N1 pandemic) triggers a major international response.
4 Evidence system around public health in LEDCs

There are limitations with the way that primary research is produced in LEDCs, how it is synthesised, disseminated and used – that is, at all four stages of the evidence system:

1. Production: Much less primary research is conducted in LEDCs than in high-income countries.
2. Synthesis: Currently available systematic reviews do not reflect developing world priorities.
3. Dissemination: LEDC policy-makers and health care professionals have limited access to information in subscription journals.
4. Use: Many interventions for which there are systematic reviews are tough to implement in resource-poor situations, though perhaps they could be with help.

4.1 Production: Limited amount of primary research in LEDCs

Although the number of new scientific articles published worldwide each year is considerable, few journals report studies in LEDCs, where most preventable deaths occur. In 2000, just 1.7% of publications on public health came from LEDCs, though recent data suggest that scientific production on public health is rising in countries in transition such as China, India, and Brazil.

The low amount of primary research in LEDCs could be linked to the health spend. In 2001, about US$3.059 trillion—approximately 9 percent of global gross domestic product (GDP)—was spent on health care worldwide. However, only 12 percent of this amount was spent in LEDCs, which accounts for 84 percent of the global population and 92 percent of the global disease burden.

4.2 Synthesis: Currently available systematic reviews do not reflect developing world priorities

Systematic reviews have tended to reflect the priorities of developed countries rather than the global disease burden. An analysis in 2003 of nearly 3000 systematic reviews found that few of them focused on diseases and aspects of health care affecting large numbers of the world’s population. Figure 5 contrasts the big gap between the global disease burden of 11 diseases which are common in LEDCs (but not in the developed world) and R & D expenditure on those diseases.
There is a relative absence of reviews of many priority health issues in developing countries, such as child and maternal under-nutrition and environmental risks (sanitation and hygiene) – which account for 3 million deaths a year. For example, the last systematic review on sanitation and hygiene was carried out a decade ago.

The largest producer of systematic reviews in healthcare is the Cochrane Collaboration, a global network of researchers. A survey of the place of residence of Cochrane reviewers from 1997 to 2007 suggests that although the number of reviewers from developing countries increased, the proportion of reviewers from developing countries compared to developed countries declined from 16 percent to 10 percent (Figure 6).

Figure 6: Number and proportions of Cochrane reviews by location of contact author in developing/developed countries, 1997 to 2007

[Graph showing the number and proportions of Cochrane reviews by location of contact author in developing/developed countries from 1997 to 2007]

There is a relative absence of reviews of many priority health issues in developing countries, such as child and maternal under-nutrition and environmental risks (sanitation and hygiene) – which account for 3 million deaths a year. For example, the last systematic review on sanitation and hygiene was carried out a decade ago.
4.3 *Dissemination*: LEDC policy-makers and health care professionals have limited access to subscription-based information

Many medical publishers have started to give selected institutions access to journals in the developing world. While these initiatives are designed to be ‘help for self-help’ for developing countries, they have to be substantially increased to redress the current imbalances and to ensure that systematic reviews play a more prominent role in improving the global health outcomes.

Members of the Structured Operational Research and Training Initiative (SORT IT) suggest that policy-makers and health workers in LEDCs cannot afford the costs of US$ 35 – 40 per published article even if they are highly relevant to their programmes. In addition, annual institutional and personal subscription rates for a closed access journal such as The Lancet are simply unaffordable for most institutions, let alone individuals. They suggest that publishing corporations waive subscription fees or offer free on-line access for individual operational research articles from LEDCs, that all taxpayer-funded research be published as open access, and that the WHO HINARI programme of free journal access in LEDCs should be expanded to include not only institutions, but also individuals.

To reduce the burden among LEDC authors to submit research to open access journals, the authors suggest that annual research budgets from governments or international bilateral / multilateral donors include a line item to cover publishing fees. They also recommend that publishers establish a dedicated annual fund for this group, that there be ‘pooled funding’ from donors, and that open access journals have waivers or subsidies with differential pricing, even if it means a marginal increase in fees for the rest of the articles from authors outside LEDCs.
4.4 Use: Many interventions reviewed cannot be initially implemented in resource-poor situations, but with help, can be implemented

Even where reviews are nominally relevant, the treatment and interventions advocated may be unavailable or inappropriate in LEDCs. This problem is potentially solvable with assistance as per the case study below.

Case study 1: Antiretroviral drugs

Antiretroviral (ARV) drugs have the potential to dramatically improve the health and extend the lives of some people with HIV/AIDS. A Cochrane review of antiretroviral treatment (ART) for reducing the risk of mother-to-child transmission of the HIV infection concludes that ART will have an immediate benefit for countries with the resources to adopt such treatment.

In developed countries, ARVs are used frequently to reduce new HIV infections and to reduce occurrence of opportunistic infections and AIDS deaths. However, the high cost and demanding clinical infrastructure necessary put them out of reach of the vast majority of people with HIV. This problem is especially acute in developing countries, where HIV infection levels are high and public resources are extremely scarce.

The new millennium saw a breakthrough in treatment provision for resource poor areas when an Indian pharmaceutical company started to produce generic ARVs identical to those made by large pharmaceutical companies, but significantly cheaper. This sparked a price war between branded and generic drug makers, which forced the large pharmaceutical companies to lower their price. This competition, coupled with pressure from activists, organisations such as the Clinton Foundation and governments of poor countries with severe HIV epidemics, dramatically reduced the price of ARVs for developing countries. By the middle of 2001, ART was available from Indian generic manufacturers for as little as $295 per patient per year (PPY).

The price of ARVs for LEDCs has continued to fall. In 2013, the average cost of first-line ART for LEDCs was $115 PPY and $330 PPY for second-line ART. The price of third-line ART has also decreased but LEDCs still pay on average, more than $1500 PPY. Many LEDCs in Asia, Latin America, Eastern Europe and Central Asia continue to pay higher prices due to their inability to access cheaper generic ARVs.

It is important to remember the difference between intervention efficacy and effectiveness. Studies need to explore more than intervention efficacy under trial conditions. In order for evidence to inform decision-making, evaluation needs to address the prevailing problems, the complex pathways to health outcomes, and the effectiveness of interventions in local contexts.
It is also important to understand the conflicts between evidence and local priorities / politics as illustrated by the following case study.

**Case Study 2: Malaria control**

Malaria transmission occurs in 106 countries, notably in Africa. In 2013, an estimated 198 million cases of malaria occurred, causing an estimated 584,000 deaths, mainly of children under five.\(^{42}\)

In recent years, the expansion of proven and cost-effective vector control interventions, has led to a 26% reduction in malaria mortality globally and an estimated 33% drop in Africa alone.

To date, larval source management has been a marginal intervention in Africa, applied mainly in urban, low-burden and elimination settings. It is not considered to have made a tangible contribution to the recent successes. Despite this, larval control interventions have attracted a significant amount of political attention and support in recent years. Most headlines have been generated by larviciding, which involves regularly applying a biological or chemical insecticide to water bodies to reduce the number of mosquito larvae and pupae.

The growing interest in larviciding has been fuelled by a desire to find new solutions for malaria control in Africa and to reduce dependence on foreign aid and on foreign-manufactured commodities. Larviciding has received priority status under a tripartite agreement, signed in 2009 between Venezuela, Cuba and the Economic Community Of West African States (ECOWAS), a regional group of fifteen countries.

Following a 2012 review by its Malaria Policy Advisory Committee, WHO now recommends larviciding only as a supplementary measure where mosquito breeding sites are few, fixed and findable.\(^{43}\) The WHO position has been supported by a recent Cochrane review.\(^{44}\)

The WHO Malaria Policy Advisory Committee will review additional evidence as it becomes available. Until there is more compelling evidence, the WHO view is that larval control should continue to be viewed as a supplementary measure for malaria control in carefully selected settings. Promoting the widespread use of larval source management in rural areas of sub-Saharan Africa would be premature, and would divert scarce resources from effective interventions.\(^{45}\)
4.5 Use: the extreme difficulty of implementation

Harvard surgeon Atul Gawande talked in his 2014 Reith Lectures of the lessons from many examples of medical advances. He cites an essay by Samuel Gorovitz and Alasdair MacIntyre in 1976 on the nature of human fallibility: causes of which include:

1. Ignorance: we don’t know what to do

2. Ineptitude: the knowledge exists but an individual or a group of individuals fail to apply that knowledge correctly. There’s “some shame or guilt to the fact that we don’t get it right all the time. And exposing it can make people more angry than exposing the fact that we’re simply ignorant”

3. Necessary fallibility: when we don’t - can’t - know all the relevant factors in enough detail.

“You go back a hundred years, and we lived in a world where our futures were governed largely by ignorance” he says, but now we’re largely “struggling with ineptitude as struggling with ignorance… not only how we close the continuing gaps of ignorance but how we ensure that the knowledge gets there”.

“I think we have been fooled by Penicillin. Discovered in 1929, it was almost 20 years before it was mass produced and could stop disease. And [then] it was like a miracle…this treatment that could eliminate whole classes of disease – a whole body of bacterial infections that we basically thought, you know, you couldn’t do much of anything about. It came as a miracle because it was so easy - just an injection. That made us imagine that this was the future of medicine, of healthcare in general; that we would just have an injection for cancers, for heart disease, for stroke.

“But from this last century of discovery, very little of it has turned out to be like Penicillin. In fact, we’re facing extreme complexity.” This gives rise to his interest in systems to (a) get the information to the right place (overcome ignorance) and (b) implement it properly and consistently (to overcome ineptitude).

“We have made tremendous discoveries, but find it’s extremely complex to deliver on them… The volume of knowledge and skill has exceeded our individual capabilities.”

This actually doesn’t seem to be the problem in education, because there has been so much less research so (arguably) the problem remains that there are many aspects of ‘what works’ that are still unknown. But nonetheless it is salutary about the need to build systems for implementation:

“The important insight is that what we have to focus on is how to deliver on the guidelines and standards and knowledge that we have discovered, how to make it easy for everybody to follow… “When we think about the transformation of human health that’s occurred during the last century, we tend to focus on the discoveries - on the new drugs and the new procedures. Julio Frenk, dean of the Harvard School of Public Health, points out that there was another major contribution during this past century: the emergence of complex, specialised institutions for delivering those discoveries to people - hospitals, clinics, professionals who could staff them.”
5 Useful tools developed in medicine

5.1 Standardised units used in medicine

As commitment to monitoring health and rational allocation of health resources has grown internationally, so too have the methods that researchers and policy-makers use to evaluate health and medical outcomes in individuals and in populations.

In medicine, to map the ‘amount’ of illness and to choose between treatments, three units are used: Quality-Adjusted Life Years (QALYs), Disability-Adjusted Life Years (DALYs) and Health-Adjusted Life Years (HALYs). We discuss them in turn and see how they came to be developed.

5.2 Quality-Adjusted Life Years (QALYs)

5.2.1 Definition

QALYs are a common unit to compare types of interventions and programmes, and hence are used in assessing the value for money of a medical intervention. They extend the idea of life expectancy by incorporating both the quality and the quantity of life lived. Rather than count every year of life lived as though they were equivalent, QALYs ‘discount’ years lived in a state of ill-health: those years are counted as being worth less than a year of healthy life.

5.2.2 Development of QALYs

QALYs were developed in the late 1960s by economists, operations researchers, and psychologists, primarily for use in Cost-Effectiveness Analysis (CEA). They followed early work to develop a descriptive measure combining time lived with functional capacity. The term QALY emerged in 1976 in an article by two economists, Zeckhauser and Shepard to indicate a health outcome measurement unit that combines duration and quality of life. The underlying concept was formally shaped in the early 1970s in the development of a health status index, while an earlier study of the treatment of chronic renal disease had used a subjective adjustment for quality of life. Early applications of the health status index included one on tuberculin screening and one on screening for phenylketonuria.

QALYs were accepted relatively fast. A review in 1992 counted 51 economic evaluations using QALYs as the outcome measure. Only a few years later, the QALY framework was widely accepted as the reference standard in cost-effectiveness analysis amid a continuing debate on its theoretical underpinnings and practical implications.

5.2.3 Use of QALYs

QALYs enable a measure of disease burden, and are routinely used in assessments of medical care, technology, and public health interventions - these studies have proliferated over the past two decades. QALYs are often used in cost-utility analysis to calculate the cost/benefit ratio for a particular healthcare intervention. This is used to allocate health care resources, with an intervention with a lower cost to QALY saved (Incremental Cost Effectiveness Ratio, ICER) being preferred over an intervention with a higher ratio.
QALYs represented an important breakthrough in conceptualising the health outcome (denominator) in a cost-effectiveness (CE) ratio. A CE ratio describes the incremental price of a unit of health effect from a health intervention—be it preventive or curative, population-based, or clinical—when compared with an alternative intervention.

\[
\text{CE ratio} = \frac{\text{cost}_{\text{new strategy}} - \text{cost}_{\text{current practice}}}{\text{effect}_{\text{new strategy}} - \text{effect}_{\text{current practice}}}
\]

When the denominator of the CE ratio is computed using QALYs, the cost-effectiveness analysis is referred to as cost-utility analyses (CUA).

\[
\frac{\text{Cost Intervention B - Cost Intervention A}}{\text{QALY B - QALY A}}
\]

Cost-utility analysis (CUA) is appropriate in situations where quality of life is “the” or “an” important outcome of health care.

Given a specific budget constraint, QALYs are maximised by increasing the “utility” of individuals and aggregates of individuals.\(^6\) The original formulation of QALYs was drawn from the theoretical underpinnings of welfare economics and expected utility theory.\(^6\) In welfare economics, social utility is the aggregate of individuals’ utilities, and economists hold that maximising social utility is the primary goal for resource allocation. QALYs are often seen as inexorably linked with utilitarianism, a social theory that dictates that policies be designed to do the greatest good for the greatest number of people. A summary of the different types of indicators is given in Table 2.
5.2.4 Role of institutions in promoting QALYs

QALYs were promoted through medical technology assessment conducted by the US Congress Office of Technology Assessment. The National Institute for Health and Care Excellence (NICE), an executive non-departmental public body of the UK Department of Health uses QALYs as the primary outcome for quantifying the expected health benefits associated with a given treatment regime. NICE regulates treatments which can be provided by the UK’s National Health Service, in large part on the cost per QALY; it will normally not pay more than £20,000 per QALY.

Table 2: Advantages and disadvantages of different type of analysis, with examples

<table>
<thead>
<tr>
<th>Type of analysis</th>
<th>Type of benefit</th>
<th>Example outcome</th>
<th>Advantages</th>
<th>Disadvantages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost-minimisation</td>
<td>Fixed outcome</td>
<td>£ per medical procedure</td>
<td>Simple to conduct</td>
<td>Restricted to one technology</td>
</tr>
<tr>
<td>Cost-effectiveness</td>
<td>Clinical outcome</td>
<td>£ per 20 mm reduction in blood pressure</td>
<td>Straightforward outcome measures (life gained, blood pressure reduced) Frequently incorporated within RCTs, therefore least biased</td>
<td>One-dimensional Limited comparability e.g. Interventions aimed at increasing life years gained cannot be directly compared with those that improve physical functioning</td>
</tr>
<tr>
<td>Cost-utility</td>
<td>QALY</td>
<td>£ per QALY</td>
<td>The use of single measure of health benefit enables diverse health care interventions to be compared</td>
<td>Complex to conduct Practical problems in valuing utility arising from different measuring instruments, health problems and interventions</td>
</tr>
<tr>
<td>Cost-benefit</td>
<td>Currency (£, $ etc.)</td>
<td>£</td>
<td>Allows comparison across sectors (e.g. road building, schools)</td>
<td>Places monetary value on life which is considered priceless Practical problems in valuing health</td>
</tr>
</tbody>
</table>
5.3 Disability-Adjusted Life Years (DALYs)

Disability-adjusted life years (DALYs) are a measure of life years lost because of disease, adjusted for assumptions about disability as well as the impact of age and future time. They were launched to widen the measurement of disease from the presence of morbidity and mortality, usually cited by the World Bank, to include the impact on disability in a commensurable way with mortality.

