Getting Better:

What education systems in less developed countries can learn from evidence-based medicine

Caroline Fiennes and Liora Wulf

ADVICE ON **Giving Evidence** BASED ON

Table of contents

> **‘I beseech you, in the bowels of Christ, think it possible
> that you may be mistaken.’**
> *- Oliver Cromwell to the synod of the Church of Scotland, 1650*

> **‘There still seems to be considerable misunderstanding
> amongst the general public [...] about the relative
> value of opinion, observation and experiment in testing
> hypotheses.’**
> **-** *Archie Cochrane, the 'father of evidence-based medicine'* [1]

> **‘Remarkably little is known about the impact of education
> policies on student outcomes in developing countries’**
> *– Conclusion of a study of 20 years of education studies* [2]

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, and what isn't worth gathering.

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, the US Center for Effective Philanthropy, in the Stanford Social Innovation Review, 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-based Center for Global Development.

Since graduating with a degree in Politics and Philosophy, **Liora Wulf** has worked in the private and charitable sectors to improve project design, clarify thinking around outcomes and impact and secure funding in excess of £1 million for a range of large and small non-profit organisations.

We are grateful to the funders who enabled this work.

1 Introduction

'It's chilling that when we think we are doing good, we may actually be doing harm.'
– Ben Goldacre, epidemiologist and campaigner for the good use of evidence in health

This Getting Better project aims to help improve education in less economically developed countries (LEDCs). It grew from the observations that medicine has made marked performance improvements in many areas in recent decades, and that it has a strong culture of producing evidence, critiquing it and basing decisions on it.[†] This evidence-based practice contrasts starkly with the norm in education in LEDCs, where attainment standards are generally flat or sometimes declining.

The purpose of producing evidence is to influence behaviour – of practitioners, policy-makers, communities and others. So this is at root a behaviour change exercise. Richard Thaler of the University of Chicago, 'father of behavioural economics' and co-author of "Nudge: Improving Decisions about Health, Wealth and Happiness" cites two 'mantras' essential for evidence-based policy and practice:

- 'You can't make evidence-based policy decisions without evidence.
- If you want to encourage some activity, make it easy'.

The activity we want to encourage is finding and heeding evidence about education in LEDCs.

With that in mind, this project explores both health and education, looking at how and what evidence is produced (is it possible for decisions to be evidence-based?), and how it is shared and used (is it easy for anybody to find, understand and use?)

Others share our interest in increasing the use of evidence in education. For example, the UK's Department for Education and Skills commissioned a report in 2013 from Dr Ben Goldacre, author and Wellcome Research Fellow in Epidemiology, entitled 'Building Evidence into Education' The World Bank, the UK Department for International Development (DFID) and US Agency for International Development (USAID), have created an initiative named Building Evidence on Education to advance more rigorous research on international education.[3]

The need for change is clear from, for example, the estimation that only half of Africa's school-aged children will attend school and learn basic skills.[4] Or the fact that, in India, educational attainment has been falling despite rising spending[5]: at least 250 million primary school-aged children there cannot read, write or count well, even those who have spent four years in school.[6]

Project aims

There are enough similarities in the delivery mechanisms for health and education for the parallel to be relevant. We are therefore seeking to understand what education in LEDCs can learn about evidence from evidence-based medicine (EBM). To be clear, the aim is to *learn* from EBM, not necessarily to *emulate* it, since clearly education and medicine differ in many respects.

[†] 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 stated otherwise.

This project aims to identify institutions, skills, resources, relationships, culture and funding systems which would bolster the use of evidence in education. It is designed to identify ways to change the behaviour of actors in education – practitioners, policy-makers, funders and others – to ensure that practice is based on good evidence. Crucially, the project is a behaviour change exercise itself, also aiming to mobilise relevant organisations to make these improvements.

'Evidence-based medicine is the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients.'[7]
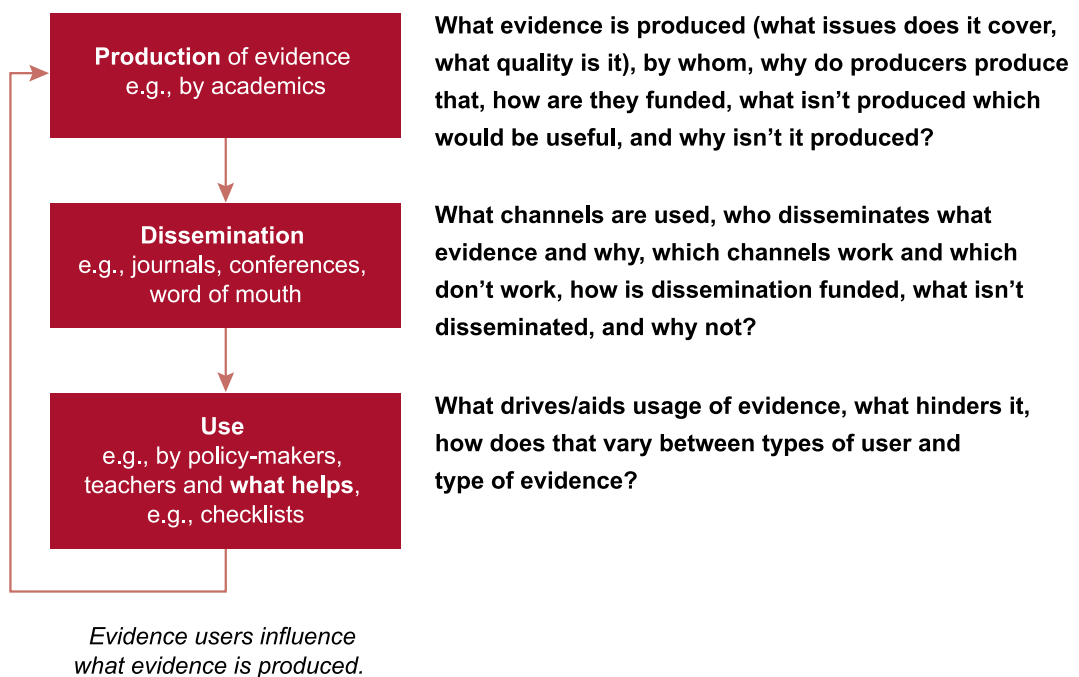
Evidence systems

By 'evidence system' we mean what evidence is **produced** (and how, by whom, why, and how it's funded), how it is **disseminated** and stored, and how it is **used**.[8]

The project also looks at the respective goals of health and education, which inform (or should inform) what evidence is produced, disseminated and used.

To allay one fear expressed about the project, we will not simply advocate for randomised control trials (RCTs) in education. In our structure, RCTs are one method for producing evidence – there are many other methods and plenty of other issues in dissemination and use.

Figure 1: Model of an evidence system[9]

Production of evidence e.g., by academics	**What evidence is produced (what issues does it cover, what quality is it), by whom, why do producers produce that, how are they funded, what isn't produced which would be useful, and why isn't it produced?**
Dissemination e.g., journals, conferences, word of mouth	**What channels are used, who disseminates what evidence and why, which channels work and which don't work, how is dissemination funded, what isn't disseminated, and why not?**
Use e.g., by policy-makers, teachers and **what helps**, e.g., checklists	**What drives/aids usage of evidence, what hinders it, how does that vary between types of user and type of evidence?**

Evidence users influence what evidence is produced.

1.1   The project plan: work so far and proposed next steps

Stage 1: Research

This report is the first output from the Getting Better project. Following discussions with senior people in education and medicine (see appendix 2) and a review of the literature, we have outlined the evidence systems in health and education. (We do not claim that this research is comprehensive. It was done on a relatively small budget and we spoke to people we could reach relatively easily.)

Based on this research, we are recommending various activities that would improve educational outcomes for students by improving the evidence system in education in LEDCs. These are listed in summary in Chapter 2 and discussed in the rest of the document.

We also identified other areas which might yield insights or ideas for improving the evidence system in education in LEDCs, such as how the evidence system in nursing works, since nurses are often more similar to teachers than are, say, surgeons. We would happily investigate these areas, and they are listed in Appendix 3.

Stage 2: Planning for action

Giving Evidence is starting to consult on the activities listed in Chapter 2 with various funders, policy-makers, analysts and practitioners in education. With them, we are prioritising the recommendations, identifying possible partners for them, and planning for action on them. Others are already working on some of the areas and we will not replicate their work but hope that our findings are useful to them.

We welcome your feedback on these recommendations, your ideas on whether and how to take them forward, and invite you to get in touch if you are interested in being involved.

Stage 3: Mobilisation

Once we have identified relevant partners and actors for the recommendations, we will begin work in those areas.

We invite your feedback on this document and the recommendations in it. Please direct comments to: Caroline Fiennes, Director, Giving Evidence caroline.fiennes@giving-evidence.com

1.2 Project scope

We have focused largely on government-funded health and education systems. We have not looked much at private healthcare systems (e.g., in the US) since the economics, incentives and resulting behaviours are so dissimilar to those in public education in LEDCs. Nor have we researched pharmaceutical companies or senior specialists in hospitals, as these differ too much from education systems in LEDCs.

In seeking to learn from the best practice in medicine, we take lessons primarily from more economically developed countries (MEDCs). Clearly this means that most of the lessons must be translated twice – between sectors, and between types of countries.

Figure 2: Applying lessons
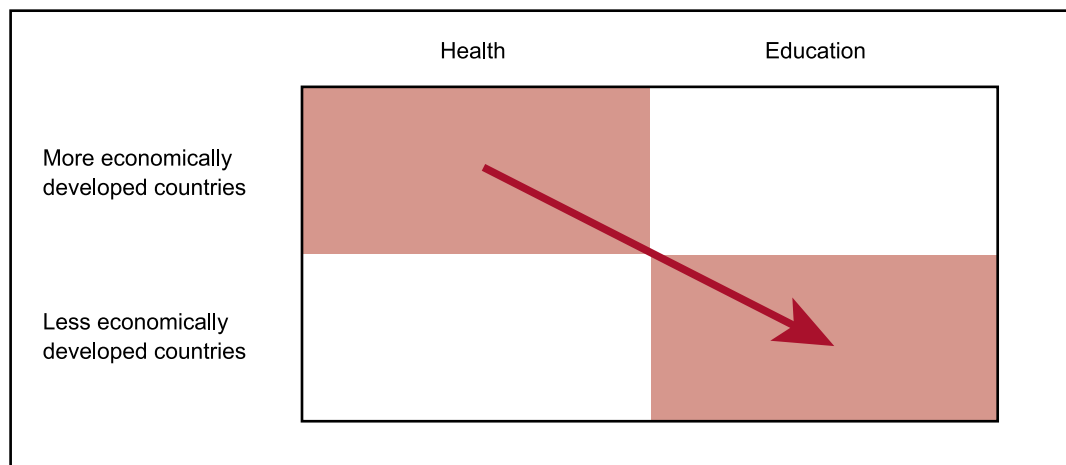
We are aware that the research does not adequately cover education specialists and qualitative research. This is partly because, to date, we reached out to more quantitative than qualitative researchers. But it's also partly because few of the qualitative researchers and education specialists we contacted acknowledged our request or agreed to talk with us.

We're also aware of the limitation that the research so far does not distinguish between conflict, post-conflict and non-conflict countries.

2 Emerging recommendations

Our investigations suggest that work on the following areas could improve the evidence system in education and hence educational outcomes for students. Hence we recommend investigating and working in the following areas. (The recommendations are ordered according to how they arise in the document, not the priority we would attach to them. In terms of sequencing implementation, the recommendations marked with an asterisk should perhaps be tackled before those without). Giving Evidence is working on some, and invite you to get in touch if you are interested in being involved.

2.1 Goals of education

1. *Improve clarity about the goals of education beyond literacy and numeracy. This will require hearing and considering the differing perspectives of stakeholders such as governments, employers, communities etc.

2. *Fund longer studies which track the effect of interventions right through to their intended effects (e.g., lifetime earnings). A possible alternative/adjunct is to explore using proximate outcomes in education research to reduce the need for lengthy studies.

2.2 Production of evidence

3. Fund work to assess (and thence improve) the quality of education research by non-academics. (This will require making the research more accessible and clearer about research methods used: see recommendations below.)

4. *Increase funding for producing better and more comprehensive baseline data in education in LEDCs. Many countries do not even know how many out-of-school children they have.

5. *Improve the ease of finding, assessing and using research by non-governmental organisations (NGOs) and other non-academics (including monitoring and evaluation). As a first step, create a repository for education research by NGOs, practitioners and any other entity, with requirements for clear description of the intervention used, the research method used, how it was used (e.g., how the sample was chosen), and the results. Funders could require entities they fund which conduct any monitoring and evaluation or other research to post it to the repository.

6. Improve the quality of research by NGOs and other non-academics. Recommendation 5 is necessary for this, but it will also require training those research-producers, training funders, and increasing partnership between them and professional researchers.

7. Develop a mechanism in education similar to the James Lind Alliance (see section 6.6) to ensure that the issues on which researchers work are those which practitioners, policy-makers, communities and other stakeholders would prioritise.

8. *Funders could encourage and facilitate more interaction between the two 'tribes' of researchers – educationalists and development economists. It seems likely that they could achieve more (and/or avoid redundant research) by collaborating better.

9. *Increase funding for researching the effectiveness of interventions, since so little currently exists.

2.3 Dissemination of evidence

10. Create better 'push' channels for disseminating education research to its intended users, such as journals, conferences.

11. *Increase the incentives on researchers (both academics and practitioners, e.g., NGOs) to share their findings with policy-makers and the field. Funders could make this a condition of funding.

12. *Make education research easier for users to 'pull' (find). This relates to recommendation 5, of publishing it in a findable place, and with standard searchable sections (as medical journal articles have standardised abstracts).

2.4 Use of evidence

13. Involve users in prioritising research topics, to ensure that research is relevant. This would secure buy-in from practitioners, making it more likely that they use the research, and reduce wastage.

14. *Produce guidance and training for education funders about the various types of evidence and their differing strengths and weaknesses.

15. Investigate mechanisms to make it easier for practitioners to consume, understand and use research. These might include 'practice guidelines' (analogous to doctors' clinical practice guidelines). They might also include checklists. Though terribly simple, checklists seem to be highly effective: in one study looking at the use of soap in Karachi, using a checklist halved cases of diarrhoea, a huge effect from a seemingly trivial intervention.[10]

16. Explore delineating binary 'conditions' in education (e.g., either a child needs remedial maths tuition or they don't) to help teachers make decisions (analogous to how doctors are aided in diagnosing conditions by binary delineation: either you have diabetes or you don't).

3 Summary findings from stage one

In summary, we found that education has a very limited evidence-base and little culture of produc-ing, critiquing or using research. Research in education is normally quite separate from practice, whereas in medicine it often happens in the same buildings and involves the same people. Remark-ably, even foundations which fund lots of education delivery and research, and some large delivery agencies, told us that they don't use research produced by others[11].