DALYs were originally developed by Harvard University for the World Bank and published in 1990 in the *Global Burden of Disease* study. DALYs were used in 1993 in the World Development Report ‘Investing in Health’. This report was a collaboration between the World Bank (who is a leading source of funding in the health sector amongst international agencies alongside its role as a development institution) and WHO. The report was ground-breaking for two reasons:

- It was the first attempt to assess the global burden of disease by region of the world in this way; and
- It was the first attempt to bring together results from various cost-effectiveness analyses of health interventions by disease using one outcome measure (DALYs) to recommend global health policies.70

The World Health Organization (WHO) subsequently adopted the method in 1996 as part of the Ad-hoc Committee on Health Research *Investing in Health Research & Development* report. It is now a key measure employed by the WHO in such publications as its *Global Burden of Disease* (see section 5.3.3)72

5.3.1 Why were DALYs developed?

DALYs were developed to help shape health policies of international institutions such as the World Bank and WHO. Their introduction was designed to broaden the usual focus of these institutions from measuring disease in terms of mortality and morbidity to including an estimate of the impact of morbidity. As the focus of the institutions was international, the DALY was intended to enable many forms of comparison: across diseases, countries, curative/preventive care as well as different time periods.

DALYs have two prime purposes: (1) as an input to the calculation of the global burden of disease; and (2) as an outcome measure for use in cost-effectiveness analysis. Both were intended to influence:

- Prioritisation of health care spending within and across countries for curative and preventive care.
- Flow of funds within health research and development.
- Identification of disadvantaged groups for targeting health interventions.
- Composition of training for clinical and health practitioners.
- Methods for assessing performance in health projects and health systems.
5.3.2 Table 3: How DALYs have been used

<table>
<thead>
<tr>
<th>Type of use</th>
<th>By DALY developers</th>
<th>By others</th>
</tr>
</thead>
<tbody>
<tr>
<td>Burden of disease</td>
<td>WHO/World Bank global burden of disease exercise</td>
<td>National (e.g., Australia) and sub-national (e.g., regions in the UK) burden of disease; individual diseases</td>
</tr>
<tr>
<td>Cost-effectiveness analysis</td>
<td>The World Bank health sector priorities review for low/middle income countries</td>
<td>Cost-effectiveness analysis of alternative treatments in specific low-income countries</td>
</tr>
<tr>
<td>Sectoral analysis</td>
<td>Mexico (country)</td>
<td>Turkey (country)</td>
</tr>
</tbody>
</table>

5.3.3 The Global Burden of Disease

A Global Burden of Disease (GBD) study aims to guide policy-makers by quantifying the burden of premature mortality and disability for major diseases or disease groups. It uses a summary measure of population health, the DALY, to combine estimates of the years of life lost and years lived with disabilities. The GBD project was commissioned in 1990 and is a collaborative effort between hundreds of experts worldwide, including researchers at the WHO, Harvard School of Public Health, the University of Auckland School of Population Health, the Institute for Health Metrics and Evaluation (IHME), and the World Bank. It provided a standardised approach to epidemiological assessment and uses a standard unit, the DALY, to aid comparisons. The original 1990 project estimated health gaps using DALYs for eight regions of the world.

The GBD report in 1996 constituted the most comprehensive and consistent set of estimates of mortality and morbidity yet produced and WHO regularly develops GBD estimates at regional and global level for a set of more than 135 causes of disease and injury. The data are also broken down by age, gender and region. WHO also supports National Burden of Disease (NBD) studies to obtain country-specific estimates for input to national policy. The national studies are based on the GBD concept and the data can be used in Environment Based Diseases (EBD) assessments to estimate the contributions that environmental risk factors make to the overall disease burden. Over 30 countries are now undertaking NBD studies.

5.3.4 Summary measures of population health

Summary measures of population health measure the health of a population by combining data on mortality and non-fatal health outcomes into a single number. Besides the DALY, several other such measures have been devised, including the Quality-Adjusted Life Year (QALY), the Disability-Adjusted Life Expectancy (DALE) and the Healthy Life Year. The benefits and challenges of these measures have been examined. As the DALY has been the most widely-used measure, and can be applied across cultures, we will focus on it in this report. The DALY measures health gaps as opposed to health expectancies. It measures the difference between a current situation and an ideal situation where everyone lives up to the age of the standard life expectancy, and in perfect health. Based on life tables, the standard life expectancy at birth is set at 80 years for men and 82.5 for women.
The DALY combines in one measure the time lived with disability and the time lost due to premature mortality. DALYs for a disease or health condition are calculated as the sum of the Years of Life Lost (YLL) due to premature mortality in the population and the Years Lost due to Disability (YLD) for people living with the health condition or its consequences:

**DALY = YLL + YLD**

The YLL basically correspond to the number of deaths multiplied by the standard life expectancy at the age at which death occurs. The basic formula for YLL (without including other social preferences which are too complex to be discussed here), is the following for a given cause, age and sex:

**YLL = N x L**

where:
- N = number of deaths
- L = standard life expectancy at age of death in years.

To estimate YLD for a particular cause in a particular time period, the number of incident cases in that period is multiplied by the average duration of the disease and a weight factor that reflects the severity of the disease on a scale from 0 (perfect health) to 1 (dead). The basic formula for YLD is the following (again, without applying social preferences):

**YLD = I x DW x L**

where:
- I = number of incident cases
- DW = disability weight
- L = average duration of the case until remission or death (years).

As noted above, the DALY measures the gap between the actual health status of a population and some “ideal” or reference status, using time as the measure. In developing the DALY indicator, two key value issues were identified:

- how long “should” people in good health expect to live?
- how should we compare years of life lost through death, with years lived with poor health or disability of various levels of severity?

The first of these choices relates to the standard life expectancy used to calculate the YLL, and the second to the development of disability weights for YLD.

**5.3.5 Health Adjusted Life Years**

Health-Adjusted Life Years (HALYs) are an umbrella term for DALYs and QALYs. They measure a population's health, allowing the combined impact of death and morbidity to be considered simultaneously. This feature makes HALYs useful for comparing illnesses, interventions, and populations. A 1998 Institute of Medicine report found these measures to be “increasingly relevant to both public health and medical decision makers” and of late, HALYs have gained higher visibility in policy circles internationally. Health-Adjusted Life Expectancy (HALE) is a related summary measure of population health which estimates the average time in years that a person at a given age can expect to live in the equivalent of full health. Life tables, such as those created by the US census, are combined with cross-sectional age-specific Health Related Quality of Life (HRQL) data.
Various concerns are voiced about HALYS:

- Sometimes HALYs can be used in a way which some people find objectionable and that approach is absence of incorporating a socially sanctioned and empirically valid “equity weight,” distributional effects of resource allocation based solely on HALYs. In broad terms, HALYs combine life years and quality of life related with health. For example, if a patient lived for 93 years but the years he lived in full health is 85, that is considered as HALYs.
- “Aggregation”, the way that values for health states and diseases are combined across individuals, as well as along the spectrum of alive to dead. Critics of HALYs say that failure to treat life-saving interventions as conceptually distinct from health-improving interventions is at odds with how society views life and death medical decisions.77
- Similarly, some people question whether minor benefits accruing to many people should be viewed as equivalent to more significant benefits accruing to few.78

5.4 Randomised controlled trials (RCTs)

5.4.1 Origins of randomised trials – in psychology and agriculture
Randomised experiments first appeared in psychology, where they were introduced by Charles Sanders Peirce,79 and in education.80 Later, randomised experiments appeared in agriculture, due to Jerzy Neyman81 and Ronald A. Fisher. Fisher’s experimental research and his writings popularised randomised experiments.82

5.4.2 Introduction of RCTs in medicine
The first randomised controlled trial (RCT) in medicine is credited to Sir Austin Bradford Hill, an epidemiologist at the UK’s Medical Research Council, who, in 1948 tested whether streptomycin is effective in treating tuberculosis. Oddly, the need to create a control group which received no treatment arose because streptomycin was in short supply (like much else, following the war). Hill devised randomisation partly to keep doctors from trying to manoeuvre their patients into the streptomycin arm of the trial.

Earlier in his career, Hill believed that simply alternating the assignment of hospital admissions to drug versus control worked well enough. Later he recognised that simple alternation led to selection bias because the sequence was too easy to predict. That realisation led to the use of a random numbers table to generate the numeric series by which patients would be assigned to conditions.

5.4.3 Principles and limitations of RCTs
The basic principles of this model are (1) comparison, under controlled conditions, of two or more therapeutic regimens (one of which may be a traditional treatment, a placebo, or the exclusion of active treatment), and (2) statistical analysis of the possibility of error. The recognised methodology is the RCT with its associated features of control groups, randomisation, and blinding.

The RCT is by no means a straightforward solution, as even its advocates agree. On the one hand, the principle of comparison often means that one set of subjects will receive a less effective treatment, or possibly none at all, a situation which may sometimes be ethically questionable. On the other, the logistics of designing and carrying out a trial within the real-world constraints of cost, time, and personnel require that the investigators select certain subjects for treatments, specify outcome measures and criteria, and set limits to the duration of treatment and follow-up. These necessary choices and exclusions may affect the statistical result of the RCT or cast doubts on its external validity.
Recognising the legitimacy of certain objections, researchers often attempt to accommodate them in the design of an RCT. In making these accommodations and implementing a study in the real (sometimes hostile) world, certain methodological allowances must be made. The researcher here has been forced by circumstances to depart from the ideal textbook design. But without these methodological accommodations, the RCT would never have been permitted in the first place. These allowances, which are forced on researchers by practical considerations, are seized upon by critics to discredit RCTs.

In the 20th century, to discover the hidden causes of unpredictable and unknown responses to treatment, medical researchers and statisticians developed a mathematical model to describe and calibrate the complex responses of the human body to therapeutic interventions.

5.4.4 RCTs as ‘gold standard’ for a trial
Despite its limitations described above, the RCT remains the “gold standard” for single studies. Its power rests on its ability to isolate a single factor.

The inferential authority of the RCT has been such that it is accepted as a standard for “rational therapeutics” by physicians and regulatory authorities and also by patients and populations at risk. In the late 1980s, for example, groups such as the Institute for Research on Women’s Health documented the exclusion or artificial restriction of women from clinical trials, even when the disease in question affected both sexes, and the scarcity of trial evidence on problems specific to women, such as menopause. In 1991, the National Breast Cancer Coalition challenged the cancer research establishment to carry out trials on new and innovative treatments. In effect, women demanded inclusion in clinical trials and the production of trial evidence specific to their needs, such as menopause. The following decade saw the creation of the US National Institutes of Health (NIH) Office for Research on Women’s Health and the institution of a number of gender-specific and gender-comparative trials. Today, guidelines for Public Health Service (PHS) grant applications specifically require the inclusion of women unless there is a valid reason for exclusion.

ACT UP (an AIDS Coalition) have also demanded RCT expansion, lobbying for more trials, for the inclusion of more subjects and more minorities, and for the use of active drugs rather than placebo controls. ACT UP proposed the redesign of trial methodologies to make them more sensitive to patient needs, in particular the replacement of life-or-death outcome criteria with surrogate markers of therapeutic efficacy, such as CD4 cell counts. ACT UP’s emphasis on “participatory knowledge making” has been adapted by advocacy groups for Lyme disease, breast cancer, and chronic fatigue syndrome. Although all these groups may criticise trial procedures, they do not reject trial evidence; rather, they seek to participate in its production and to find ways to combine statistical rigor with sensitivity to patient needs. As one activist told sociologist Steve Epstein, “It’s about having good science that develops good therapies so that we may have a cure or therapy someday.”

As the clinical trial has evolved in the last 100 years, physicians and scientists—and subjects as well—have faced the same challenge: how to develop “good therapies” based on “good science,” science that imposes order on, but neither distorts nor devalues, individual human experience. The RCT is a dynamic methodology, and its present and future are informed by its history.
Archie Cochrane and social RCTs

The introduction of RCTs into medicine was very difficult, as were some early attempts by Archie Cochrane to use them in non-medical settings.

Cochrane’s interest in reliable evidence began when he himself was referred for medical treatment and becomes suspicious that there was no evidence for it. He runs tests whenever he can, even in a ‘hospital’ in POW camp in Germany in World War II: a situation so dreadful that the latrine routinely overflows, an orderly is randomly executed at the door and at one point, he himself feels a German bullet whizz through his own hair. On one trial in such circumstances, he berates himself that “it was a pretty awful trial…the numbers were too small and they were not randomised”.

Later, in the 1960s, he is director of the new epidemiology unit at the University of Cardiff and suggests evaluating the teaching of anatomy (“which I thought too detailed and time consuming”) by randomising students between Cardiff, which didn’t teach it, and Bristol, which did. The proposal “was laughed out of court”.

He suggests RCTs to various civil servants he encounters in the Home Office, who he finds “consistently exhibited hysterical reactions to the mention of randomised controlled trials”.

He discusses with several magistrates that punishments for boys “caught breaking and entering” seemed to follow no pattern and that they could be randomised to determine effectiveness. “I…was horrified when they reacted like elderly physicians…They too suffered from the God complex. They knew what to do without the help of any trials. It was depressing.”

He tries to find a headmaster to randomise caning. “Few were prepared to listen for long.” One did, briefly, but then declined, “saying ‘No. When a boy is caught smoking, I know whether he should be caned or not!’” He then more or less gives up. “I had never imagined that sociological research could prove so difficult.”

Introducing rigorous research to medicine is much easier. He gets interested in the 1970s in why some countries do RCTs and others do not. The answer is perhaps relevant to introducing them into other disciplines. He concludes that “the answer has much…to do with medical education. In the Protestant North West [of Europe, medical education] has for some time been more scientific than elsewhere.” As discussed elsewhere, teacher training is rarely terribly scientific. Though the pattern isn’t uniform: anomalies include Germany where “an attempt was made to have [RCTs] declared illegal” – quite possibly from revulsion at Nazi experiments on live people during the war – and Sweden, “where medical education is very scientific, but they do few trials”, which he can’t explain.
5.5 Checklists for reporting trials and experiments

5.5.1 Evolution of Introduction, Methods, Results, And Discussion (IMRAD) structure for reporting experiments and results

Since its origin in the 17th century, the scientific paper has been through many changes. Although during the first two centuries, its form and style were not standardised, the letter form and the experimental report coexisted. The letter was usually single authored, written in a polite style, and addressed several subjects at the same time. The experimental report was purely descriptive, and events were often presented in chronological order. It evolved to a more structured form in which methods and results were incipiently described and interpreted, while the letter form disappeared. Method description developed during the second half of the nineteenth century, and an overall organisation known as “theory—experiment—discussion” appeared. In the early 20th century, standards began to emerge and the literary style waned.

By the 1920s, the structure of Introduction, Methods, Results, And Discussion (IMRAD) had emerged and was considered the ideal outline for scientific writing. However it was not widely used. Until 1945, articles were organised in a manner more similar to a book chapter, mainly with headings associated with the subject, and did not follow any particular structure. From 1950 to 1960, the IMRAD structure was partially adopted, and after 1965, it began to predominate, attaining leadership in the 1980s.

A small group of editors of general medical journals met informally in Vancouver, British Columbia, in 1978 to establish guidelines for the format of manuscripts submitted to their journals and became known as the Vancouver Group. Its requirements for manuscripts, including formats for bibliographic references developed by the National Library of Medicine, were first published in 1979.

Neither the rate at which the use of this format increased nor the point at which it became the standard for today’s medical scientific writing is well established. Sollaci et al carried out a 50 year survey of four leading medical journals of internal medicine, but did not find definite reasons explaining the rise of the IMRAD structure in the literature. They suggested that sciences other than medicine (e.g., physics) might have influenced the growing use of this structure.

5.5.2 Standardised format for reporting clinical trials

A well designed and properly executed RCT is the best design for a single study of the efficacy of a healthcare intervention. However, trials are often badly designed, and/or badly executed, and inadequate methods are associated with bias, especially exaggerated treatment effects. Critical appraisal of the quality of clinical trials is possible only if the design, conduct, and analysis of RCTs are thoroughly and accurately described in the report.

However, reporting of RCTs was (and remains) incomplete, as documented many times. For example, information on the method used in a trial to assign participants to comparison groups was reported in only 21 percent of 519 trial reports indexed in PubMed in 2000 and only 34 percent of 616 reports indexed in 2006. Similarly, only 45 percent of trial reports indexed in PubMed in 2000 and 53 percent in 2006 defined a primary end point, and only 27 percent in 2000 and 45 percent in 2006 reported a sample size calculation. Reporting is not only often incomplete but also sometimes inaccurate. Of 119 reports stating that all participants were included in the analysis to which they were originally assigned (intention-to-treat analysis), 15 (13 percent) excluded patients or did not analyse all patients as allocated.
5.5.3 Developing standardised content for trial reports: the origins of CONSORT

In 1993, 30 experts including medical journal editors, clinical trialists, epidemiologists, and methodologists met in Ottawa, Canada to develop a new scale to assess the quality of reports of clinical RCTs. However, during preliminary discussions, participants felt that many of the suggested scale items were irrelevant because they were not normally reported by authors. Therefore, the group unanimously agreed to focus on improving reporting of RCTs.