Table 1: Outline of the evidence systems in health and education

Evidence system stage	Health	Education
Production	Large amount of evidence produced.	Very little evidence produced.
	Significant funding of research (far exceeding education).	Scant funding for research.
	Some intermingling of qualitative and quantitative research methods.	Two 'tribes' of researchers, researching causal and non-causal parts of education. Neither tribe is particularly close to practitioners or policy-makers.
	Demand from practitioners for evidence is high.	Little demand from teachers, policy-makers and other practitioners.
	Intermingling of practice, teaching and research.	Little intermingling of practice, teaching and research.
Dissemination	Many dissemination channels reach practitioners and policy makers. Notable channels are medical journals, conferences, training, drug sales reps although there are a number of biases that exist in dissemination mechanisms.	Few (if any) dissemination channels reach practitioners or policy-makers. Journals are not aimed at practitioners, but at other researchers.
	The research that is published is generally quite well indexed and findable.	Published information is dispersed, not standardised or well indexed and so is hard to locate.
Use	Guidelines and checklists help practitioners to overcome cognitive and other barriers to use of evidence.	Few (if any) guidelines or translation of evidence, so research is hard to interpret or use for practitioners.

Thus it is currently probably impossible to run an evidence-based education system in an LEDC. This is partly because of history, and partly culture and funding. Some of the factors are probably amenable to change, and so, on the positive side, it seems that much could be improved relatively swiftly.

3.1 Goals: there are few agreed goals in education

In health, goals are reducing incidence, mortality and symptoms of disease. These goals are generally agreed, and lend themselves to measurement, such as whether the incidence of flu has changed.

By contrast, goals in education are contested, beyond basic literacy and numeracy. People variously argue that the goals for education are high-school graduation rates, college graduation rates, lifetime earnings, results in exams, vocational skills, knowledge for its own sake, and so on. Different goals would lead to quite different education systems. Clearly, researchers can't identify interventions that achieve education's goals without clarity on what those goals are.

Education's focus on inputs
Education policy tends to revolve around inputs, such as access, enrolment, attendance, and availability of teachers, rather than outcomes such as learning or employment. For instance, whereas the Millennium Development Goals (MDGs) relating to health focus on outcomes (better maternal health, reducing incidence of HIV/AIDS), the education MDG concerns an input – school attendance (specifically, "Ensure that children everywhere will be able to complete a full course of primary schooling"[12]), irrespective of what, if anything, happens in school.

**Table 2:** Millennium Development Goals – inputs or outcomes?

Health and education MDGs	Input?	Outcome?
Education		
Enrolment in primary education	✓	
Health		
Reduce child mortality		✓
Improve maternal health		✓
Combat HIV/AIDS, malaria and other diseases		✓

This focus on inputs would be fine if a clear link had been established between the input and desired goals of education, but this is not the case.

3.2 Production of evidence: comparatively little is produced in education

The overwhelming finding is that **there is much less evidence about education than there is about health**. On most education topics there is no evidence at all. Taking just one type of evidence, RCTs, the 20 years between 1990 and 2010 saw only 13 RCTs on interventions in education in LEDCs, against tens of thousands in medicine.[13] The graph below indicates the extent to which RCTs in health outnumber those in social fields.

Figure 3: The number of RCTs produced over time in health compared to the number of RCTs produced in social welfare, education, and crime and justice fields combined [14]
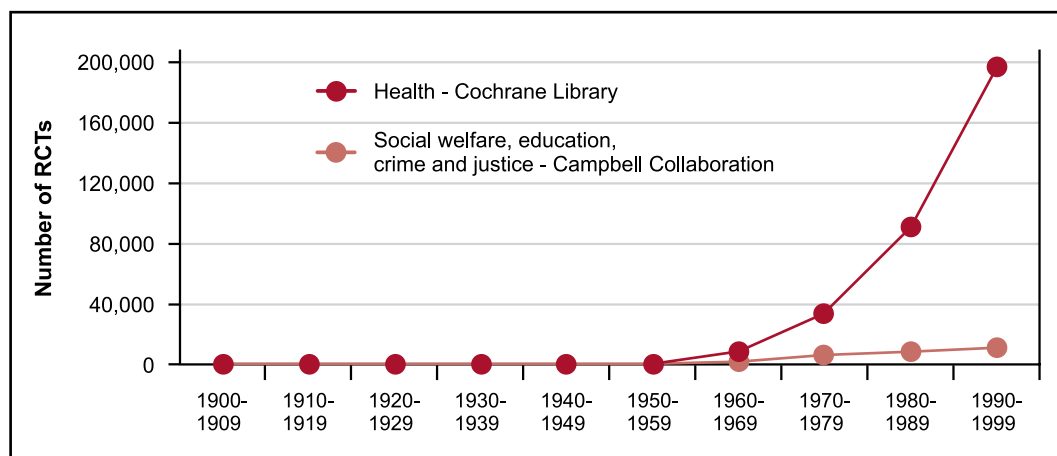
3.2.1 Funding of health research vastly exceeds that of education

This is partly due to the pharmaceutical industry, though government and charitable funding for research is also greater in health. In the UK for example, government spending on health research (domestically) was nearly 60 times greater than that on education research in 2010 – 2011.[15] In 2014, DFID spent £75 million on research into LEDCs, but less than £0.5 million on education research.[16]

This lack of funding obviously contributes to the lack of research in education. However we suspect that it is not the sole factor.

3.2.2 Education has patchy 'baseline' data

Baseline data provide a snapshot, showing how large a problem is and where it is. In health, they include data on incidence and mortality: e.g., how many people have flu and where are they? In education, baseline data might catalogue enrolment or reading levels in various locations. They are invaluable for prioritising problems though offer no guidance about improving performance.

Baseline data are surprisingly patchy in education: many countries do not have a clear picture of learning levels in various subjects. This obviously prevents policy-makers and practitioners knowing what to prioritise and seeing whether policies and processes are working.

3.2.3 There is a mismatch between the evaluation time-periods of interventions and outcomes

Most education research measures short-term outcomes such as test scores, despite the fact that the ostensible goals (e.g., labour market participation, political engagement) are much longer term. This would be acceptable if there were evidence to connect the short-term outcomes with longer-term goals, but currently there isn't.

A related issue is that policy-makers often need advice on decisions in quite different time-scales than researchers work to. For example, investigating the effect of a new two-year programme will (obviously) take at least two years, but a policy-maker or practitioner can't wait that long. They often need advice on the best material available, which creates tensions between them and researchers.

3.2.4 Research, practice and teaching are separate in education, whereas intermingled in medicine

Since Aristotle, medicine has intermingled teaching, research and practice: senior doctors often teach trainee doctors, do research and see patients. Most senior jobs (e.g., in Royal Colleges) require a strong research record. It supports evidence-based practice because evidence is generated by its users, who hence understand and can evaluate it.

This intermingling doesn't exist in education (nor indeed in most other disciplines). Consequently, education research can seem very distant an irrelevant to a teacher's work. We look briefly at policing, which also has no history of this intermingling but yet where evidence is increasingly important.

3.2.5 There is a much greater culture of research in health than in education

This intermingling means that doctors are taught to conduct research, to assess it and are exposed to it and involved in it throughout their careers. Hence it's natural for them to seek and use it. They're scientists by training and by nature: and science fundamentally involves acknowledging the limits of intuition and experience and looking for empirical findings. It's obvious that we can't understand the body intuitively and hence need experiments, the culture of which spills over to create experiments and empirical investigations in other areas of medicine, such as the best ways of teaching medicine or persuading patients to take more exercise.

Education (like policing) isn't like this: teachers aren't selected on their scientific ability, and the learning process isn't obviously unintuitive. It's more of a delivery culture, and hence demand for research is lower.

The causation isn't clear. Perhaps the lack of demand arises from scarce funding (nobody will seek research if there is none), or causes it (nobody's convinced funders to fund education research). Certainly demand may be limited by education research being (often rightly) perceived as context-specific: if the results of a trial in Kenya are not transferable to Uganda or Sri Lanka, it is unlikely that practitioners in those countries will bother to seek it out.

3.2.6 Education has two quite distinct groups of researchers

The interests of the first group, education specialists (educationalists), include pedagogy and context. The second group tends to comprise development economists, interested in the effectiveness of interventions. The latter therefore run studies that investigate causation, such as case control studies and RCTs. It's remarkable how little these groups work together, value, or even read each other's work. For instance, their journals are separate, as are their conferences, and we found them noticeably derogatory about each other.

3.2.7 Findings are less widely applicable in education than in medicine

The two issues here are fidelity and external validity (defined in Table 3 below).

Table 3: Barriers to generalisability

Fidelity	Did all beneficiaries receive the same intervention? Fidelity is high if 20 patients are given the same pill in the same dosage and in the same way. It is lower if they all have breast cancer removed, because (a) the surgery may be performed differently by different surgeons, and (b) the patients may have different conditions. Fidelity is lower still for mentally ill patients receiving a 'talking therapy', since clearly the therapist responds to what the patient says, which will be different in each case.
	Fidelity is low whenever there is scope for people to change what is done. Fidelity is generally low in educational settings, because:
	  - Teachers differ in their styles
	  - Teachers respond to their students, so effectively students get a range of interventions
	  - The resources available in schools vary e.g., no blackboard
	  - Local contexts such as parental involvement in education vary

External validity	If an intervention (say, providing books) works in one context, will it work elsewhere? External validity considers whether the results of an experiment will be valid outside the context in which the experiment was done. External validity is high in physics – the speed of light is the same in Japan as in Jamaica (although external validity is not always high even in physics: gravity's strength varies between Sweden and Singapore). People often say that external validity is high in medicine ('everybody reacts the same way to a pill') though this isn't always true – for example, people of different races, sizes and genders metabolise drugs differently. External validity is low whenever context is important, as it is for most for 'social' interventions, including education. For instance, treating children for intestinal worms dramatically reduces illness (and hence school absenteeism) in Kenya,[17] but won't achieve much in the Bronx since school absenteeism there has very different causes.

3.3 Dissemination: not effective in education

Dissemination 'pushes' material to users, and also allows them to 'pull' (find) it when they want.

Formal 'push' channels are much less established in education than in medicine: few journals or conferences reliably reach users (teachers, policy-makers etc.). In part, this is probably because there is much less evidence to share, and because medical 'push' channels are subsidised (sometimes covertly) by pharmaceutical companies.[18] But evidence in education is shared amongst academics and other producers: however there is so little attempt to share it with users that one senior interviewee derided education research as 'an echo chamber'.

Education also seems to have a poor 'pull' system, in that it's hard to find evidence. The separation between the two groups of academic researchers have different journals. And much non-academic material is unpublished or published in unlikely places and incomparable formats.

3.4 Use: evidence little used in education

The primary barrier to using evidence in education is simply its scarcity, along with ignorance about what has been produced and a range of cognitive biases.

Barriers to usage in medicine include some of the same biases and barriers as in education, including the facts that evidence is complicated and practitioners are busy. So medicine has 'clinical practice guidelines' – short summaries produced for practitioners by governments, regulators and professional bodies. There is no equivalent in education (as far as we know); teachers normally have a curriculum, and sometimes scripted lessons, though these are quite different.

Education could perhaps benefit from evidence-based guidelines for handling particular situations, e.g., a child often removed from school by parents, falling behind in mathematics, or whose behaviour suddenly changes (which may indicate a problem at home, such as domestic violence). Similarly, some areas of health have introduced checklists with surprisingly large success, yet these have not been tried (to our knowledge) in education.

4 Evidence-based medicine as a model for education

> **'I would like to stress how very far ahead the medical profession [...] is of other professions. What other profession encourages publications about its error and experimental investigations into the effect of their actions? Which magistrate, judge, or headmaster has encouraged RCTs into their 'therapeutic' and 'deterrent' actions?'**
>
> *- Archie Cochrane[19]*

Lessons from medicine are particularly relevant to education because health and education have such similar delivery systems. For example, every functioning government in the world has ministries for health and education, both health and education are delivered by trained practitioners (teachers, doctors), operating in institutions (schools, hospitals) which provide interventions (operations, lessons) to beneficiaries (students, patients) who know less about what they need than practitioners do. This strengthens our suspicion that lessons can be transferred.

Figure 4: Similarities between delivery systems in health and education
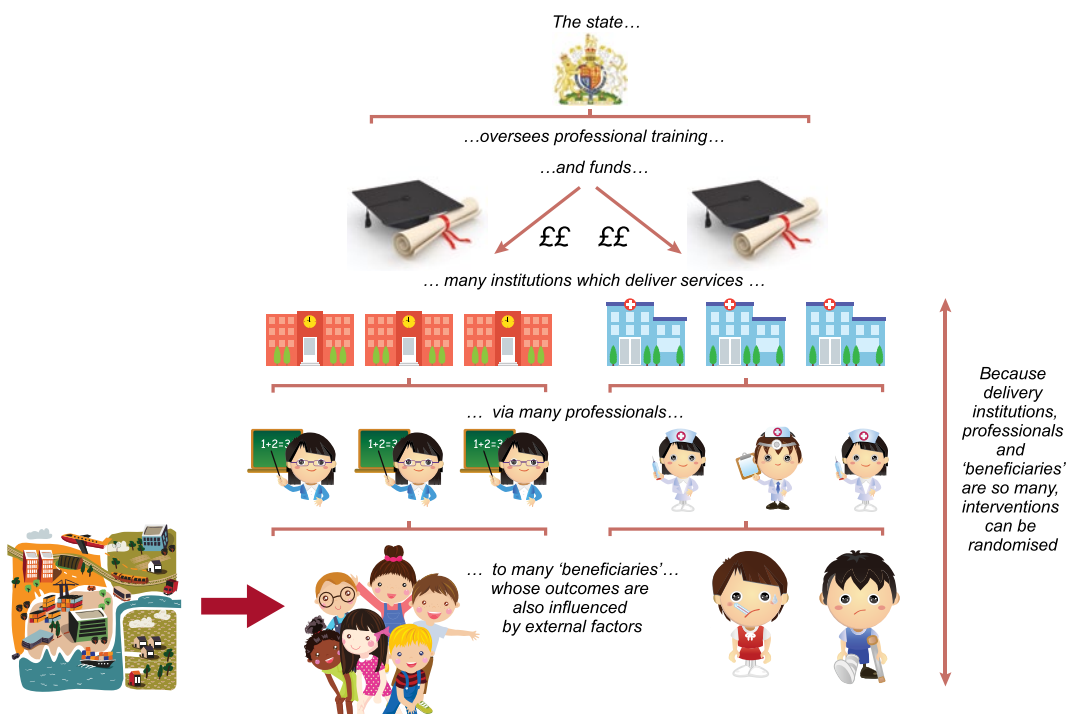
Current debates in education about performance and measurement are similar to those in medicine about 60 years ago. Teachers and educationalists often object to measurement on the grounds of the individuality of the people they serve, on the basis that they know what's best, because the true value of 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, and all have been overcome.

It is particularly surprising that, of all professions, medicine was so early to become evidence-driven. Doctors have always had high status – their opinions have always attracted respect – so it's remarkable that they willingly make their judgements subservient to the data.

**'What's the difference between a doctor and God?
God doesn't think he's a doctor.'**
*– standard joke*

We do not pretend that medicine is perfect; many significant problems with evidence in medicine remain, and we recite some in this document. Some arise from commercial interests of the pharmaceutical industry, for which there is no parallel in education and hence may be avoidable in education.