Participants nominated items to be included in a checklist for reporting trials: items which, if not reported, are often associated with poor design and biased estimates of the effects of the intervention under investigation. One outcome of the meeting was the Standardized Reporting of Trials (SORT) statement. This consisted of a 32-item checklist and a flow diagram in which investigators were encouraged to report on the various aspects of how an RCT was conducted.

Concurrently, and independently, another group of experts, the Asilomar Working Group on Recommendations for Reporting of Clinical Trials in the Biomedical Literature, convened in Asilomar, California, was working on a similar mandate. This group also published a proposal which included a checklist of recommended items for authors to consider when reporting RCTs.

In a subsequent editorial, the two groups were urged to develop a common set of recommendations. The group met in 1996 to merge the best of the SORT and Asilomar proposals into a single, coherent evidence-based recommendation, which might better appeal to journals. The meeting resulted in the Consolidated Standards of Reporting Trials (CONSORT) Statement, first published in 1996.

CONSORT also comprises a checklist and flow diagram for reporting an RCT (Appendix 1). They are primarily intended for use in writing, reviewing, or assessing reports of simple two-group parallel RCTs. Extensions to the CONSORT checklist for reporting trials with some other designs have been published, as have those for reporting harms, other types of interventions (non-pharmacological treatments, herbal interventions), and abstracts. A version of CONSORT is being developed at Oxford University for social and psychological interventions.

5.5.4 Endorsement and enforcement

Since its publication in 1996, CONSORT has been supported by more than 400 journals and several editorial groups, such as the International Committee of Medical Journal Editors (ICMJE - the Vancouver Group), the Council of Science Editors (CSE), and the World Association of Medical Editors (WAME). CONSORT is published in Dutch, English, French, German, Japanese, and Spanish. It can be accessed together with other information on the CONSORT website.

However, many journals do not enforce CONSORT – and even some which endorse it do not in fact enforce it. Hence compliance is patchy (see below).

5.5.5 The effect of CONSORT

While the assessment of RCTs described below indicates that using CONSORT helps to improve the quality of reports of RCTs, the effect “on practice and on patients” of CONSORT is far from clear.

In an assessment of 71 RCTs published in three journals in 1994, allocation concealment was not clearly reported in 43 (61 percent) of the trials. Four years later, after these three journals required that authors reporting an RCT use CONSORT, the proportion of papers in which allocation concealment was not clearly reported had dropped to 30 of 77 (39 percent). Of course, CONSORT itself is a non-randomisable intervention and hence that drop can’t be unambiguously attributed to CONSORT.
The usefulness of CONSORT is increased by continuous monitoring of biomedical publications, which allows it to be modified dependent on the merits of maintaining or dropping current items, and including new items. For example, when Meinert observed that the flow diagram did not provide important information about the number of participants who entered each phase of an RCT (i.e., enrolment, treatment allocation, follow-up, and data analysis), the diagram was modified to accommodate the information.

The checklist is similarly flexible. This iterative process makes the CONSORT statement a continually evolving instrument. Although participants in the CONSORT group and their degree of involvement vary over time, members meet regularly to review the need to refine CONSORT. For instance, at the 1999 meeting, participants decided to revise the original statement, partly in response to emerging evidence on the importance of various elements of RCTs. Further meetings of the group in 1999 and 2000 led to the revised CONSORT statement 2001. Following a meeting in 2007, a further revision was developed into the CONSORT 2010 statement (Appendix 1 and Appendix 2).

The most telling indicator of CONSORT’s impact is its uptake by journals and effect on the completeness of reporting of trials. CONSORT is currently known to be endorsed by over 600 biomedical journals and several prominent editorial organisations including the ICMJE and WAME. Journal endorsement of CONSORT and other reporting guidelines typically occurs in the form of a supportive statement in a journal’s Instructions to authors.121

A 2012 Cochrane systematic review assessed the effect of journals’ endorsement of CONSORT on the reporting of trials they publish.122 In 50 included studies evaluating the reporting of 16,604 trials, 25 of 27 CONSORT-related items measured were more completely reported in trials published in endorsing journals than those in non-endorsing journals; five items were significantly better reported. As an example, how “allocation concealment” was achieved, a defining feature of randomised trials was reported in 81 percent more trials published in endorsing versus non-endorsing journals. Similar findings were yielded for many items when comparing trials published in journals before and after CONSORT endorsement.

However, upon closer examination, these numbers are not altogether encouraging. Even among endorsing journals, the reporting of key methodological items was still dismal. Allocation concealment, for instance, while significantly better reported in endorsing journals, was reported in only 45 percent of trials compared to 22 percent of trials in non-endorsing journals. Many other key features of trials such as reporting the methods of sequence generation and defining the primary outcome were also reported less than 50 percent of the time in endorsing journals.
Figure 7 and Figure 8 indicate the success in adherence to CONSORT.

Figure 7: Differences in the completeness of reporting of CONSORT items between trials published in endorsing vs non-endorsing journals.123

<table>
<thead>
<tr>
<th>CONSORT Checklist Item</th>
<th># of Evaluations</th>
<th># of RCTs</th>
<th>RR</th>
<th>99% CI</th>
<th>Favours Non-Endorsement</th>
<th>Favours Endorsement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title and Abstract</td>
<td>7</td>
<td>1,233</td>
<td>1.13</td>
<td>(0.96, 1.33)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>5</td>
<td>513</td>
<td>1.07</td>
<td>(1.01, 1.14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>6</td>
<td>693</td>
<td>0.95</td>
<td>(0.56, 1.62)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interventions</td>
<td>6</td>
<td>638</td>
<td>1.00</td>
<td>(0.95, 1.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objectives</td>
<td>6</td>
<td>540</td>
<td>1.01</td>
<td>(0.96, 1.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcomes</td>
<td>8</td>
<td>1,302</td>
<td>1.17</td>
<td>(0.95, 1.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample Size</td>
<td>11</td>
<td>1,843</td>
<td>1.61</td>
<td>(1.13, 2.29)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequence Generation</td>
<td>14</td>
<td>2,231</td>
<td>1.59</td>
<td>(1.38, 1.84)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allocation Concealment</td>
<td>16</td>
<td>2,396</td>
<td>1.81</td>
<td>(1.25, 2.62)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implementation</td>
<td>5</td>
<td>486</td>
<td>1.47</td>
<td>(0.65, 3.32)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinding of Participants</td>
<td>5</td>
<td>711</td>
<td>1.39</td>
<td>(0.87, 2.23)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinding of Intervention</td>
<td>5</td>
<td>710</td>
<td>1.39</td>
<td>(0.74, 2.12)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinding of Outcome Assessor</td>
<td>5</td>
<td>719</td>
<td>1.72</td>
<td>(0.69, 4.30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinding of Data Analyst</td>
<td>3</td>
<td>497</td>
<td>3.56</td>
<td>(0.40, 31.6)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinding Any Description</td>
<td>8</td>
<td>1,851</td>
<td>1.23</td>
<td>(0.93, 1.52)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical Methods</td>
<td>9</td>
<td>894</td>
<td>1.03</td>
<td>(0.90, 1.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant Flow</td>
<td>16</td>
<td>2,461</td>
<td>1.16</td>
<td>(0.94, 1.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recruitment</td>
<td>6</td>
<td>959</td>
<td>1.03</td>
<td>(0.75, 1.41)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline Data</td>
<td>5</td>
<td>528</td>
<td>1.07</td>
<td>(0.94, 1.22)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numbers Analysed</td>
<td>13</td>
<td>2,145</td>
<td>1.23</td>
<td>(0.98, 1.55)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcomes and Estimation</td>
<td>6</td>
<td>617</td>
<td>1.00</td>
<td>(0.95, 1.05)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ancillary Analyses</td>
<td>4</td>
<td>378</td>
<td>1.31</td>
<td>(0.48, 3.58)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse Events</td>
<td>8</td>
<td>911</td>
<td>1.14</td>
<td>(0.86, 1.52)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpretation</td>
<td>5</td>
<td>540</td>
<td>1.01</td>
<td>(0.96, 1.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generalisability</td>
<td>5</td>
<td>540</td>
<td>1.22</td>
<td>(0.88, 1.70)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Evidence</td>
<td>4</td>
<td>317</td>
<td>1.03</td>
<td>(0.91, 1.17)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pooled Risk ratios and 99% CI

0.5 1 2

Figure 8: Differences in the completeness of reporting of CONSORT items between trials published before and after CONSORT publication

<table>
<thead>
<tr>
<th>CONSORT Checklist Item</th>
<th># of Evaluations</th>
<th># of RCTs</th>
<th>RR</th>
<th>99% CI</th>
<th>Favours Non-Endorsement</th>
<th>Favours Endorsement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title and Abstract</td>
<td>3</td>
<td>532</td>
<td>1.41</td>
<td>(0.63, 3.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>2</td>
<td>457</td>
<td>1.04</td>
<td>(1.00, 1.08)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>4</td>
<td>622</td>
<td>0.98</td>
<td>(0.88, 1.09)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interventions</td>
<td>4</td>
<td>630</td>
<td>1.02</td>
<td>(0.97, 1.07)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Objectives</td>
<td>2</td>
<td>517</td>
<td>1.04</td>
<td>(0.91, 1.16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcomes</td>
<td>5</td>
<td>716</td>
<td>1.43</td>
<td>(0.85, 2.40)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sample Size</td>
<td>6</td>
<td>983</td>
<td>1.30</td>
<td>(0.71, 2.37)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sequence Generation</td>
<td>8</td>
<td>1,085</td>
<td>1.46</td>
<td>(0.88, 2.44)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Allocation Concealment</td>
<td>6</td>
<td>855</td>
<td>1.23</td>
<td>(0.55, 2.75)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Implementation</td>
<td>2</td>
<td>517</td>
<td>1.94</td>
<td>(0.15, 24.7)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinding of Participants</td>
<td>1</td>
<td>75</td>
<td>0.77</td>
<td>(0.45, 1.31)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinding of Intervention</td>
<td>1</td>
<td>75</td>
<td>0.26</td>
<td>(0.09, 0.74)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinding of Outcome Assessor</td>
<td>1</td>
<td>75</td>
<td>0.66</td>
<td>(0.34, 1.30)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinding of Data Analyst</td>
<td>1</td>
<td>75</td>
<td>0.27</td>
<td>(0.02, 3.17)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Blinding Any Description</td>
<td>4</td>
<td>926</td>
<td>0.96</td>
<td>(0.61, 1.51)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical Methods</td>
<td>5</td>
<td>1,111</td>
<td>0.86</td>
<td>(0.62, 1.19)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Participant Flow</td>
<td>8</td>
<td>992</td>
<td>1.33</td>
<td>(0.95, 1.87)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recruitment</td>
<td>3</td>
<td>628</td>
<td>1.77</td>
<td>(0.48, 6.49)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline Data</td>
<td>2</td>
<td>529</td>
<td>1.42</td>
<td>(1.24, 1.62)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Numbers Analysed</td>
<td>6</td>
<td>1,005</td>
<td>1.72</td>
<td>(1.18, 2.50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Outcomes and Estimation</td>
<td>3</td>
<td>532</td>
<td>1.35</td>
<td>(0.73, 2.50)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ancillary Analyses</td>
<td>1</td>
<td>442</td>
<td>3.46</td>
<td>(2.47, 4.84)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adverse Events</td>
<td>3</td>
<td>507</td>
<td>1.39</td>
<td>(1.12, 1.73)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interpretation</td>
<td>1</td>
<td>442</td>
<td>1.01</td>
<td>(0.98, 1.04)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generalisability</td>
<td>1</td>
<td>442</td>
<td>1.77</td>
<td>(2.12, 1.48)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall Evidence</td>
<td>2</td>
<td>517</td>
<td>1.31</td>
<td>(0.99, 1.73)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Pooled Risk ratios and 99% CI

0.5 1 2

www.giving-evidence.com
5.5.6 Funding and staffing of CONSORT

CONSORT is not funded by the publishing industry as one might expect but instead by a rather surprising coalition: the (UK) Medical Research Council, Family Health International (a nonprofit human development organisation), the Ottawa Hospital Research Institute and the University of Oxford. The lead researcher, Professor Doug Altman, is a medical statistician at the University of Oxford: Altman has also led EQUATOR (section 5.5.9) and in 2015 received the British Medical Association’s lifetime achievement award for his work.

5.5.7 Reporting systematic reviews: QUOROM statement and its evolution into PRISMA

Systematic reviews and meta-analyses are essential tools for summarising evidence accurately. They help clinicians keep up to date; provide evidence for policy makers to judge risks, benefits, and harms of healthcare behaviours and interventions; gather together and summarise related research for patients and their carers; provide a starting point for clinical practice guideline developers; provide summaries of previous research for funders wishing to support new research; and help editors judge the merits of publishing reports of new studies. Recent data suggest that at least 2500 new systematic reviews reported in English are indexed in Medline annually.

Unfortunately, there is considerable evidence that key information is often poorly reported in systematic reviews, thus diminishing their potential usefulness. As for all research, systematic reviews should be reported fully and transparently to allow readers to assess the strengths and weaknesses of the investigation. That rationale led to the development of the QUOROM (Quality of Reporting of Meta-analysis) statement, a detailed reporting recommendation published in 1999.

The QUOROM statement, developed in 1996 and published in 1999, was conceived as guidance for authors reporting a meta-analysis of randomised trials. Since then, much has happened. First, knowledge about the conduct and reporting of systematic reviews has expanded considerably. For example, the Cochrane Library’s Methodology Register (which includes reports of studies relevant to the methods for systematic reviews) contained more than 11,000 entries in March 2009. Second, there have been many conceptual advances, such as “outcome-level” assessments of the risk of bias, that apply to systematic reviews. Third, authors have increasingly used systematic reviews to summarise evidence other than that provided by randomised trials.

However, despite advances, the quality of the conduct and reporting of systematic reviews remained well short of ideal. All of these issues prompted the need for an update and expansion of the QUOROM statement. Of note, recognising that the updated statement now addresses the above conceptual and methodological issues and may also have broader applicability than the original QUOROM statement, this changed the name of the reporting guidance to PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses).

5.5.8 Development of PRISMA

The PRISMA statement was developed by a group of 29 review authors, methodologists, clinicians, medical editors, and consumers. They attended a three day meeting in 2005 and participated in extensive post-meeting electronic correspondence. A consensus process that was informed by evidence, whenever possible, was used to develop a 27-item checklist. Items deemed essential for transparent reporting of a systematic review were included in the checklist. The flow diagram originally proposed by QUOROM was also modified to show numbers of identified records, excluded articles, and included studies. After 11 revisions the group approved the checklist (Appendix 3) and the flow diagram (Appendix 4).
5.5.9 Advancing the work of CONSORT to other areas of medical research - EQUATOR

EQUATOR (Enhancing the Quality and Transparency of health Research programme) grew out of the work of CONSORT and other guideline development groups. The project began in March 2006. Initially funded for one year by the UK NHS National Knowledge Service, the project had three major objectives: to map the current status of all activities aimed at preparing and disseminating guidelines on reporting health research studies, identify key individuals working in the area, and establish relationships with potential key stakeholders.

EQUATOR was formally established in 2008 as an international initiative that seeks to improve the reliability and value of published health research literature by promoting transparent and accurate reporting and wider use of robust reporting guidelines. It is the first coordinated attempt to tackle the problems of inadequate reporting systematically and on a global scale; it advances the work done by individual groups over the last 15 years. Table 4 gives the reporting guidelines for main study types, compiled by EQUATOR.

Table 4: Reporting guidelines for main study types

<table>
<thead>
<tr>
<th>Study Type</th>
<th>Guidelines</th>
</tr>
</thead>
<tbody>
<tr>
<td>Randomised trials</td>
<td>CONSORT</td>
</tr>
<tr>
<td>Observational studies</td>
<td>STROBE</td>
</tr>
<tr>
<td>Systematic reviews</td>
<td>PRISMA</td>
</tr>
<tr>
<td>Case reports</td>
<td>CARE</td>
</tr>
<tr>
<td>Qualitative research</td>
<td>SRQR</td>
</tr>
<tr>
<td>Diagnostic / prognostic studies</td>
<td>STARD</td>
</tr>
<tr>
<td>Quality improvement studies</td>
<td>SQUIRE</td>
</tr>
<tr>
<td>Economic evaluations</td>
<td>CHEERS</td>
</tr>
<tr>
<td>Animal pre-clinical studies</td>
<td>ARRIVE</td>
</tr>
<tr>
<td>Study protocols</td>
<td>SPIRIT</td>
</tr>
</tbody>
</table>

We now examine the transition of medicine to EBM, and the historical tradition and ‘building blocks’ on which it drew.
6 How did medicine come to be so evidence-based? Excerpts from the history of Evidence-Based Medicine (EBM)

Medical decision making has gone through a fundamental change in the last 50 years or so. Simply put, the foundation for decision making has shifted away from subjective judgments and reliance on authorities toward a formal analysis of evidence.