## 4.1 Changes in performance: some examples in health and education

Health

Despite the problems, it's clear that performance in medicine has dramatically improved, and below are some examples. (These graphs don't establish causation, of course, though there are many documented instances of practice changing as the result of better evidence).
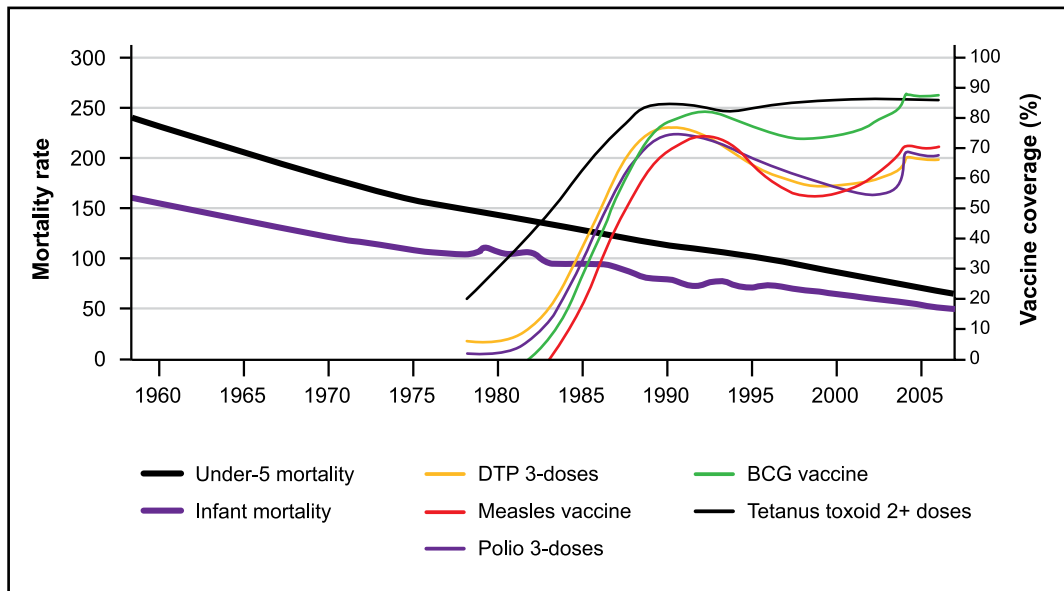
Figure 5: Crude death rate for infectious diseases – United States 1900-1996[20]
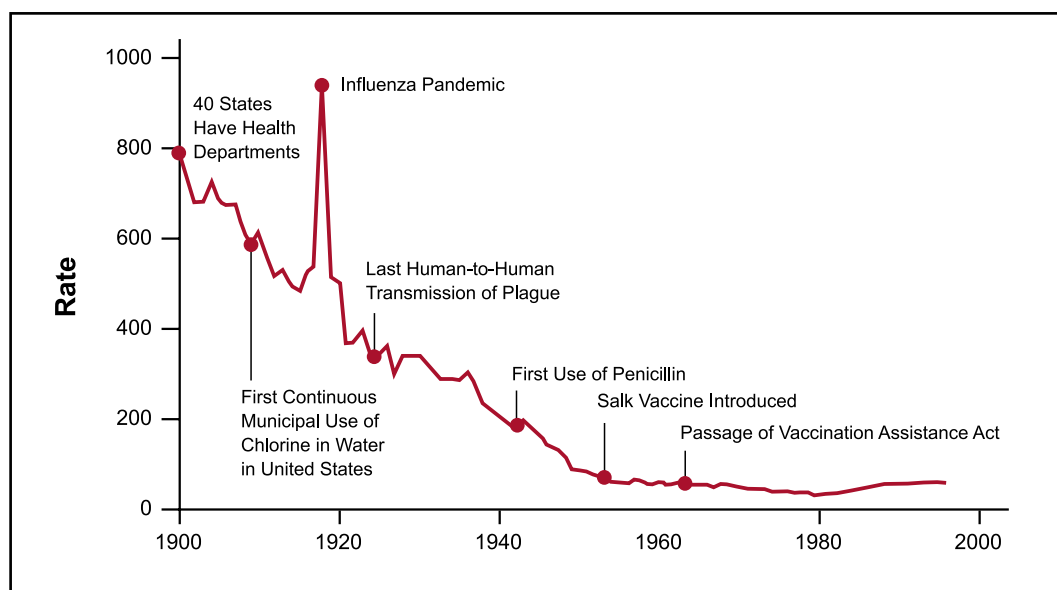
Figure 6: Global annual reported cases of Polio, and Pol3 vaccine coverage, 1980 – 2012 (World Health Organization, 2012)[21]
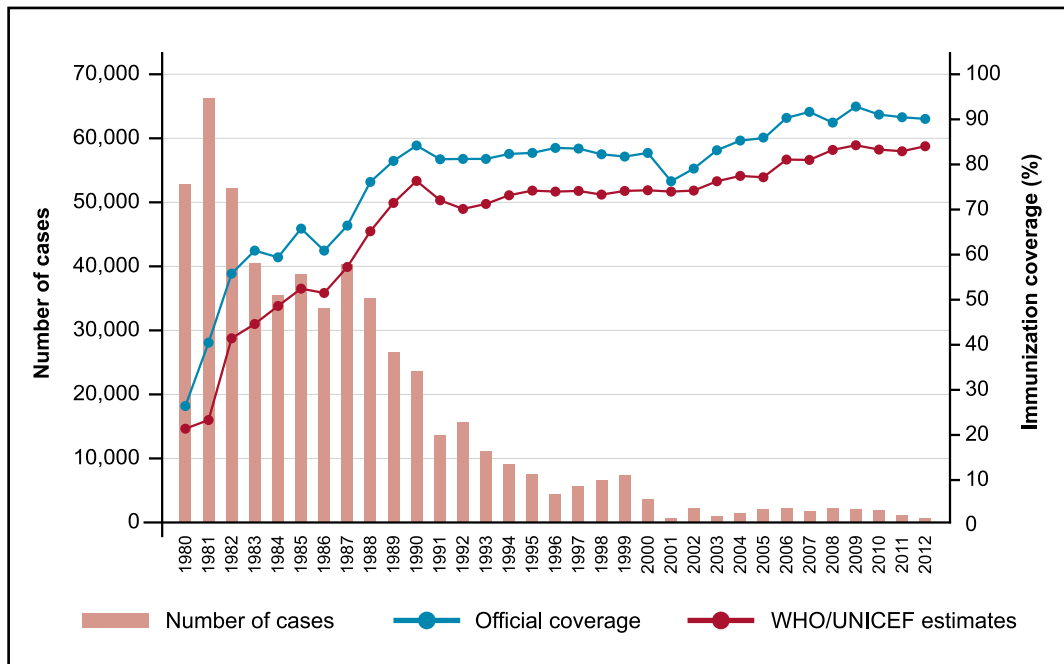

Figure 7: Under five mortality and infant mortality against vaccination coverage in India[22]

Education

Taking the example of India (for which good statistics are available), educational attainment has *decreased*, despite the introduction of a tax and legislation to ensure that children have access to education.[23]

Figure 8: Declining educational attainment from 2008 – 2012 in Indian government schools[24]
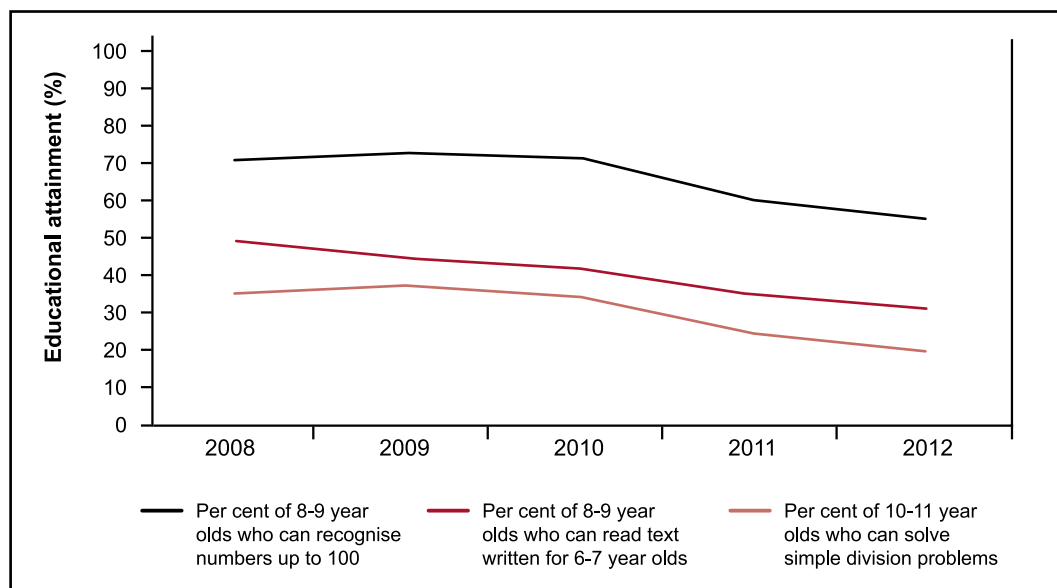
Live Mint and the Wall Street Journal reported that 'the proportion of all children in class five [10 years old] who can read a class two [seven years old] level text has declined by almost 15 percentage points since 2005. Similarly, the proportion of students in class 8 [13 year olds] who can do divisions [sic] has declined by almost 23 percentage points during the same period.'

4.2 Parallels between delivery systems in health and education

The significant parallels between the delivery systems in health and education include:

- **Funding:** In most countries, some level of healthcare and education is provided by governments, provided universally, and funded by taxation. These are often supplemented by private sector providers. Both systems absorb large amounts of public funding.

- **Delivery model:** Both have many practitioners (teachers, doctors, nurses) in institutions (schools, hospitals, clinics), serving beneficiaries (students, patients). As a result, many interventions in both systems can be randomised.

- **Training:** Generally, practitioners in both systems are trained, and the training is overseen by government, and often practitioners need to be registered with a professional body overseen by government.

- **Information:** In both systems, there is substantial information asymmetry – doctors and teachers (should) both know more about what the patient or student needs than the patient or student does. Hence the patient or student often cannot assess the treatment they receive. (This is quite different to business, where the customer states what they want and the provider isn't asked for a view of whether they're correct.) This asymmetry can lead to patrician behaviour, and hence requires regulation or inspection to ensure quality.

- **Incentives:** Providers of both public healthcare and education have little financial incentive to perform well (in contrast to, say, businesses).

- **Lifestyle influences on outcomes:** In both health and education, outcomes are affected not only by the delivery system (teachers in schools, nurses in clinics), but also outside influences such as parenting style and culture, or diet and exercise.

- **Potential to randomise:** Interventions in both health and education can be tested empirically because they both have many similar 'units': there are many similar deliverers (teachers, doctors), many similar recipients (students, patients) and many similar institutions (schools, clinics). This allows evaluations of many interventions to have adequate sample sizes, and can be randomised. This wouldn't be the case for, say, species conservation or legislative systems which are too few and too heterogeneous.

These parallels make medicine much more analogous to education than many public services (e.g., providing roads) and other work funded philanthropically (e.g., species conservation or defending human rights).

4.3 Differences between delivery systems in health and education

Of course there are many ways in which education differs from health, which is why this project seeks only to learn from health rather than to transplant its entire evidence system into education. Differences include:

- **Delivery model:** Education is largely a 'batch process' in that students are taught in groups, whereas healthcare is generally delivered to an individual. Education is also normally delivered proactively: children are schooled irrespective of their circumstances. Healthcare is generally reactive, with patients being seen only when they have a condition. (Public health work, e.g., vaccination campaigns, is both proactive and batch, and hence a subject for further study in this project).

- **Referral to specialists:** In health, patients are often referred by a generalist to a specialist. The education system generally does not work on a referral basis (with the possible exception of remedial classes): there are few specialist teachers who provide support for 'conditions' such as poor spelling or high potential in mathematics.

- **Goals/desired outcomes:** There is less agreement about the goals of education beyond basic literacy and numeracy (the accumulation of knowledge? high incomes? better jobs?) than in health. Furthermore, health outcomes are more apparent (the disappearance of my rash) and often transpire more quickly. By contrast, outcomes of education may take many decades.

- **Level of education of practitioners:** Teachers are generally not trained as long or as well as health professionals: in many LEDCs, many teachers are not trained at all.[25] Health practitioners (in MEDCs and to some extent elsewhere) are encouraged or required to have continuing professional development (CPD) and take exams as they become more specialised. Teaching hospitals allow students to learn from real cases, and doctors in MEDCs are trained to use

evidence and interpret clinical trials. This, with other factors, encourages practitioners to produce evidence. However, training for nurses and physiotherapists is lower than for doctors and hence these groups are more interesting for this study.

- **Role of commercial interests:** In education, the role of commercial interests is relatively low – except for private schools, these interests are limited to providing books and other equipment to schools. In contrast, the pharmaceutical industry is pivotal to the evidence system in health. This is so dissimilar to education that we have mostly excluded it from this study. Commercial interests play a smaller role in other parts of health, such as talking therapies and physiotherapy, which are consequently more interesting to this study.

- **Generalisability of findings:** As discussed, external validity and fidelity are generally more problematic in education than in health.

- **Politicisation and 'barriers to entry':** Unqualified outsiders often opine about educational methods and politicians often introduce education interventions. This is much rarer in medicine (though does happen in relation to the management of healthcare) – politicians rarely override doctors' opinions about medical interventions. The 'barriers to entry', so to speak, are lower in education than in health.

4.4 Are the differences between health and education too great?

Some commentators suggest that education and health are so fundamentally different that it is impossible to learn from each other. We see those differences, yet retain our suspicion that much can be learnt, as outlined in Table 4.

Table 4: Addressing differences that might challenge the analogy between education and health

Difference between education and medicine: Suggested barrier to learning from medicine	Response: Reason to still pursue learning from medicine
Context is paramount in education, unlike in health, precluding generalisations about education.	This project is particularly interested in parts of health which are highly context-specific, such as mental health. Perhaps one culture views mental illness as possession by the devil, while another views it as a purely physical disease. Clearly this will affect a patient's situation (they may get locked away) and hence the effect of an intervention. Conversely, some findings about human behaviour have been proven generally applicable (such as response to incentives) which can be applied to teachers and education.

Doctors in MEDCs are highly educated, able to follow complex arguments, trained and self-selected to think of themselves as scientists, and hence can reasonably be expected to read scientific papers.	This may be true for doctors in MEDCs, but health has some practitioners whose levels of education are similar to teachers', nurses', midwives' and physiotherapists'.
Health interventions have very high fidelity. For example, it is easy to be sure that every patient took the same pill.	This is true of pharmaceuticals, on which this project therefore will not focus. But it isn't true of, for example, talking therapies and physiotherapy.
Pharmaceutical companies have strong financial incentives to prove that their drugs work, and so are willing and able to spend on producing supportive evidence.	Indeed, and we largely exclude them from our analysis since they are so dis-analogous. Clearly it's harder to exclude their wider influence e.g., on institutions, conferences, journals and regulation, though we do so where possible.
Medicine is delivered to an individual and/or reactively, making it too different from education.	Though true for some areas of health (e.g., surgery), this isn't true of public health (e.g., vaccinations or behaviour change campaigns) which hence are interesting to this project.
Medicine's intermingling of practice, research and teaching is too different from education to make comparisons meaningful.	Again, this is not true of all medical professions, e.g., nursing or physiotherapy. We also take a brief look at policing, which also has no such intermingling.
It is easy to see whether a medical intervention is working because outcomes are clearly defined and relatively quick (e.g., neonatal survival), unlike in education.	There are many instances in healthcare where outcomes are difficult to detect and/or slow e.g., in mental health or public health. We look at the helpful development in health of making many conditions binary (either you have diabetes or you don't), from which education could perhaps learn.
Medicine has a culture of performance measurement, unlike education.	This was not always the case, and the emergence of EBM involved many stand-up rows. As noted, the objections are remarkably similar to those made now in education, and in some ways it's particularly surprising that medicine, of all disciplines, has welcomed evidence. Hence we look briefly at the history of how evidence gained traction in medicine, and will research this further in the next stage of our work.