There is a long tradition of collecting medical statistics and some experimentation in medicine prior to 1960s. However, though evidence has been used in medicine since ancient times, the evidence was “unfiltered information”, and the practice in medicine was not systematically evidence-based.

Much of what we say about ‘evidence-based medicine’ can be said about ‘evidence-based practice’ in other disciplines, such as education. Table 5 below highlights some of the differences between the world of clinical medicine before and after the introduction of EBM. Our interest here is in what caused this change. EBM as we know it today is the result of several related but initially separate lines of research, and everyone who participated in the movement has his or her own story.

Using an arbitrary timeline of 1960 to divide our story (as a legal regulatory framework overseen by the US Food and Drug Administration (FDA) requiring proof of efficacy of new drugs was introduced in 1962), we have divided this story into the following parts:

- Use of evidence in medicine prior to the 1960s
- Development of the field of Clinical Epidemiology
- Use of evidence in medicine after 1960s
- Emergence of the field of clinical epidemiology
- Role of institutions in promoting EBM
- Indexing of medical literature
- Backlash against EBM
- Making it easy to access evidence
Table 5: Has evidence-based practiced changed medicine?  

<table>
<thead>
<tr>
<th>Before EBM</th>
<th>After EBM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source of knowledge is expert opinion</td>
<td>Source of knowledge is systematic review of evidence</td>
</tr>
<tr>
<td>Clinical skills are seen as semi-mystical</td>
<td>Clinical skills can be audited and managed</td>
</tr>
<tr>
<td>Research is marginal to practice</td>
<td>Research and practice go together</td>
</tr>
<tr>
<td>Analysis of research is haphazard</td>
<td>Analysis of research is systematic</td>
</tr>
<tr>
<td>Not important to gather new evidence from patients routinely</td>
<td>Patients should be included in trials wherever possible</td>
</tr>
<tr>
<td>Main information sources are experts, selected journals, and books</td>
<td>Essential to have immediate (electronic) access to systematically collected evidence</td>
</tr>
<tr>
<td>Most of what doctors need to know is in their heads</td>
<td>Doctors must use information tools constantly</td>
</tr>
<tr>
<td>Only lip service is paid to keeping up to date and learning new skills</td>
<td>Essential to keep learning new skills</td>
</tr>
<tr>
<td>Most medical care is assumed to be beneficial</td>
<td>Widespread recognition that the balance between doing good and harm is fine</td>
</tr>
<tr>
<td>Clinical performance is not systematically audited</td>
<td>Clinical performance is regularly reviewed and managed</td>
</tr>
<tr>
<td>Managers have little involvement in clinical processes</td>
<td>Managers are involved in clinical processes</td>
</tr>
<tr>
<td>Organisational model is hierarchical</td>
<td>Organisational model is much more democratic, based on ability to use evidence</td>
</tr>
<tr>
<td>Doctor patient relationship is essentially master / pupil</td>
<td>Patient partnership is the norm</td>
</tr>
<tr>
<td>Patients do not have easy access to the knowledge base of doctors</td>
<td>Patients have as much access to the evidence as doctors do</td>
</tr>
</tbody>
</table>
6.1 Use of Evidence in Medicine prior to 1960s

The study of anatomy began at least as early as 400 BC, with anatomists such as Hippocrates and Herophilus. Experiments in medicine started early: in 1753, James Lind published his landmark experiment into providing vitamin C to avoid scurvy. However, it took Lind 41 years to convince the British Royal Navy to implement his recommendation, after which the incidence of scurvy among the British sailors sharply declined.

Andrew Duncan, a Scottish doctor, launched a publication 20 years later summarising research for clinicians (which is like the EBM journal produced by BMJ (Section 6.4.2.1)). He recognised even then, that there was “lack of open access to information relevant to the wellbeing of patients”.

Semmelweis, a doctor working in the Vienna Maternity Hospital in 1846, observed that the rates of maternal mortality were much higher in the clinic attended by doctors rather than by midwives (98.4/1000 vs 36.2/1000), and that the doctors tended to come directly from postmortems without washing their hands. Semmelweis pointed this out – and was fired for his pains.

L'Institut Pasteur (The Pasteur Institute) is a nonprofit private research institution founded by Louis Pasteur in 1887 in Paris. Louis Pasteur served as the Institute’s first director until his death in 1895. He strongly supported the practice of EBM. The Institute’s research focuses on the study of infectious diseases, micro-organisms, viruses, and vaccines. Pasteur researched bacteria, contributed to the germ theory of disease (discussed in Section 7.2), and developed methods of sterilisation aimed at reducing pathogens in food, a process eventually called pasteurisation.

The smallpox vaccine was discovered in 1796 by the British physician Edward Jenner. Pasteur further developed the vaccination technique during the 19th century, extending its use to vaccinations against anthrax and rabies.
Box 1: Role of curiosity

Harvard surgeon Atul Gawande has many stories of the ‘extreme curiosity’ that has enabled many discoveries in medicine. In his 2014 Reith Lectures he recounted the tale of Werner Forssmann, a surgical intern in Eberswalde in Germany, who, in 1929, made an observation.

‘He was reading some medical journals - an obscure one actually, it had animal studies - and it had the picture of a horse where they’d threaded a tube up the leg of the horse all the way into the heart and then described what was going on from taking blood from there. And he said, “Well if we could do that to a horse, what if we did that to a human being?”’

And he went to his superiors, to his bosses, and said, “How about we take a tube and thread it into a human being’s heart?” And they said, “You’re crazy. You can’t do that. We know whenever you touch the heart, when people have attempted it in surgery, it goes into fibrillation and the patient dies. You cannot do this.” And he said, “Well what about in an animal?” “There’s no point and you’re just an intern anyway. Who says you should even deserve to get to ask these questions? Go back to work.”

He stole into the x-ray room, took a urinary catheter, made a slit in his own arm, threaded it up the vein and into his own heart and convinced a nurse to help him take a series of nine x-rays showing the tube inside his own heart.

He was fired.

And then in 1956, he was awarded the Nobel Prize with Andre Cournand who took his findings some 20 years later at Columbia University and then recognised that you could not only put the catheter into a person’s heart but shoot dye into the heart and that would let you take pictures and you could see the living heart and how it actually worked from the inside. What they’d done was they had founded the field of cardiology.’

Penicillin was discovered in 1928 by Sir Alexander Fleming, a Scottish biologist, pharmacologist and botanist and used extensively in World War II. Sir Austin Bradford Hill performed the first medical RCT in World War II (discussed in Section 5.4.2)
6.2 Use of Evidence in Medicine after 1960s

In 1960s and 1970s, a number of physicians were raising concerns about the lack of evidence and fears that patients were not receiving best treatments. They began to push for change in the traditional medical system and laid the foundation of clinical epidemiology and EBM as it is practiced today. A few of these stories are described below.

**Archie Cochrane** was an obstetrician. In 1935, as a lone medical student, he marched through London carrying a homemade placard that read, “All effective treatments must be free.” According to him, nobody noticed. In a seminal book “Effectiveness and Efficiency: Random Reflections on Health Services”, first published in 1972, he called for an international register of randomised controlled trials, and for explicit quality criteria for appraising published research, but neither goal was achieved in his lifetime. Today, the Cochrane Controlled Trials Register has more than 400,000 entries, and an international movement (Section 6.4.4) to improve the methodology of research synthesis also bears his name.

**Richard Smith**, former editor of the British Medical Journal (BMJ), writes of his disappointment in a medical education (in 1970 at the University of Edinburgh) which focused on facts, rather than the scientific method; and describes it as “anti-intellectual”.

**Sir Iain Chalmers**, a British health services researcher, one of the founders of the Cochrane Collaboration and coordinator of the James Lind Initiative told his story in *EBM: An Oral History*. He explained that in the 1960s, his medical education was geared towards enabling students to regurgitate answers for examinations, rather than towards becoming capable of independently assessing the evidence for a particular intervention. Working in Gaza as a young doctor, he followed rules such as never treating viral infections with antibiotics, only to discover afterwards that a number of trials had demonstrated that antibiotics in measles patients serve to reduce bacterial pneumonia. “In retrospect, I’m angry about that” he said, since suffering and deaths could have been avoided had the system been different.

**Brian Haynes**, a professor of clinical epidemiology and biostatistics at McMaster University in Canada, described his shock at asking a lecturer in 1969 about the evidence for Freud’s theories, only to be told that the lecturer was extremely sceptical that any evidence existed and that he had been ordered to give the lecture by the Freudian department chair. Archie Cochrane had experienced much the same when, in the 1930s, he had sought help from a Freudian psychologist for a condition himself. Haynes claimed that at that point he began to wonder about the evidence for the rest of his medical education - a concern that led him to pursue a doctorate at McMaster University with David Sackett, who turned out to be one of the founders of EBM (section 6.4).

**David Eddy**, physician, mathematician, and healthcare analyst, describes how he prepared a talk on physicians’ decision-making, expecting to find strong evidence with respect to his chosen topic of diagnostic mammography. However, he found “very few numbers, no formal rationale, and blatant errors in reasoning”.

Eddy then turned to ocular hypertension, reasoning that perhaps with more established interventions, more reliable evidence would exist. In fact he found that despite tens of millions of people receiving the treatment, only eight controlled trials had been conducted. Six of those eight trials demonstrated that patients became worse, rather than improved, with treatment. In trying to publish his findings, Eddy encountered considerable resistance from editors unwilling to publish about poor evidence,
decision-making or physician uncertainty. Eddy went on to work towards more formal use of evidence throughout his career, including his 20 years as chief scientist of the Medical Coverage Advisory Committee to the Blue Cross Blue Shield Association.

6.2.1 Introduction of key concepts
Along with these individuals pushing for change, a few key papers were published which talked about the following concepts, which though routine now, were not in common use 40 years ago.

6.2.1.1 Unwanted variation
Unwanted Variation refers to differences in treatment that do not arise from illness, or medical need. In 1967, while working in the Regional Medical Program (US) created with a $350,000 grant from President Lyndon B. Johnson, Dr. Jack Wennberg analysed Medicare data to determine how well hospitals and doctors were serving their communities. He found evidence of ‘unwanted variation’ in the Medicare population in diverse conditions such as treating back pain, surgical decisions and treatment for chronic conditions, such as use of beta blockers for individuals with congestive heart failure or lipid testing for those with diabetes.

6.2.1.2 Cost effectiveness
Cost effectiveness and QALYs (Section 5.2) were also introduced in the 1970s to address the issue that the resources available to meet the demands for health care are limited. Effectiveness was also pushed by Archie Cochrane in his book “Effectiveness and Efficiency: Random Reflections on Health Services” (although we think he did not use the term “cost effectiveness”).

6.2.1.3 Inappropriate interventions
In 1987, well conducted community studies in the Medicare population in the US showed that one in six cases of coronary angiography were being performed inappropriately, and gave high quality evidence for an increased professional effort to improve the appropriateness of this (and thereby other) procedure(s).

6.3 Emergence of the field of clinical epidemiology
Clinical epidemiology is a field which extends the principles of epidemiology (essentially, how disease is distributed in a population) to clinical practice. It incorporates the location, evaluation and application of the best evidence to patient care by clinicians as well as the generation of high quality research evidence by clinical researchers.

Clinical epidemiology became a formal course of study first at McMaster University's new medical school in 1967, headed by Dr David Sackett. Sackett, who had trained at Harvard, defined clinical epidemiology as “the application, by a physician who provides direct patient care, of epidemiological and biometric methods to the study of diagnostic and therapeutic process in order to effect an improvement in health” and that this was absolutely essential to improving medical care. The term ‘clinical epidemiology’ is sometimes used interchangeably with EBM.
6.4 Role of Institutions in promoting EBM

The key events in the history of EBM regarding organisations are shown on the timeline in Figure 9.

**Figure 9: Policy and academic milestones in the development of trials and the science of reviewing trials**

---

**A. Regulations & Organisations**

A legal regulatory framework overseen by the US FDA requiring proof of efficacy of new drugs was introduced in 1962, and other countries followed suit. These developments made it inevitable that randomised trials would increasingly become an important component of the evidence base.

Government health technology assessment agencies were also established because policymakers sought more reliable evidence of the effects of other forms of health care interventions. The Society for Medical Decision Making was founded in 1979 with a mission to improve health outcomes through advancing proactive systematic approaches to clinical decision making and policy-formation in health care by providing a scholarly forum that connects and educates researchers, providers, policy-makers, and the public.

Explicit insistence on evidence of effectiveness was introduced by the American Cancer Society in 1980. The U.S. Preventive Services Task Force began issuing guidelines for preventive interventions based on evidence-based principles in 1984. In 1985, the Blue Cross Blue Shield Association applied strict evidence-based criteria for covering new technologies. Beginning in 1987, specialist societies such as the American College of Physicians, and voluntary health organisations such as the American Heart Association, wrote many evidence-based guidelines. In 1993, the Cochrane Collaboration created a network of 13 countries to produce systematic reviews and guidelines. In 1997, the US Agency for Healthcare Research and Quality (then known as the Agency for Health Care Policy and Research, or AHCPR) established Evidence-based Practice Centers (EPCs) to produce evidence reports and technology assessments to support the development of guidelines.

---

**B. Publications**

---

www.giving-evidence.com
In the same year, a National Guideline Clearinghouse that followed the principles of evidence-based policies was created by Agency for Healthcare Research and Quality's (AHRQ), the American Medical Association (AMA), and the American Association of Health Plans (now America's Health Insurance Plans). In 1999, the National Institute for Clinical Excellence (NICE) was created in the UK and has acquired a high reputation internationally as a role model for developing clinical guidelines.

As the number of clinical trials grew, so too did the science of reviewing trials. Systematic reviews and meta-analyses endeavouring to make sense of multiple trials began to appear in a variety of health fields in the 1970s and 1980s (see Box 1). An important early example showed that postoperative radiotherapy after surgical treatment of breast cancer was associated with a previously unrecognised increased risk of death. Another challenged beliefs about vitamin C and the common cold. A third suggested a previously unrecognised advantage of some forms of fetal monitoring during labour in reducing neonatal seizures.

Box 2: Early systematic reviews of the effects of health care interventions

- Effects of ascorbic acid on the common cold (1975).
- Routine administration of iron and vitamins during pregnancy (1978).
- Randomized controlled trials of fetal monitoring (1979).
- Bipolar disorder, a state of the science report (1979).
- Aspirin after myocardial infarction (1980).
- Should every survivor of a heart attack be given a beta-blocker (1982).
- Effect of intravenous streptokinase on acute myocardial infarction (1982).
- Should mild hypertension be treated (1985).
- Beta blockade during and after myocardial infarction (1985).

By the mid-1980s, the need to minimise the likelihood of being misled by the effects of biases and the play of chance in reviews of research evidence was being made evident in articles and textbooks. In 1988, regularly updated electronic publication of systematic reviews and meta-analyses, along with bibliographies of randomised trials, began in the perinatal field. This provided a model for the international Cochrane Collaboration (launched in 1993) to prepare, maintain, and disseminate systematic reviews of the effects of health care interventions.
We now discuss a few notable institutions (and individuals) which had a major role in the development of EBM.

6.4.1 McMaster University (Canada)

McMaster University can be called the birthplace of the term “EBM”. At McMaster University in the 1970s, Clinical Epidemiology was re-invented as EBM through consolidation of principles and by structuring methods, as summarised in a series of articles in 1983. The focus was to reform Tradition Based Medicine (TBM), one doctor at a time through critical appraisal of literature (“evidence”) and by evaluating personal experience.

In 1985 a member of the McMaster group, David Sackett, published an influential book “Clinical Epidemiology: A basic science for clinical medicine” that described tools to help doctors use research from a study population to inform their care of an individual patient. The McMaster group went on to coin the term “evidence based medicine” and established the EBM Working Group in 1992 that published a series of articles (subsequently compiled in a manual) for doctors in the Journal of the American Medical Association (JAMA).