4.5 Areas of medicine particularly interesting for this project

As a result of the differences and similarities discussed above, areas of medicine vary in their relevance to education and usefulness to this project, as outlined below.

Figure 9: Least and most analogous areas of health in relation to education

Least analogous parts of health	More analogous parts of health (likely to be the focus of future study)
- Drugs: high fidelity and external validity, commercial incentives. - Pharmaceutical companies: greater impact of commercial interests on the evidence system. - Hospital doctors and surgeons: highly educated, used to reading journals and critiquing evidence.	- Public health: proactive, batch process. - Mental health, nursing, physiotherapy: practitioners are similar to teachers in level of education and familiarity with evidence; fidelity and importance of context are more similar to education. - Medical education: process of training doctors is highly relevant to education.

4.6 The long road to evidence-based medicine

'I had learned that even with the best of intentions, it's possible to do great harm, even with some of the treatments I had been taught in medical school. I became rather obsessed by finding out what the quality of evidence was that we used to guide our practice.'

- Iain Chalmers, co-founder of the Cochrane Collaboration[26]

In many ways it is amazing that evidence now dominates medical practice, given that the medical profession once relied almost exclusively on the 'eminence, charisma and personal experience' of doctors to determine the best course of medical action for a patient.[27] For example, Ignac Semmelweis, working in Vienna in 1847, discovered that the cause of the higher death rate of women giving birth in hospitals instead of at home (20 per cent compared to 1 per cent) was Streptococcus bacteria carried by doctors who had recently carried out autopsies. Instead of being hailed for this life-saving breakthrough, Semmelweis was roundly mocked, and dismissed.[28]

'Murderous bastard'

At around the same time, Joseph Lister discovered that using carbolic acid to sterilise equipment during surgery could dramatically reduce infection, and published this evidence in *The Lancet* in 1867. However, Atul Gawande, a surgeon at Harvard Medical School and expert on patient safety, recounts how even 20 years later surgeons 'continued to operate in black frock coats stiffened with the blood and viscera of previous operations', despite the evidence on its detrimental effect.[29] It took an entire generation for proper sterilisation techniques to be introduced, despite clear evidence of its effectiveness.[30]

Academics saw research as something separate from practice, and not relevant to practitioners: Dr Archie Cochrane, the 'father of EBM' and one of its earliest and most significant proponents, recalled from his time at Cambridge University in 1948: 'I remember being advised by the most distinguished people that the best research should be utterly useless'.[31]

Cochrane highlighted throughout his career the areas of medicine least likely to rely on RCT evidence[32], challenging people in the medical community who believed that the opinion of doctors should take precedence over evidence. Cochrane challenged this 'god complex'[33] in one instance by instigating a trial to determine where patients who had suffered a heart attack should recover: in hospital, as assumed by cardiologists of the day, or at home.

Having gathered a group of cardiologists to reveal the interim results of the trial, Cochrane passed around a table comparing the progress of the two patient groups. Making clear that none of the results were statistically significant as the trial was not completed, Cochrane suggested that the figures indicated that it was indeed safer to be in hospital rather than recover at home.

Immediately the cardiologists demanded that the trial be stopped, calling Cochrane a 'murderous bastard', as patients at home were being harmed unnecessarily.[34] At this point Cochrane revealed that he had swapped the figures on the table around, and it was in fact the patients at home who showed the largest improvement. He inquired whether the doctors would like to close the trial immediately or wait for the final results. The doctors allowed the trial to continue.[35]

Theories vary as to how the broad acceptance of EBM has evolved over the past 60 years.[36] Some historians argue that EBM has existed for centuries: the James Lind Library has evidence from 1550 BC on how to reduce a dislocated mandible (lower jawbone).[37] Others dispute the idea that there is a genuine culture of EBM even now, citing recent claims that only 10 – 25% of practitioners' decisions can be reasonably called 'evidence based'.[38]

Despite this, it is reasonable to suggest that the concept of EBM as taught today in medical schools and incorporated into medical literature, has become far more prominent since the mid-20th century.

Perhaps this 'revolution' resulted from the bloody-mindedness of a few individuals, such as Cochrane.[39] David Sackett of McMaster University suggests that EBM took root because it empowered younger doctors, nurses and other practitioners to challenge their superiors.[40] He suggested that senior doctors who were secure in their knowledge were happy to be challenged, but the backlash against EBM was led by "middle-level guys who were used to making pronouncements."[41] Other political, sociological and technological factors might have also contributed. A sociologist suggested that the Second World War was crucial, as the model shifted from 'the lone scientist working in a modest laboratory to large work groups, [to] massive funding and the support of the establishment and industrial concerns.'[42]

The rise of the RCT
The first RCT was carried out by Sir Austin Bradford Hill in the UK in the 1940s, to evaluate an antibiotic for treating tuberculosis.[43] RCTs began to take precedence as the way to test medical interventions, and regulation of the pharmaceutical industry became stricter.[44] Though RCTs began in pharmaceuticals, clinical epidemiology also grew in importance.[45]

Evidence-based medicine emerges as standard

In 1967, the American physician and mathematician Alvan R. Feinstein published his groundbreaking work, "Clinical Judgment," which together with Cochrane's book "Effectiveness and Efficiency" (1972), increased acceptance during the 1970s and 1980s of controlled studies. The explicit scientific methods used to determine 'best evidence' were largely established by the McMaster University research group led by Sackett and Gordon Guyatt. Guyatt coined the term 'evidence-based' in 1990[46] and 'evidence-based medicine' appeared in the medical literature in 1992.[47]

In 1996, Sackett wrote what has become the standard definition of EBM: 'the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients'.[48]

Evidence as practice is now an established medical habit

Other significant steps were the creation of the Centre for Evidence-based Medicine (CEBM) at Oxford University, and of The Cochrane Collaboration, a network of researchers who create systematic reviews (SRs) of high-quality evidence (solely RCTs) to guide clinicians. Now 20 years old, The Cochrane Collaboration has produced over 5,000 SRs, and has over 100,000 researchers in over 120 countries.[49 50 51]

The UK's National Health Service (NHS) is regarded by some as the world leader in evidence-based practice, through its support of research and development and, for example, support for The Cochrane Collaboration.[52] The CEBM was one of the institutions that kick-started EBM. It may be that the UK provides a good example of the environment needed for EBM to flourish, as its universal health coverage enables comparisons, analysis and randomisation across virtually the entire population.