While the idea was not new – and had been championed by, for instance, Archie Cochrane and David Eddy even in the early 1970s – it was the McMaster group which made the important step of turning the research methods into a toolkit: this allowed them to be practically applied on a widespread basis.

6.4.2 Centre for EBM (UK)

In 1994 Sir Muir Gray, who was then director of NHS Research and Development (R & D) for Anglia and Oxford Regional Health Authority, set about developing an EBM movement in the UK. Using a very small start-up grant from the R & D scheme he initiated several centres around the country. By 1995 the Centre for Evidence Based Medicine (CEBM) was established in Oxford University headed by David Sackett. The initial centre consisted of two rooms in the John Radcliffe Hospital, Oxford. The broad aim was to promote evidence-based health care and provide support and resources to anyone who wants to make use of them. The Centre started building a network of tutors locally and across the world that shared the philosophy of CEBM. Its course ‘Teaching EBM’ runs to this day.

The initial remit of the centre was:

1. To support the teaching and practice of evidence based health care throughout the UK and Europe (see Evidence Cart in Box 3 below).

2. To create formal graduate education in the conduct of randomised controlled trials and systematic reviews at the University of Oxford.

3. To conduct applied, patient-based and methodological research to generate the new knowledge required for the practice of evidence-based health care.
Box 3: Evidence cart, taken on hospital rounds and its use for clinical care of patients (developed by David Sackett)

CEBM emphasised that it needed to use evidence about five times for every admitted patient and twice for every out-patient. It was essential to use evidence at the bedside and for this it needed to be accessed it within seconds.

As a result the evidence cart was born (pictured here). The contents of the cart included:
- An infra-red simultaneous stethoscope with 12 remote receivers to allow junior doctors to share the experience of listening to, for example, an unusual heart sound;
- Textbooks on making physical diagnosis and reprints of the evidence based (JAMA Rational Clinical Exam) series;
- A laptop to search PubMed, a notebook computer and projector and pop-out screen to display the results of the evidence searches; and
- A rapid printer for the junior doctors and the medical students, and at times the patients, to take information away.

In the month before the cart, an audit of 72 clinical cases that needed a search for evidence of their optimal management revealed only 19 (26 percent) of searches were actually carried out. After its arrival, the evidence cart was used 98 times over a month to search for and use evidence – alongside clinical judgement – to try and improve patient care. An audit of these searches revealed that 81 percent were for evidence that could affect diagnostic and/or treatment decisions and, more impressively, 90 percent of these searches were successful in finding evidence that was used, in a timely manner, to make decisions with the patient about their care. Further details of the audit revealed that, of these searches, 52 percent confirmed diagnostic and/or management decisions, 23 percent led to changes in decisions and 25 percent led to additional decisions. When the cart was removed from the clinical wards, an audit revealed that of 41 cases where a search for evidence should have been carried out, only 5 (12 percent) were.

A 1995 Lancet paper analysed the treatments given to all 109 emergency admission patients managed during a one month period. At the time the consultant teams were split into 3 teams: firm A, B and C. Firm A (headed by David Sackett) had ten shifts of 12 hours per month seeing on average 25 to 30 patients per take. Of the 109 treatments, 82 percent were classified as evidence-based. Firm A noted that the majority of patients were offered (and accepted) evidence-based interventions. They also noted that over half of those patients were receiving treatments where the evidence had shown more harm than benefit.

Other centres were then set up by other universities, e.g., in Austria and Australia. Their ability to rapidly influence practice partly relies on the proximity of research, teaching and practice, which is unusual to medicine.
6.4.2.1 Summarising relevant research to change practice - The Journal of Evidence-Based Medicine

In 1995 the EBM journal was launched by CEBM and quickly became a huge success. It was based on searching the top journals for articles that were most likely to change practice and then summarising their content. The EBM journal systematically searches a wide range of international medical journals applying strict criteria for the validity of research. Experts critically appraise the most clinically relevant articles and summarise them including commentary on their clinical applicability. Evidence-Based Medicine also publishes articles relevant to the study and practice of evidence-based medicine. The EBM Journal is owned by BMJ and was edited by Paul Glasziou for 10 years. Glasziou was then director of the CEBM and now leads work on research waste and clarity of reporting in Bond University, Australia.

Along with the ACP journal it was right for its time, as it tried to overcome the barriers of the ever increasing numbers of RCT published each year.

6.4.3 Agency for Health Care Policy and Research (USA)

The Agency for Healthcare Research and Quality (AHCPR) is one of 12 agencies within the US Department of Health and Human Services. The Agency was established as a constituent unit of the Public Health Service (PHS) under the Omnibus Budget Reconciliation Act of 1989 to enhance the quality, appropriateness, and effectiveness of health care services and access to care by conducting and supporting research, demonstration projects, and evaluations; developing guidelines; and disseminating information on health care services and delivery systems.

Establishing legislative framework for EBM

One of the first moves by AHCPR was to build the outcomes / effectiveness agenda into an amendment to the PHS Act that would elevate AHCPR’s activities and personnel into a new Public Health Service (PHS) agency alongside National Institute of Health (NIH), the Food and Drug Administration (FDA), the Centers for Disease Control and Prevention (CDC), and the Health Resources and Services Administration (HRSA).

Several pieces of strategic thinking by AHCPR leadership were incorporated in the legislation:

- First, the new agency was to address practical concerns that already had a constituency. The agency (and HSR) thus became identified with the quality and cost concerns for which a constituency had been built.
- Second, the legislation incorporated bureaucratic strategy to improve the lot of health service research—its movement into an NIH-level PHS agency and the incorporation of two new sources of funding.
- Third, the legislation reflected marketing strategy, with the abandonment of the phrase “health services research” and the suggestion that the agency did more than research.

The legislation had several important effects:

- First, the movement of health services research from NCHSR into the new agency did enhance its prominence and visibility.
- Second, the legislation substantially changed HSR’s budgetary fortunes. The funding moved to a new plane—from $53 million in 1989 to $97 million in 1990 - a remarkable change in view of the field’s history.
- Third was the creation of a new programme for developing practice guidelines and an enhanced and focused programme of outcomes research.
6.4.4 Cochrane Collaboration (worldwide)

The Cochrane Collaboration was founded in 1993 under the leadership of Iain Chalmers. It was developed in response to Archie Cochrane’s call for up-to-date, systematic reviews of all relevant RCTs of health care.\textsuperscript{189}

Cochrane’s suggestion that the methods used to prepare and maintain reviews of controlled trials in pregnancy and childbirth (his field) be applied more widely was taken up by the Research and Development Programme, initiated to support the UK’s National Health Service. Through the NHS R&D programme, funds were provided to establish a “Cochrane Centre”, to collaborate with others in the UK and elsewhere to facilitate systematic reviews of RCTs across all areas of health care.\textsuperscript{190}

The collaboration was formed to organise medical research information in a systematic way to facilitate the choices that health professionals, patients, policy-makers and others face in health interventions according to the principles of evidence-based medicine.\textsuperscript{191}

The collaboration formed an official relationship in January 2011 with the World Health Organization (WHO) as a partner non-governmental organisation with a seat on the World Health Assembly to provide input into WHO resolution.

The collaboration has centres in many countries and regions, e.g., the South Asian centre is based in Vellore, Tamil Nadu.

6.4.5 National Institute for Health and Care Excellence (NICE) (UK)

Since the early 1990s, evidence-based health care has become a cornerstone in UK policies to identify effective, and in particular cost-effective, practices and to move away from decisions based on opinion or current practice to a greater use of scientific research and evidence.\textsuperscript{192} A key step came with the introduction of the National Institute of Health and Clinical Excellence (NICE) in 1999. NICE was initially established in England and Wales to help the NHS meet three continuing objectives: (i) to improve continually the overall standards of care; (ii) to reduce unacceptable variation in clinical practice; and (iii) to ensure the best use of resources so that patients receive the greatest benefit.\textsuperscript{193}

In 2005, after merging with the Health Development Agency, it began developing public health guidance to help prevent ill health and promote healthier lifestyles.

In April 2013 it took on responsibility for developing guidance and quality standards in social care. To reflect the role of social care, the name was changed to National Institute for Health and Care Excellence, while the initials remained the same (NICE).

The way NICE was established in legislation means that the guidance is officially England only. However, it has agreements to provide certain products and services to Wales, Scotland and Northern Ireland. Decisions on how the guidance applies in these countries are made by the devolved administrations, who are often involved and consulted with developing NICE guidance.

NICE’s role is to improve outcomes for people using the NHS and other public health and social care services by:

- **Producing** evidence-based guidance and advice for health, public health and social care practitioners.
- **Developing** quality standards and performance metrics for those providing and commissioning health, public health and social care services.
- Providing a range of informational services for commissioners, practitioners and managers across the spectrum of health and social care.

**NICE guidance framework**

**NICE Guidelines:** make evidence-based recommendations on a wide range of topics, from preventing and managing specific conditions, improving health and managing medicines in different settings, to providing social care to adults and children, and planning broader services and interventions to improve the health of communities. These aim to promote integrated care where appropriate, for example, by covering transitions between childrens and adult services and between health and social care. NICE guidelines use cost (£) per QALY as its metric (discussed more in section 5.2.3)

**Technological appraisals:** assess the clinical and cost effectiveness of health technologies, such as new pharmaceutical and biopharmaceutical products, but also include procedures, devices and diagnostic agents. This is to ensure that all NHS patients have equitable access to the most clinically- and cost-effective treatments that are viable.

**NICE International:** helps to raise standards of healthcare around the world by providing advice and support to encourage the use of clinically- and cost-effective treatments. NICE International - which operates on a strict not-for-profit, fee-for-service basis - also carries out research such as generating case studies, preparing tools to help data analysis and encouraging shared learning through international meetings.

### 6.5 Repositories and indexing of medical literature

The rapid expansion of medical information was a key feature of the development of EBM. Prior to the widespread development of electronic databases it was not possible to gain rapid access to research evidence to inform practice. The precursor of the US National Library of Medicine (NLM) began indexing the medical literature in 1865. Between 1865 and 2006, the index grew from 1,600 references to nearly 10 million. Even with the assistance of electronic databases such as NLM's MEDLINE, the problem of having to trawl through and sift vast amounts of data has grown.

#### 6.5.1 Role of National Library of Medicine in EBM

The world's largest biomedical library, the US National Library of Medicine (NLM), on the campus of the National Institutes of Health in Bethesda, Maryland, has been a centre of information and innovation since its founding in 1836. NLM maintains and makes available a vast print collection and produces electronic information resources on a wide range of topics that are searched billions of times each year by millions of people around the globe. It also supports and conducts research, development, and training in biomedical informatics and health information technology. In addition, the library coordinates a 6,000-member National Network of Libraries of Medicine that promotes and provides access to health information across the US. NLM's information services and research programmes serve the world by supporting scientific discovery, clinical research, education, health care delivery, public health response and the empowering of people to improve their personal health.
6.5.2 Milestones in the creation of repositories and indexing of medical research

Some of the key milestones in evidence generation and research are given in Table 6.

Table 6: Milestones in evidence generation and research:

<table>
<thead>
<tr>
<th>Date</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>1879</td>
<td>First volume of the Index Medicus is published, establishing the Library’s groundbreaking role in systematic indexing of medical journal articles.</td>
</tr>
<tr>
<td>1960</td>
<td>1st edition of the Medical Subject Headings (MeSH) released for use in both indexing and cataloguing. Likely the first major thesaurus developed for use in an automated bibliographic system, MeSH is still used in data creation and information retrieval in PubMed, the NLM Catalog, and other NLM databases.</td>
</tr>
<tr>
<td>1964</td>
<td>The Medical Literature Analysis and Retrieval System (MEDLARS), developed under contract by the General Electric Company, became operational at NLM. Over time, this pioneering computerised bibliographic system would produce many printed indexes, catalogues, and subject bibliographies for NLM and partner organisations, including the American Dental Association, the American Nurses Association, the American Hospital Association, and the World Health Organization. It also provided the first large-scale, computer-based retrospective search service available to the public. NLM quickly made data from the system available on magnetic tapes to selected institutions to establish decentralised search services across the country.</td>
</tr>
<tr>
<td>1965</td>
<td>The Medical Library Assistance Act of 1965 signed into law, authorising NLM's extramural programmes of grant assistance to improve that nation’s medical library and health communication resources, including the establishment of the Regional Medical Library Network (now called the National Network of Libraries of Medicine).</td>
</tr>
<tr>
<td>1970</td>
<td>Lister Hill Centre begins range of networking experiments: remote online retrieval via commercial telecommunications network; use of satellite communications (in cooperation with NASA) for medical consultation between remote Alaskan areas and academic medical centres; video networking for medical education in rural areas.</td>
</tr>
<tr>
<td>1971</td>
<td>MEDLINE (MEDLARS Online) began national operation, following an R&amp;D project directed by NLM’s Lister Hill Centre. MEDLINE, which evolved from the MEDLARS system which had been installed in 1964, was the first successful collaboration of a large reference database with a nationwide commercial telecommunications network.</td>
</tr>
<tr>
<td>1972</td>
<td>The Library joins with the Bureau of Health Manpower Education to support training of physicians and other health scientists in the use of computer technology for medical education and the provision of health care. This was the beginning of NLM-funded informatics training, which now includes 18 informatics research training programmes in academic centres, fellowships at NLM, and highly regarded short courses at Woods Hole.</td>
</tr>
<tr>
<td>1978</td>
<td>Hazardous Substances Databank released, with extensive data about the properties, toxic, and environmental effects of hazardous substances.</td>
</tr>
<tr>
<td>1985</td>
<td>The DOCLINE document request system implemented for free use by health sciences libraries in the National Network of Libraries of Medicine. Key features included the ability to import journal article reference data from MEDLINE so it did not have to be rekeyed and automatic routing of requests to appropriate libraries, based on an algorithm that consulted a large database of summary holdings data for US health sciences libraries and tables of information about preferred borrowing patterns.</td>
</tr>
<tr>
<td>1989</td>
<td>Congress established the Agency for Health Care Policy and Research (AHCPR) and instructed NLM to work with the new agency to improve access to health services research and health care technology assessment information.</td>
</tr>
<tr>
<td>Date</td>
<td>Details</td>
</tr>
<tr>
<td>------</td>
<td>---------</td>
</tr>
<tr>
<td>1993</td>
<td>National Information Center on Health Services Research and Health Care Technology (NICHSR) established at NLM. (Public Law 103-43) to improve “the collection, storage, analysis, retrieval, and dissemination of information on health services research, clinical practice guidelines, and on health care technology, including the assessment of such technology.</td>
</tr>
<tr>
<td>1998</td>
<td>NLM began a telecommunications programme for the Multilateral Initiative on Malaria, heading an international effort to provide malaria researchers in Africa with full access to the Internet.</td>
</tr>
</tbody>
</table>

### 6.5.3 National Network of Libraries of Medicine (NN / LM)

NN / LM is a nationwide partnership of health sciences libraries, information centres, public libraries and community-based resource centres. Administered by the National Library of Medicine, the network consists of eight regional medical libraries, 160 resource libraries and over 5,000 members. Together, network libraries strive to improve the health of all American citizens.

Though the internet helps to share reliable health information with people around the world through Medline Plus, the network’s partner libraries give NLM a vast reach, from cities to rural areas and even across oceans to the users in Alaska, Hawaii and the territories. The members help spread the word that good information is available, and there’s a ripple effect as that word goes out to consumers and health professionals.

NN / LM Network members work together to make sure that the most accurate biomedical information is available to both health professionals and the public.

The Regional Medical Libraries coordinate a variety of services for the network’s eight geographic regions. Working with the members of the network as well as community-based organisations, faith-based organisations, tribal colleges, public health agencies and other institutions, they promote good health information and the many online resources of the National Library of Medicine, such as Medline Plus. Free RML-sponsored Medline Plus workshops and presentations introduce health professionals and the public to a wealth of reliable and timely biomedical information in English and Spanish.
### 6.6 Backlash against EBM

The backlash against EBM was formidable and aggressive. Criticism has ranged from EBM being old hat, impossible to practise, ‘cookbook medicine’, the creature of managers and purchasers, too concerned only with randomised trials. Some accused it - and still do - of being a dangerous innovation to curtail clinical freedom and minimise health care costs. A 1995 Lancet editorial suggested that EBM go back “to its place”, a suggestion that David Sackett found particularly wounding given its racial implications and his own involvement in the civil rights movement. (In the US, the statement that African Americans should be kept “in their place” so as not to challenge existing political structures was often heard in response to demands for increased civil rights).

Sackett and his supporters chose to respond in an editorial which confronted each of the perceptions about EBM. The response is summarised below:

<table>
<thead>
<tr>
<th>Accusation</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td>EBM is old-hat and impossible to practice</td>
<td>EBM is possible to practise, and clinicians are already doing it</td>
</tr>
<tr>
<td>EBM can be conducted only from ivory towers and armchairs</td>
<td>EBM is being carried out by front-line teams in general medicine, psychiatry and surgery: “busy clinicians who devote their scarce reading time to selective, efficient, patient-driven searching, appraisal and incorporation of the best available evidence can practice evidence-based medicine”</td>
</tr>
<tr>
<td>EBM is “cook-book” medicine which treats every clinical case identically</td>
<td>EBM is not “cook-book” medicine; external evidence informs, but does not replace, individual clinical expertise</td>
</tr>
<tr>
<td>EBM is cost-cutting medicine</td>
<td>EBM is not cost-cutting medicine, because practitioners will find the best interventions to maximise benefits for patients, which may even raise costs</td>
</tr>
<tr>
<td>EBM is restricted to randomised trials and meta-analyses</td>
<td>These are only the starting points; it involves tracking down the best external evidence with which to answer clinical questions</td>
</tr>
</tbody>
</table>

However, clearly the 1996 Sackett editorial failed to assuage all concerns about the evidence-based movement. A 1997 journal article minced no words in stating that:

“Clinicians need not be bashful about putting ‘evidence-based’ mania in its place…Overweening and unjustifiable ambition should be opposed; just as we ought to impose the imposition of any system of rationalistic dictatorship based upon simplistic and incomplete analysis…I see EBM, in its present form, as a dangerous delusion: erroneous in both rationale and conclusions, and a potentially lethal weapon in the hands of misguided regulators and reformers.”

A 1997 BMJ article entitled “Evidence-based medicine: Socratic dissent” parodies a dialogue between Socrates and the over-enthusiastic practitioner of EBM. This concludes with Socrates warning Enthusiasticus that politicians and managers are not concerned about the reliability of evidence. Instead, he suggests that these groups rather see “your beloved evidence-based medicine as a means to shackle the doctors and bend them to their will…Beware, Enthusiasticus, that you are not used as a dupe in a political game of health economics. Remember, hemlock may be down the line.”
How did such a movement manage to gain footing against a backlash “among middle-level guys who were used to making pronouncements”, as Sackett put it? In *EBM: an Oral History*, Sackett attributes the surge in popularity of EBM to it being:

(1) Supported by senior clinicians who were secure in their practice and happy to be challenged, and

(2) Empowering young doctors—and subsequently nurses and other clinicians because it didn’t rely on authority or tenure.

Essentially, then, EBM was supported by two discrete groups, and hence the opponents of EBM were challenged on two fronts. The notion of empowerment and of basing treatment decisions on evidence, rather than authority, caught on.

Another perspective would suggest that the dawning of EBM is not, per se, a revolution but a result from other drivers of change such as increased emphasis on patient autonomy and access to information; the increasing availability of the internet in the 1990s for medical students and then patients for access to medical information and evidence, both for practitioners and patients; desires from a number of stakeholders to better manage the clinical process; and the cost constraints on health care that were emerging in the 1970s and 80s.

### 6.7 Making it easy to access evidence

A lot of effort has gone into making reliable, high quality evidence easy to access for the busy practitioner. A few reliable, reputed easy-to-use sources of medical evidence are listed in Box 4.

**Box 4: Sources of easy-to-use digests of medical research**

- The National Institute for Health and Care Excellence (NICE) [https://www.nice.org.uk/](https://www.nice.org.uk/) (see section 6.4.5)

- The Scottish Intercollegiate Guidelines Network (SIGN) [http://sign.ac.uk/index.html](http://sign.ac.uk/index.html)

  was established in 1993 and develops evidence based clinical practice guidelines for the National Health Service (NHS) in Scotland. SIGN guidelines are derived from a systematic review of the scientific literature and are designed as a vehicle for accelerating the translation of new knowledge into action to meet our aim of reducing variations in practice, and improving patient-important outcome

- The Oxford Centre for Evidence-Based Medicine [http://www.cebm.net/](http://www.cebm.net/) (see section 6.4.2)

- The Centre for Reviews and Dissemination ([https://www.york.ac.uk/crd/](https://www.york.ac.uk/crd/)), based at York University, is funded by the National Institute for Health Research (NIHR), England; the Department of Health, Public Health Agency, Northern Ireland; NICE and Welsh Assembly Government. It provides research-based information about the effects of health and social care interventions via databases and undertakes systematic reviews evaluating the research evidence on health and public health questions of national and international importance. This includes managing four databases of research evidence in health and social care as a means of informing decision making and encouraging uptake of research.

- The Cochrane Collaboration ([www.cochrane.org](http://www.cochrane.org)) (see section 6.4.4)
6.8 Ongoing problems in EBM

Many now argue that despite its successes, EBM is facing a serious crisis and some of the major problems are summarised in Box 5.

**Box 5: Major problems in EBM today**

- The volume of evidence, especially clinical guidelines, has become unmanageable
- Statistically significant benefits may be marginal in clinical practice
- Inflexible rules and technology driven prompts may produce care that is management driven rather than patient centred
- Waste in the research process, perhaps £100bn every year
- Evidence based guidelines often map poorly to patients who have more than one problem e.g., diabetes and hypertension, or kidney problem and asthma etc

**Publication bias** is a major problem. This arises because not all evidence which is produced is published, and the published material seems skewed. For example (and as discussed in the first Getting Better document), clinical trials published by pharmaceutical companies are four times as likely to be favourable than the same trials conducted independently. As Richard Smith, former editor of the BMJ, points out in *EBM: an Oral History*, we don’t know what is happening to these unreported studies. Is there a conscious decision not to publish? Are there missing outcomes? Are investigators losing interest after trials? The fact that we know so little about the failure to report raises the questions of what is the evidence and can we rely on what we do know about.

**Various problems with EBM** were outlined in a 2014 BMJ article which argued for “better understanding of how clinicians and patients find, interpret and evaluate evidence from research studies, and how (and if) these processes feed into clinical communications, exploration of diagnostic options, and shared decision-making”. The problems cited include:

- “Distortion of the evidence-based brand”: through drug and research industries setting the research agenda; manipulation of trials; and factoring of political conviction in development of “evidence-based” policies
- “Too much evidence”: sheer volume renders decision-making difficult
- “Marginal gains and a shift from disease to risk”: As interventions that promise large improvements were implemented first, research questions today focus more on marginal gains
- “Overemphasis on following algorithmic rules”: to the detriment of local, individualised and patient-initiated elements of the clinical consultation
- “Poor fit for multi-morbidity”: with prevalence of chronic degenerative diseases, few patients have one condition that maps to one guideline.
Waste in the research process is huge, estimated at around £100bn every year. It arises from four main factors:

- research which asks questions which don’t matter (either because nobody cares about the answer or because the answer is already known)
- research which is too poor quality to answer the question properly
- research which is never published at all (or at least, nowhere that people can find)
- research reports which are too unclear to show what the intervention actually was and/or has too few details of the research method to assess the research quality.

Most, if not all, of these problems afflict impact research by charities and the social sector, almost certainly including education in LEDCs.

Finally, another important issue relates to translatability of evidence-based practice to public health. As a simple example, it is easier to establish whether a particular medication will improve health in a single patient than it is to assess whether reducing added sugar in processed foods will reduce obesity in a population. The factors affecting the success of the intervention are more complex, the timescales are longer, and having a control group may be impossible; nonetheless, evidence should play an important part in determining appropriate interventions at this level. One BMJ article noted:

“At although we have focused on individual clinical care, there is also an important evidence base relating to population level interventions aimed at improving public health (such as pricing and labelling of consumables, fluoridation of water, and sex education). These are often complex, multifaceted programmes with important ethical and practical dimensions, but the same principles apply as in clinical care. Success of interventions depends on local feasibility, acceptability, and fit with context – and hence on informed, shared decision making with and by local communities, using summaries and visualisations of population level metrics.”

How did medicine come to be so evidence-based?
7 Case study of discovery and deployment of developments: Oral rehydration therapy

The case study below describes research over more than 110 years, which lead to the discovery, understanding of biological mechanism, and subsequent mass adoption of Oral Rehydration Therapy (ORT; also known as oral rehydration solution, ORS). Oral Rehydration Therapy has been referred to as the “magic bullet” in the treatment of diarrhoea.

7.1 Cholera epidemics in the 19th century

Epidemic Cholera caused profuse and violent cramps, vomiting and diarrhoea, with dehydration so rapid and severe that the blood thickens and the skin becomes deathlike and blue. Cholera victims can die in a matter of hours. Because 19th-century transformations in industrial, urban, political, and cultural life were intimately connected with discussions of proper public health practices and causes of disease, attempts to explain epidemic Cholera involved every part of society.

7.2 Causes of Cholera

7.2.1 Miasma theory

For much of the 19th century, most European and American physicians believed that Cholera was a locally produced miasmatic disease. The miasma theory (also called the miasmatic theory) held that diseases such as Cholera, chlamydia or the Black Death were caused by a miasma (Μίασμα, ancient Greek: “pollution”), a noxious form of “bad air”, also known as “night air”. The theory held that the origin of epidemics was a miasma emanating from rotting organic matter. The miasma theory was accepted from ancient times in Europe, India, and China.

7.2.1.1 From miasma to germ theory

The miasmatic theory was challenged by John Snow, suggesting that there was some means by which the disease was spread via a poison or morbid material in the water. He suggested this before and in response to an epidemic on Broad Street in central London in 1854 (Figure 10).
7.2.2 Discovery of Cholera organism

Cholera came to Florence in 1854 during the Asiatic Cholera Pandemic of 1846-63, and Filippo Pacini, an Italian scientist became very interested in the disease. Immediately following the death of Cholera patients, he performed an autopsy and with his microscope, conducted histological examinations of the intestinal mucosa (Figure 11). During such studies, Pacini first discovered a comma-shaped organism. He published a paper in 1854 entitled, “Microscopical observations and pathological deductions on Cholera” in which he described the organism and its relation to the disease.

7.2.3 The epidemics subside

By the end of the 19th century, Cholera epidemics no longer appeared in Europe and North America. The reasons for this are uncertain, but standards of living had risen and many communities had made major changes in sanitation and established permanent boards of health. Medical thought had changed in many ways as well: whereas in 1831, most physicians believed Cholera to be a nonspecific, non-contiguous miasmatic condition that favoured the morally and physically predisposed, by the end of the 19th century, Cholera was primarily understood to be a specific contagious disease caused by a particular microscopic organism.
7.2.4 Ignoring evidence about ORS
Several doctors published studies in the 1950s that claimed that oral solutions containing a wide range of ingredients—carob flour, bananas, and even carrot soup—had benefited Cholera patients, but these did not explain their results in terms of known physiological mechanisms. Therefore until the 1960s, the standard treatment for Cholera consisted of starvation, along with intravenous solutions administered in hospital settings under the close supervision of health professionals. Such treatment was essentially unavailable to the millions of people who were affected by seasonal epidemics of Cholera every year.

7.2.5 Role of institutions in developing ORS
International research centres in Dacca, East Pakistan (now Dhaka, Bangladesh), and Calcutta (now Kolkata), India, sites of frequent Cholera epidemics, worked intensively beginning in the early 1960s on a Cholera vaccine. While the two research centres have since vied for credit, especially in the area of oral rehydration therapy, unlike the stereotypical cutthroat competition of today, the two laboratories freely shared their results.

Improvements in the understanding of the fluid electrolyte balance during diarrhoea, and a trial and learning approach by the scientists in the two research centres, led to the recognition of the potential importance of oral rehydration solution as more than just a supplement to hospital-administered IV fluids dawned slowly upon a research community focused on developing a vaccine for Cholera, not on treatment.

The researchers realised that the intravenous solution needed to be extended to the oral solution: the patient needed to drink enough Oral Rehydration Solution (ORS) to exactly equal the volume of fluid being lost. In 1968, young doctors were allowed to implement a large-scale, closely monitored, and ultimately successful clinical field trial with an all-oral rehydration solution. Seasonal epidemics of Cholera in Asia and in emergency settings in other regions of the world meant that the possibility of an inexpensive treatment administered by nonmedical personnel was of great interest to international organisations such as WHO and UNICEF.

7.2.6 Adoption of ORS
7.2.6.1 Role of institutions in promoting ORS
In 1994, a year after the 25th anniversary of the discovery of ORS, four institutions were acknowledged for promoting ORS worldwide:

1. BRAC (formerly Bangladesh Rural Advancement Committee), an international development organisation based in Bangladesh
2. UNICEF
3. UNDP (United Nations Development Programme), and
4. USAID.

In this case study, we mainly explore the role played by BRAC.

7.2.6.2 Role of BRAC
After the discovery of ORS, WHO did not encourage home-available non-standardised rehydration solution, especially if it was formulated with just sugar and salt, without other electrolytes such as potassium and bicarbonate.
However from its experience in refugee camps, BRAC concluded that family members could dispense ORS and that too many lives were at stake not to attempt a stopgap solution despite its imperfections. BRAC developed the “Lobon-Gur Solution”, and designed a roll out suitable to the rural context.

Stage 1 Brain-storming before field trial
BRAC had to ensure exact proportions of salt, sugar, and water in ORS in rural households - not just to be effective, but also to avoid potentially fatal concentrations of salt. By trial and error, BRAC scientists came up with the combination of salt (lobon) one pinch; and unrefined sugar (gur) one fistful; in half seer (unit of liquid used in cooking; about a litre) of water to develop a consistent, effective lobon-gur solution. This solution was based on research which had shown that most rural Bangladeshi women could estimate a seer of water accurately.

Developing the message: BRAC’s deliverable was an educational message covering
- how to make the lobon-gur solution,
- when to use it,
- how to manage childrens nutrition afterward, and
- how to prevent diarrhoea in the future.

Delivery and dissemination: Since mothers, daughters, and mothers-in-law provided most of the child care in Bangladesh, these women were identified as the best ‘vector’ for delivering ORS and disseminating the message. In the 1970s, radio was not widely available in Bangladesh, and gathering women at a central place was not practical in rural areas as women were discouraged from appearing in public. BRAC decided that women would have to be trained individually, in their homes, face-to-face. This training would have to be delivered by women, who had a good reputation in the community.

Monitoring: As the margin of error between an effective solution and one dangerously high in salt was narrow, a system for monitoring how well the target group could make and use the solution had to be devised.

Stage 2 Field trial
BRAC conducted its first three-month field experiment in two villages sufficiently close to BRAC’s local office so that the programme supervisor and two female oral rehydration workers could meet every day to share experiences and work to improve the protocol. The results (successes and failures) were reported not just to donors but to the international scientific community. The field trials showed that:

- Illiterate women could learn to make an effective lobon-gur solution
- Young women with 8 to 10 years of education could be effective trainers for making and administering lobon-gur solution
- A cost-effective monitoring system could determine how much mothers absorbed the lobon-gur solution lesson and how the training might be improved
- Necessary management and logistics—such as shipping sample solutions to the reference centre—were feasible
- The field experiment produced a refined message - “10 Points to Remember” - which, properly delivered by a trained Rehydration Worker, could communicate the definition of diarrhoea, the way to make and use lobon-gur solution to treat it, and some nutritional advice, all in about 20 minutes.
Stage 3 Large scale piloting

BRAC began a larger pilot project, the Oral Therapy Program (1979–80, OTP). Oral Rehydration Workers (ORWs) recruited and trained locally resident women. To keep the programme focused on effective learning, not just on the number of women trained, BRAC introduced an incentive system for the ORW. A monitor would visit 10 percent of the women taught in the last month. The ORW who trained them would receive a bonus for each woman who could make an effective lobon-gur solution, and then additional bonuses based on the number of messages the woman could recall. The incentive system motivated a new round of improvements in teaching methods. Before the incentives, for example, the ORW simply demonstrated how to make the lobon-gur solution; now they insist that the mothers make a batch themselves, in front of the ORWs.

The number and content of “Points to Remember” was revised several times during OTP. Trial and error showed that 10 messages were too many for many mothers to remember, so several were combined. As a result, the main training tool eventually became “Seven Points to Remember.”

National Roll Out - Bangladesh National Diarrhoea Control Programmes

After further successful larger trials by BRAC, Bangladesh became the first country in the world to accumulate large-scale experience using ORS. In 1981, the government created the National Oral Rehydration Project and distributed packets of ORS to health centres in 100/509 sub-districts. Between 1980 and 1990, BRAC trained 12 million women (approximately half of all women in the country) to prepare and use sugar–salt solution, and still trains community health workers.

Starting in 1985, Population Services International (PSI) and, later, the Social Marketing Company (SMC) promoted branded ORS through multi-channel social marketing, spending US$ 1 million/year. In the early 2000s, SMC built its own manufacturing facility and services 220,000 retail outlets. In addition to SMC, there are now 30 - 40 ORS suppliers. ORS is supplied for free in the public sector and is very cheap in the private sector (US$ 0.06/kit). Bangladesh relied on the family unit to sustain ORS use – the majority of mothers now educate their children on ORS, removing the need for repeated marketing campaigns. The most recent Demographic and Health Survey by USAID provides clear evidence of successful ORS scale-up efforts: 78% of recent diarrhoea episodes were treated with ORS (and 41% were treated with zinc), and only 2% of all under-five deaths were attributed to diarrhoea.

7.2.7 Global campaign on ORS

In 1978 WHO established the Diarrhoeal Diseases Control programme focusing on children under five years old. From the outset, WHO worked closely with UNICEF, as well as aid agencies of individual governments, notably that of Denmark, Sweden, the UK and the USA, along with NGOs. Countries were helped to develop their own national diarrhoeal disease control programmes, and five years after WHO’s Diarrhoeal Diseases Control programme came into being, 52 countries were launching their own plans. By 1986 that number had risen to 100 and eventually swelled to 130 countries, which were home to more than 99 percent of all children in the developing world. Some countries wanted to develop ORS in pre-prepared packets. Working together, WHO and UNICEF made sure that these countries had a reliable supply of these, giving technical expert assistance where it was needed, helping them build factories and providing raw materials and packaging. By the end of 1985 more than 40 developing countries were manufacturing their own salts.
7.2.8 Global adoption of ORS

Twenty countries (out of a total of 96 with data from the standard surveys series that track ORS coverage) have entirely failed in promoting rational diarrhoea management, with less than 25% paediatric diarrhoea episodes treated with ORS. Yet many other countries have done much better: 29 countries use ORS in half or more of all episodes; eight countries use ORS in two-thirds or more of all episodes. As can be seen in the map, by this criterion, high performing countries are found in every region of the world. The vast majority of countries in sub-Saharan Africa, as well as many Indian states, have neither particularly low nor encouragingly high rates of ORS use, but rather fall somewhere in the middle.

Figure 12: Map of ORS use rates around the world.²²⁵

7.2.9 Challenges to adoption of ORS

Resistance and opposition to rehydration salts formula came (and continues to come) from paediatricians and child health specialists trained in the developed world. It required years of patient persuasion to win their support. To get ORS to the masses, researchers faced huge obstacles: a medical culture that clung to intravenous therapy as superior to what they perceived as a primitive oral form; a very high prevalence of illiteracy, especially among women; and no way to distribute ORS packets to remote, roadless areas.
8 Lessons for education in LEDCs

As per the introduction, perhaps the most striking feature of the history of medicine is the commitment to science: the curiosity to investigate patterns, the tenacity to painstakingly collect data and eliminate biases, and the openness to the notion that one might be wrong.

“The real purpose of scientific method is to make sure Nature hasn’t misled you into thinking you know something you don’t actually know.”
- From Zen and the art of motorcycle maintenance by Robert Pirsig

Medicine also has a marked focus on results – on outcomes – having long shed the focus on inputs which still characterises much discussion about education (e.g., the MDGs).

8.1 Short-cuts

The most obvious short-cut is that education does not have to identify the various biases which misled us into thinking we know (or see) things which we actually don’t nor does it need to invent experimental tools and research methods to overcome them. Cohort studies, natural experiments, controlled studies, randomisation to deal with selection bias, follow-ups to overcome survivor bias, plus newer tools around analysing big data sets: these all exist now. Education can also use knowledge from elsewhere about how ideas disseminate and diffuse (e.g. Everett’s work on adoption of technologies, broadly defined, which coined the term ‘early adopter’): this is useful for aiding research uptake. In terms of how ideas are presented and what aids uptake, we now have many behavioural insights, and knowledge of barriers other than ignorance – such as inertia and ideology. We know more about how organisations develop and learn. Education can use these to avoid the trap of simply broadcasting information (which perhaps nobody asked for) and expecting change to ensue.

The need for that research to be clear and to be findable are evident now, as are some tools for doing so. Information infrastructure (such as repositories of papers, and pre-registration of trials) in education doesn’t need to be invented, just created. The same for standardised reporting formats (like CONSORT and the others). These kinds of infrastructure are perhaps uniquely useful places for philanthropic money interested in making and sharing ‘discoveries’ of what works in education.

8.2 Initiatives to prioritise and try

Surely the ‘global (eventual) aim’ is to improve educational attainment by making educational practice better based on sound evidence. ‘Intermediate aims’ on which funders and practitioners can usefully work include creating more of a culture of curiosity and experimentation in education. Steps towards that could include:

8.2.1 ‘Show your working’

Asking educational practitioners to explain clearly and publicly (a) what their intervention is (perhaps using some variant of the TIDIER framework used to describe medical interventions in enough detail that they can be replicated) and (b) the evidence on which that intervention / their work draws. Such an approach is soon to be tried with the UK government policies. If the goal is to encourage educational practitioners to use evidence, this ‘disclosure mechanism’ may valuably prompt them to look for evidence relevant to their work. Its main value may be in normalising amongst practitioners the notion that research is useful to them, and having them consider it.
8.2.2 Getting more evidence – about both prevalence and effectiveness

As discussed in the first Getting Better report, there is certainly need for more research into education, of a type which can guide funding, policies and practice. On prevalence, in many regions, the educational attainment of students of various ages seems to be unknown, so evidence-based decisions about allocating resources between them are simply impossible. On effectiveness, many interventions are unstudied or under-studied, even ubiquitous ones such as blackboards.

One effect of encouraging practitioners to ‘show their working’ may be to highlight the gaps in knowledge – to highlight that to practitioners themselves, funders and policy-makers.

8.2.3 Monitoring the quality of evidence produced by and about education providers

As we have seen, a great deal is known about the quality of medical research, and therefore the problems in it which need addressing. ‘Research about research’ (meta-research) like this could be very valuable in education for both identifying what needs to change and also for raising awareness that the quality of research matters and that research is not all automatically correct.

The meta-research might include both quality and coverage, e.g., ‘evidence gap maps’ (3ie is starting to produce these for various areas of international development). Such gap maps could be usefully combined with the ‘show your working’ initiative, to indicate where evidence does exist but isn’t being used, and also to indicate where additional research is more sorely needed.

8.2.4 Involving teachers and practitioners in prioritising and producing research

In education, practitioners (teachers) are normally quite separate from the research process. They may be studied (like passive lab-rats) but are rarely involved in deciding what gets researched, nor when, nor designing the studies, nor in deploying the findings.

If well-designed, perhaps the research process could itself become a useful tool for raising curiosity amongst teachers. They could be consulted about research priorities (in a process somewhat like the James Lind Alliance), and involved in research design. Teachers could be encouraged (and paid) to be involved in experiments to find what works in research uptake, e.g., in writing lay summaries of research, or in taking messages (acting as ‘evidence vectors’) to their communities.

Bearing in mind the value of ‘freelance’ experiments by self-starting doctors such as Semmelweis and Forssmann into maternal deaths and blood circulation, perhaps teachers in LEDCs should be encouraged to do more experiments themselves. There could be a small fund for this and perhaps a prize to raise awareness. Tools such as RandomiseMe226, an easy online tool for running RCTs, may be useful, e.g., for experiments in a single class or involving several schools. The STIR network of teachers might be a good place to start.

Given the difficulty of getting adoption of standards for good quality research and for good quality reporting of research, education should define and deploy such standards early.
8.2.5 Experiments with disseminating evidence and research uptake

Since the aim is influencing practice rather than just publishing research, education should work on this directly. Medicine has done various experiments to identify the most effective ways of sharing research such that it is used (e.g., five-minute podcasts vs two-page summaries), and education should do likewise. Set up high-quality experiments, involve teachers in designing and running them, and monitor their effect on practice and attainment (which assumes some usable attainment data). Not all the experiments should assume the ‘read it and heed it’ model of research uptake: people can heed research for plenty of reasons other than having read it, e.g., if it is easy, or if instructions arising from it are on some checklist that they use. Models might include: dissemination of written material, of recorded material, through word-of-mouth networks, through checklists, through campaigns, though engaging communities to require more of their children’s teachers, etc.

Research uptake is a growing field, and it may be useful for the education community to have a synthesis of its findings and insights.

8.3 Things to avoid

The main thing to avoid is assuming that research will produce action. There is (appears to be) no magic impact fairy to whisk insights from a journal to a classroom. Hence the dissemination and mobilisation pieces are crucial to changing anything beyond the research industry. For example, publication bias was first documented in 1980, yet thirty years on, remains horribly unfixed, mainly because there has (until now) been no ‘ground war’ of activism: there have simply been many papers documenting it, all of which can be safely ignored by people hiding their research.\textsuperscript{227}

In the production of research, it’s pretty clear that incentives matter. Simply making research available, clear and findable is inadequate, as is providing infrastructure. For example, adherence to standardised research reporting only becomes normal if there are incentives. Even the journals which endorse CONSORT don’t enforce it – and a commercial journal offered a dramatic but non-CONSORT compliant paper which will be purchased many times can scarcely be expected to do so.

We have talked little about incentives in this document, and that is perhaps a gap to investigate further. However, we can see the following:

The commercial interests of for profit journals are clearly a problem – and hence non-profits have set up their own, such as the Wellcome Trust and Max Planck Institute. Commercial journals have incentive to, for example:

- keep material behind pay-walls, which obviously hinders dissemination
- favour papers with eye-catching large effect estimates, which are likely to be spurious outliers, or just wrong. A related problem is journals encouraging researchers to look for sub-groups in which effects are large (‘p-hacking’, named after p, the chance of that result given the null hypothesis). Richard Peto recounted an example of The Lancet requiring sub-group analysis – ‘we wouldn’t do it because sub-group analysis kills people’ – but they eventually conceded, looking at groups by star sign which (presumably by random chance) did show some ‘significant’ results
- disproportionately publish papers by commercial entities (e.g., drug companies) which will buy reprints to use as sales collateral, or sponsor conferences.

Hence to the extent that education funders can encourage or mandate research to be published in free publications (like PLoS), they should do so.
The incentives on academics are rather bizarre, in that they are encouraged to publish (a lot) in ‘high impact’ journals in order to get tenure, even though it’s widely recognised that:

1. ‘High impact’ in the research community does not correlate with ‘high impact’ in anybody else’s lives. Academics appear to have literally zero incentive to engage with anybody who can use their research: academics presenting at the University of Chicago’s Science of Philanthropy Initiative conference which convenes them and practitioners from charities and elsewhere commented on how ‘embarrassingly rare’ it was to meet people who can use their findings. Hence, though some do, many don’t. Ben Goldacre started musing recently about ‘modelling [the] total number of deaths from academics failing to do anything useful to get their findings into practice’
2. Volume of publications has little to do with quality
3. The language in which academic papers are written (particularly the weird maths in many economics papers) is not designed to be accessible outside that research community: in fact it often appears to be designed precisely to exclude outsiders
4. Journals in which academics are rewarded for publishing are mainly interested in findings which are large or surprising or new. There is strong incentive to do the first study of a particular type but very little to do the second or third (since it’s hard to get them published). This is a huge problem – and bizarrely unscientific – since science is fundamentally about phenomena which always pertain, not which are only observed once
5. There is no reason to believe that the topics in which academics and their journals are interested in bear any resemblance to those which interest policy-makers or practitioners – or the people they serve. As Daniel Kahneman rather scathingly observes: ‘the main reason that decision theorists study simple gambles is that this is what other decision theorists do’. This is not just a theoretical problem: the mismatch in medical research between what academics left to their own devices will study, and what patients and their carers want them to study, is huge, as we have seen.

Hence it is not obvious that the ideal set-up is for problems and solutions in education to be studied by academics. They are not specialised or set-up for the ‘ground war’ of influencing policy or mobilising for action. Better alternatives may be others who can use rigorous methods but who can (i) address problems felt by policy-makers irrespective of their propensity to produce academically interesting papers and (ii) mobilise to get the answers enacted. They include Ideas42, Innovations for Poverty Action, ID Insight, and specific campaigns.

Of course the incentives on academics and those on commercial journals can combine to really hinder action. Academics may be happy to put material in journals which publishers keep pay-walled, and neither of them has much reason to care if readership is low. Professor Simon Chapman, an Australian public health academic, reflected on 38 years of public health campaigning that: ‘As researchers, we undertake research and systematically review it to provide evidence to lever policy and practice change or defend existing policies and practices. But there is only a small number of people who have the power to effect change or defend good policies. The most important of these are politicians. And guess what? They don’t read research journals!’
Incentives on commercial providers is the most distorting feature of medicine – most notably drug trials. We recounted the problems in Getting Better, which include:

- Only publishing trials which are flattering. The bias which this introduces into the literature is laughable / horrifying when one considers the lengths researchers go to eliminate bias within individual studies
- Not pre-registering trials (since this makes an unpublished one harder to hide), and obstructing attempts to introduce pre-registration
- Measuring numerous outcomes and only publishing details of those which flatter
- Not publishing full protocols stating all outcomes, and not including full experimental details in the eventual post-research reports (since this makes selective publication of outcomes harder to hide), and obstructing attempts to require this
- Rigging trials to produce flattering answers. This happens in two main ways:
  i. Comparing the treatment with nothing, when there are treatments known to be much better than nothing
  ii. Comparing the treatment with something known not to work - or not to work well or not to work well in that population and with that problem. The vast majority of trials by pharma companies are against products produced by the same pharma company, and in 97 percent of head-to-head trials by pharma companies, the drug comes out well
- Regulatory capture.

Where education / educational products are provided by commercial providers – and particularly when their effectiveness is researched by those providers – great care should be taken to ensure that the research is accurate, fully reported and free from bias. Commercial incentives may eagerly mislead us into thinking that we know things which we actually don’t. Simply requiring researchers to state their commercial interests is manifestly inadequate.

These are potentially solvable by regulating trials – to ensure that trials are pre-registered, fully described, designed around a problem which some real world person actually has, not rigged, and fully reported.

That said, regulation seems to have been badly applied in medicine, leading to over-regulation (in some parts, but still under-regulation in others). For instance, doctors can easily administer a drug of uncertain effect to a whole population, but administering it to half that population as part of a trial to rigorously assess its effect requires complicated regulatory burden. Education is nowhere near this point now, but care should be taken as it becomes more experimental to ensure that the aiding and impeding effects of regulation are as one would wish.

Lastly, medicine shows the need to avoid looking for silver bullets. There are few silver bullets: few, that is, which stand up to scientific scrutiny and aren’t just the product of being misled. Even those which are considered silver bullets (like vaccines and ORS) require a complex health system to achieve desired outcomes. It is therefore, better to change our expectations to look for modest improvements, and then implement them unstintingly.

Repeating trials, and running large trials is essential to avoiding being misled, since a single trial in a single place and/or with a small number of people can produce an atypical answer. Such trials are perfectly normal in medicine, yet unusual in education (and elsewhere). This may account for the ‘worm wars’: the argument about educational benefits of treating children for soil-transmitted helminths (worms) which arose from a single study in an area of Kenya with particularly high wormloads. Repeat trials need funding. They also need incentives – which are weak for academics.
Prioritising research topics to identify the interests of policy-makers and practitioners is essential, as is incentivising researchers to work on them. We heard criticism of some trials in education as ‘addressing a question that nobody’s asking’ such as choosing between deworming and hiring more teachers: such criticism is that research resource is being wasted. We talked in Getting Better about the James Lind Alliance’s process for eliciting research priorities for ‘beneficiaries’ (here, patients, their carers and doctors), and an analogous process could be used in education.

8.4 Things to investigate

There are a few pieces of infrastructure in medical research which ought to dramatically improve the system but which actually are underperforming. Rather than the obvious response of mimicking them, education might do better to investigate or experiment with alternative systems.

Pre-registration is one. This is supposed to deter publication bias, but the rampant publication bias in medicine shows that it’s clearly failing. Perhaps a better system would combine funding with pre-registering: if research grants were divided into (i) a small portion to plan a study, and (ii) a second portion to execute the study (and the second part depended on the plan being full and pre-registered somewhere sensible) then presumably pre-registration would become more common. However, this would only work if all (or at least, most) research funders adhered to this system. Though large ones (such as the World Bank) might, many small charitable foundations who do fund research may not self-identify as ‘research funders’ and may need reaching and ‘educating’ to adhere: that could be hard. Furthermore, pre-registering doesn’t solve publication bias: it just makes it easy to spot, but only if somebody is looking. Medical research has some ‘police’ doing meta-research to spot non-reporting, poor reporting, poor practice etc. – and education will need such ‘police’ but – evidently not enough. It’s not clear how to fund them. The All Trials campaign is considering having members of the public police whether pre-registered trials ever get published. [One would think that policing of pre-registering and publishing of research would be done by regulators, but the regulatory capture impedes this.]

In short, the incentives on medical researchers to always publish are too weak, so if education replicates medicine’s system, it may just replicate the problem.

Reporting standards is another. Again, medicine has frameworks (e.g., CONSORT) but incentives (on both researchers and publishers) are weak and hence compliance is low. Again, replicating the system may simply replicate the problem.

Some interesting developments are afoot in medicine. One is automating the collection of data about research process and findings from within the researchers’ workflow. This would (a) reduce the cost to them of reporting, (b) potentially make compliance unavoidable, both of which should increase compliance. A third effect is to speed the production of systematic reviews: because producing reviews takes around 6-9 months as the process is somewhat manual, reviews are invariably out of date, omitting the most recent trials. Education could join the work on auto-collection of material to avoid this problem.
8.5 Red herrings

A few reasons are commonly given for not taking lessons from medicine about evidence, which this document clearly shows are red herrings. They do and will arise as objections in education to learning from medicine. For example:

“You obviously need evidence in medicine because it is all about life and death. The stakes are so high. And it’s so clear if somebody dies”.

- First, most medicine isn’t about life and death: it’s about skin conditions or heart murmurs or dementia
- Second, many conditions which are actually fatal take a long time to become so, and hence we are easily misled into thinking the death was caused by something else. Long trials are needed to reveal the harm
- And third, as we’ve seen, most improvements are tiny: it’s their collective impact which is impressive.

**External validity**: this gets stated as ‘drugs work the same in all people, so trials are possible and their results widely used, whereas education is highly context dependent’. It’s just not true: different groups respond quite differently to events and treatments, and hence the first basic part of reporting a trial is to describe the population in which it was done. (The four basic parts are PICO: population, intervention, control / comparison, outcome.)

8.6 Conclusion

Medicine has made vast strides in improving the length and quality of lives around the world. There is much which education can learn from about uncovering what works and implementing that – and which education needs to learn, given its poor rate of improvements. Some lessons are easy to absorb, such as the importance of studies which stay the course from intervention through to final outcome. But it should not import everything wholesale, since many substantial problems in medical research and practice remain.

Education should learn from medicine, not blindly mimic it. Building the infrastructure and incentives for creating, sharing and using sound evidence is a good place to start.
## Appendix 1: CONSORT 2010 checklist of information to include when reporting a randomised trial, and flowchart

<table>
<thead>
<tr>
<th>Section/Topic</th>
<th>Checklist item</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Background and objectives</strong></td>
<td>Identification as a randomised trial in the title</td>
</tr>
<tr>
<td></td>
<td>Structured summary of trial design, methods, results, and conclusions (for specific guidance see CONSORT for abstracts)</td>
</tr>
<tr>
<td></td>
<td>Scientific background and explanation of rationale</td>
</tr>
<tr>
<td></td>
<td>Specific objectives or hypotheses</td>
</tr>
<tr>
<td><strong>Trial design</strong></td>
<td>Description of trial design (such as parallel, factorial) including allocation ratio</td>
</tr>
<tr>
<td></td>
<td>Important changes to methods after trial commencement (such as eligibility criteria), with reasons</td>
</tr>
<tr>
<td><strong>Participants</strong></td>
<td>Eligibility criteria for participants</td>
</tr>
<tr>
<td></td>
<td>Settings and locations where the data were collected</td>
</tr>
<tr>
<td><strong>Interventions</strong></td>
<td>The interventions for each group with sufficient details to allow replication, including how and when they were actually administered</td>
</tr>
<tr>
<td><strong>Outcomes</strong></td>
<td>Completely defined pre-specified primary and secondary outcome measures, including how and when they were assessed</td>
</tr>
<tr>
<td></td>
<td>Any changes to trial outcomes after the trial commenced, with reasons</td>
</tr>
<tr>
<td><strong>Sample size</strong></td>
<td>How sample size was determined</td>
</tr>
<tr>
<td></td>
<td>When applicable, explanation of any interim analyses and stopping guidelines</td>
</tr>
<tr>
<td><strong>Randomisation:</strong></td>
<td>Method used to generate the random allocation sequence</td>
</tr>
<tr>
<td></td>
<td>Type of randomisation; details of any restriction (such as blocking and block size)</td>
</tr>
<tr>
<td><strong>Allocation concealment</strong></td>
<td>Mechanism used to implement the random allocation sequence (such as sequentially numbered containers), describing any steps taken to conceal the sequence until interventions were assigned</td>
</tr>
<tr>
<td><strong>Implementation</strong></td>
<td>Who generated the random allocation sequence, who enrolled participants, and who assigned participants to interventions</td>
</tr>
<tr>
<td><strong>Blinding</strong></td>
<td>If done, who was blinded after assignment to interventions (for example, participants, care providers, those assessing outcomes) and how the blinding was done, with reasons</td>
</tr>
<tr>
<td><strong>Statistical methods</strong></td>
<td>Statistical methods used to compare groups for primary and secondary outcomes</td>
</tr>
<tr>
<td></td>
<td>Methods for additional analyses, such as subgroup analyses and adjusted analyses</td>
</tr>
<tr>
<td><strong>Participant flow</strong></td>
<td>For each group, the numbers of participants who were randomly assigned, received intended treatment, and were analysed for the primary outcome</td>
</tr>
<tr>
<td><strong>Recruitment</strong></td>
<td>Dates defining the periods of recruitment and follow-up</td>
</tr>
<tr>
<td></td>
<td>Why the trial ended or was stopped</td>
</tr>
<tr>
<td><strong>Baseline data</strong></td>
<td>A table showing baseline demographic and clinical characteristics for each group</td>
</tr>
<tr>
<td><strong>Numbers analysed</strong></td>
<td>For each group, number of participants (denominator) included in each analysis and whether the analysis was by original assigned groups</td>
</tr>
<tr>
<td>Section/Topic</td>
<td>Checklist item</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>-----------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Outcomes and estimation</td>
<td>For each primary and secondary outcome, results for each group, and the estimated effect size and its precision (such as 95% confidence interval)</td>
</tr>
<tr>
<td></td>
<td>For binary outcomes, presentation of both absolute and relative effect sizes is recommended</td>
</tr>
<tr>
<td>Ancillary analyses</td>
<td>Results of any other analyses performed, including subgroup analyses and adjusted analyses, distinguishing pre-specified from exploratory</td>
</tr>
<tr>
<td>Harms</td>
<td>All important harms or unintended effects in each group (for specific guidance see CONSORT for harms42)</td>
</tr>
<tr>
<td>Limitations</td>
<td>Trial limitations, addressing sources of potential bias, imprecision, and, if relevant, multiplicity of analyses</td>
</tr>
<tr>
<td>Generalisability</td>
<td>Generalisability (external validity, applicability) of the trial findings</td>
</tr>
<tr>
<td>Interpretation</td>
<td>Interpretation consistent with results, balancing benefits and harms, and considering other relevant evidence</td>
</tr>
</tbody>
</table>

**Other information**

| Registration                | Registration number and name of trial registry                                                                                                           |
| Protocol                    | Where the full trial protocol can be accessed, if available                                                                                            |
| Funding                     | Sources of funding and other support (such as supply of drugs), role of funders                                                                     |
Appendix 2: Flow diagram of the progress through the phases of a parallel randomised trial of two groups (that is, enrolment, intervention allocation, follow-up, and data analysis)
### Appendix 3: Checklist of items to include when reporting a systematic review or meta-analysis, and flowchart

<table>
<thead>
<tr>
<th>Section/topic</th>
<th>Checklist item</th>
<th>Reported on page No</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Title</td>
<td>Identify the report as a systematic review, meta-analysis, or both</td>
</tr>
<tr>
<td>Abstract</td>
<td>Structured summary</td>
<td>Provide a structured summary including, as applicable, background, objectives, data sources, study eligibility criteria, participants, interventions, study appraisal and synthesis methods, results, limitations, conclusions and implications of key findings, systematic review registration number</td>
</tr>
<tr>
<td>Introduction</td>
<td>Rationale</td>
<td>Describe the rationale for the review in the context of what is already known</td>
</tr>
<tr>
<td></td>
<td>Objectives</td>
<td>Provide an explicit statement of questions being addressed with reference to participants, interventions, comparisons, outcomes, and study design (PICOS)</td>
</tr>
<tr>
<td>Methods</td>
<td>Protocol and registration</td>
<td>Indicate if a review protocol exists, if and where it can be accessed (such as web address), and, if available, provide registration information including registration number</td>
</tr>
<tr>
<td></td>
<td>Eligibility criteria</td>
<td>Specify study characteristics (such as PICOS, length of follow-up) and report characteristics (such as years considered, language, publication status) used as criteria for eligibility, giving rationale</td>
</tr>
<tr>
<td></td>
<td>Information sources</td>
<td>Describe all information sources (such as databases with dates of coverage, contact with study authors to identify additional studies) in the search and date last searched</td>
</tr>
<tr>
<td></td>
<td>Search</td>
<td>Present full electronic search strategy for at least one database, including any limits used, such that it could be repeated</td>
</tr>
<tr>
<td></td>
<td>Study selection</td>
<td>State the process for selecting studies (that is, screening, eligibility, included in systematic review, and, if applicable, included in the meta-analysis)</td>
</tr>
<tr>
<td></td>
<td>Data collection process</td>
<td>Describe method of data extraction from reports (such as piloted forms, independently, in duplicate) and any processes for obtaining and confirming data from investigators</td>
</tr>
<tr>
<td></td>
<td>Data items</td>
<td>List and define all variables for which data were sought (such as PICOS, funding sources) and any assumptions and simplifications made</td>
</tr>
<tr>
<td></td>
<td>Risk of bias in individual studies</td>
<td>Describe methods used for assessing risk of bias of individual studies (including specification of whether this was done at the study or outcome level), and how this information is to be used in any data synthesis</td>
</tr>
<tr>
<td></td>
<td>Summary measures</td>
<td>State the principal summary measures (such as risk ratio, difference in means)</td>
</tr>
<tr>
<td></td>
<td>Synthesis of results</td>
<td>Describe the methods of handling data and combining results of studies, if done, including measures of consistency (such as I²) for each meta-analysis</td>
</tr>
<tr>
<td></td>
<td>Risk of bias across studies</td>
<td>Specify any assessment of risk of bias that may affect the cumulative evidence (such as publication bias, selective reporting within studies)</td>
</tr>
<tr>
<td>Section/topic</td>
<td>Checklist item</td>
<td>Reported on page</td>
</tr>
<tr>
<td>-------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>Additional analyses</td>
<td>Describe methods of additional analyses (such as sensitivity or subgroup analyses, meta-regression), if done, indicating which were pre-specified</td>
<td>No</td>
</tr>
<tr>
<td>Results</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Study selection</td>
<td>Give numbers of studies screened, assessed for eligibility, and included in the review, with reasons for exclusions at each stage, ideally with a flow diagram</td>
<td></td>
</tr>
<tr>
<td>Study characteristics</td>
<td>For each study, present characteristics for which data were extracted (such as study size, PICOS, follow-up period) and provide the citations</td>
<td></td>
</tr>
<tr>
<td>Risk of bias within studies</td>
<td>Present data on risk of bias of each study and, if available, any outcome-level assessment (see item 12)</td>
<td></td>
</tr>
<tr>
<td>Results of individual studies</td>
<td>For all outcomes considered (benefits or harms), present for each study (a) simple summary data for each intervention group and (b) effect estimates and confidence intervals, ideally with a forest plot</td>
<td></td>
</tr>
<tr>
<td>Synthesis of results</td>
<td>Present results of each meta-analysis done, including confidence intervals and measures of consistency</td>
<td></td>
</tr>
<tr>
<td>Risk of bias across studies</td>
<td>Present results of any assessment of risk of bias across studies (see item 15)</td>
<td></td>
</tr>
<tr>
<td>Additional analysis</td>
<td>Give results of additional analyses, if done (such as sensitivity or subgroup analyses, meta-regression (see item 16)</td>
<td></td>
</tr>
<tr>
<td>Discussion</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Summary of evidence</td>
<td>Summarise the main findings including the strength of evidence for each main outcome; consider their relevance to key groups (such as health care providers, users, and policy makers)</td>
<td></td>
</tr>
<tr>
<td>Limitations</td>
<td>Discuss limitations at study and outcome level (such as risk of bias), and at review level (such as incomplete retrieval of identified research, reporting bias)</td>
<td></td>
</tr>
<tr>
<td>Conclusions</td>
<td>Provide a general interpretation of the results in the context of other evidence, and implications for future research</td>
<td></td>
</tr>
<tr>
<td>Funding</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Funding</td>
<td>Describe sources of funding for the systematic review and other support (such as supply of data) and role of funders for the systematic review</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 4: Flow of information through the different phases of a systematic review

**Identification**
- No of records identified through database searching
- No of additional records identified through other sources

**Screening**
- No of records after duplicates removed
- No of records screened
- No of records excluded

**Eligibility**
- No of full-text articles assessed for eligibility
- No of full-text articles excluded, with reasons

**Included**
- No of studies included in qualitative synthesis
- No of studies included in quantitative synthesis (meta-analysis)
10 References

2 Reith Lectures 2014.
4 at Evidence Live, Oxford April 2015.
10 AllTrials.
14 Consort long-form statement
16 Systematic review in production at the time of writing.
24 Source: http://www.nsf.gov/statistics/srvyherd/#sd
25 Source: American Medical Association.
36 SORT IT is a collaboration between TDR, The Union, and Médecins sans Frontières (MSF).
40 See more at: http://www.avert.org/antiretroviral-drug-prices.html#sthash.qwyNOHt1.dpuf


87 Ibid. p. 183.
88 Ibid. p. 193.
89 Ibid. p. 256.
90 Ibid. p. 257.
98 Trelease, S.F., Yule, E.S. (1925) Preparation of scientific and technical papers. Baltimore, MD: Williams & Wilkins.


120 Personal Communication from Douglas Altman


134 Smith, R. Powerpoint presentation to the University of Zurich. [Online] resources.bmj.com/files/talks/ebm_zurich.ppt
138 Ibid.
142 Smith, R. Unpublished paper (on file with Giving Evidence)


190 ‘Chronology of the Cochrane Collaboration’.


196 Smith, R., unpublished papers.


200 Smith, R. Zurich lecture.


204 Sushruta, an ancient Indian medical scholar and the father of Ayurveda (1500 B.C.) had prescribed that cholera victims are to be: “given to drink a profuse quantity of tepid water in which rock salt and molasses have been dissolved; or clarified water combined with rice gruel.” Sushruta Samhita III, verse II. taken from [Online] http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1036912/


216 Chabbott, C. Institutionalizing Health and Education for All: Global Goals, Innovations, and Scaling Up (International Perspectives in Education Reform Series) [Online] https://kindle.amazon.com/work/institutionalizing-health-education-all-international-ebook/B00QMRP6RC/B00QMQWBWG


218 Chabbott, C. Institutionalizing Health and Education for All: Global Goals, Innovations, and Scaling Up (International Perspectives in Education Reform Series) [Online] https://kindle.amazon.com/work/institutionalizing-health-education-all-international-ebook/B00QMRP6RC/B00QMQWBWG


220 Chabbott, C. Institutionalizing Health and Education for All: Global Goals, Innovations, and Scaling Up (International Perspectives in Education Reform Series) [Online] https://kindle.amazon.com/work/institutionalizing-health-education-all-international-ebook/B00QMRP6RC/B00QMQWBWG


225 Data sources: Demographic and Health Surveys, Multiple Indicator Cluster Surveys and national survey series.


227 Private and public discussions with Ben Goldacre, plus his comments at John Ionnandis’ lecture, Oxford, June 2015.

228 Goldacre, B. (2015) tempted to write a paper modelling total number of deaths from academics failing to do anything useful to get their findings into practice. Personal Twitter Feed. [Online] https://twitter.com/bengoldacre/status/614549765016145920 [Accessed: 18.02.16].


233 http://www.prisma-statement.org/statement.htm