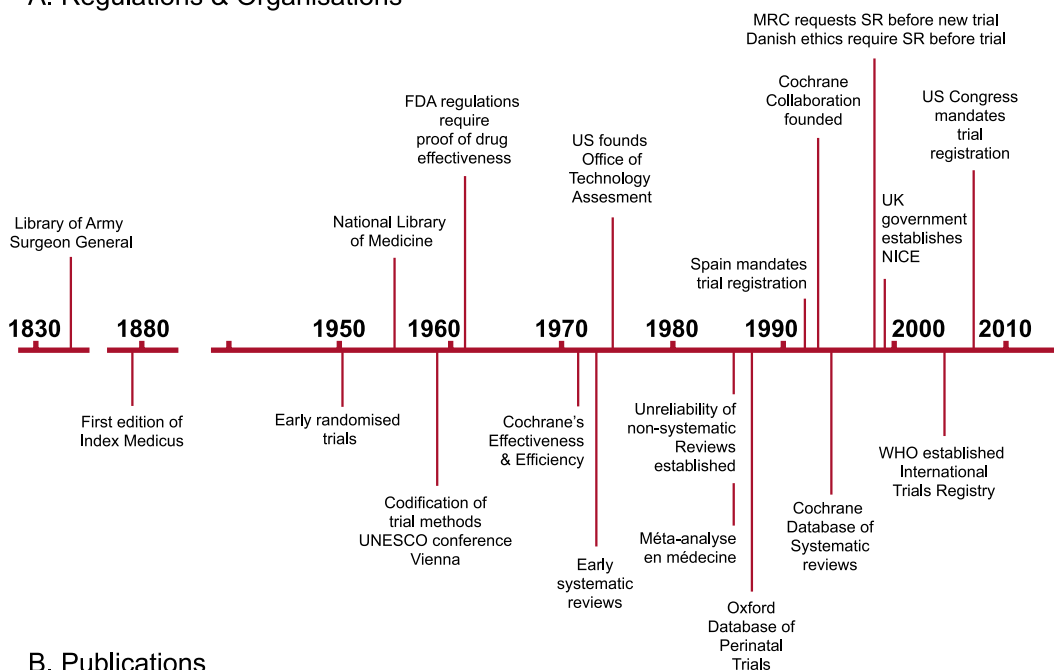### Information technology has become central to EBM

The ability to capture and large amounts of data on cheap desktop computers,[53] and compile medical records using database software,[54] allowed practitioners to draw from a vast number of medical records in a way impossible before the 20th century. More recently, the internet has offered practitioners access to a huge amount of data globally,[55] and powerful computers and electronic tools make analysis and meta-analysis of data practical and efficient.

Patients' behaviour and expectations are affected by their increased access to information, with many now taking a more proactive approach to their treatment. Given aging populations, the rising cost of healthcare and relatively small improvements in health in relation to expenditure, patients have become much more proactive in demanding evidence for a course of treatment.[56]

In 1997, Dr Muir Gray wrote in his book "Evidence-based Healthcare" that 'in the 21st century, the healthcare decision-maker, that is anyone who makes decisions about groups of patients or populations, will have to practice evidence-based decision-making. Every decision will have to be based on a systematic appraisal of the best evidence available. To accomplish this, the best available evidence relating to a particular decision must be found.'[57]

Figure 10: Policy and academic milestones in the development of trials and the science of reviewing trials[58]

A. Regulations & Organisations

One of the most influential contributions towards evidence-based practice in the US is the "User's Guide to the Medical Literature", published by The Evidence-Based Medicine Working Group in the *Journal of the American Medical Association*[59]. The guide was first published in the 1990s, designed to 'free the clinician from practicing medicine by rote, by guesswork, and by their variably-integrated experience [and]…use the literature as a tool to solve the patient's problems.'[60]

Remaining problems with the medical model

We are well aware that many problems persist with using evidence in medicine (see Appendix 3). They are so many and so profound that some senior doctors pronounce that 'medicine is broken'.[61] Hence, we do not claim that education should emulate medicine, though we do maintain that it offers valuable lessons. Perhaps as the evidence system in education develops, it can avoid some of medicine's problematic structures and biases.

Box 1: Evidence-based policing

Like education, policing has no long history of evidence-based practice and there is no structural intermingling of research and practice. Yet evidence-based policing (EBP) has been gaining ground recently and may offer lessons for education

Policing is similar to both health and education in its delivery: like health, it is about delivering 'interventions' to solve a problem. For example, should police 'treat' re-offending by (a) locking people up, (b) restorative justice (e.g., where the perpetrator meets the victim and hears about the effect on them), (c) using cognitive behavioural therapy or (d) increasing surveillance? The effects will be measurable and hence comparable. Policing also some-

---

times lends itself to randomisation as there are many similar 'units' (estates, hotspots, perpetrators, victims). However, unlike education, the outcomes for policing are pretty evident, in contrast to the disputed goals of education.[62]

In 1998 Professor Larry Sherman, the 'father of EBP', published a landmark article in the journal *National Institute of Justice*. In 2010 the Society of Evidence-Based Policing was founded to look at the production, dissemination and use of evidence in policing, which now has around 1300 members, largely police and researchers.

Some parts of the policing community are resistant to EBP, for much the same reasons as people are in education – they argue that trials take a long time and that they are difficult to carry out and interpret.[63] Some people suggest resistance arises from inertia, politics and EBP's 'newness' to many in the sector, as is the notion that policing is a 'craft' in which causal relationships can't be studied.

Despite this, the following steps have been taken to integrate evidence into policing:

- Extensive baseline data already exists in policing, collected by police forces and compiled by local, regional and national government organisations.[64]
- The UK College of Policing was recently designated the What Works Centre for Crime Reduction, to review research on policing and give an indication of its quality, cost and impact.[65]
- The UK College of Policing usefully translates evidence into 'guidelines' (although in its early stages), and journals and conferences disseminate evidence to practitioners.

Some Chiefs of Police have trialled new approaches. For example Alex Murray, Chief Superintendent of West Midlands Police (UK), is carrying out a trial with new offenders, who either go through the normal police system or go through a programme similar to Turning Point, where a good behaviour contract is agreed with the offender in lieu of a criminal sentence (the charges are withheld on condition of particular behaviours).[66] Each group is then measured against outcomes for prevalence and seriousness of offending.[67]

Research topics are generally prioritised based on the passion and interest of individual police officers and/or researchers. As well as that decentralised process, there is some centralisation: the UK College of Policing has a list of strategic research priorities, as do some police forces. The West Midlands police force holds match-making events for police with research questions to meet academics who can help to answer them.[68]

Barriers to the production and dissemination of evidence include the lack of funding for EBP, as well as public and media reluctance to accept policies (even in trials) that appear to be 'soft on crime', even if supported by good evidence.[69] Furthermore, some people create a false choice between reliance on science (seen as abstract) and common sense.

The International Association of Chiefs of Police in the US hosted a round-table on EBP in 2003, bringing together experts in academia and the police force.[70] The meeting agreed that collaboration between researchers and police works best when:

- 'The problem being researched is relevant to the law enforcement agency.
- Researchers and law enforcement leaders share responsibility and devote time and interest to the overall project.
- Researchers and law enforcement leaders are qualified to work in the partnership.
- Researchers offer practical recommendations to law enforcement leaders.
- Law enforcement leaders are in positions of power and can act upon the recommendations for operational change that resulted from research.'[71]

Other recommendations included:

- Selecting and supervising skilled researchers.
- Training police to evaluate and carry out research.
- Designing and formalising the research agenda and evaluating research results.
- Developing and sustaining relationships between police and researchers.
- Managing funding for research partnerships.[72]

Many of these recommendations can be readily applied to education. Significantly, few education researchers and practitioners have entered into a dialogue similar to the round-table described above.

5 Establishing goals in education

The World Health Organization (WHO) defines the goals for health as 'good health, responsiveness to the expectations of the population and fair financial contribution'.[73] There appears to be broad agreement on these goals between institutions, practitioners and policy-makers.

By contrast, **education has no widely-agreed goals**. Its goals are subject to extensive debate and disagreement. Remarkably, much focus in education is currently on inputs, despite there being little evidence that increased inputs lead to better outcomes.

Professor Karthik Muralidharan of the University of California, San Diego observed of the education system in India that 'Improvements in school quality as measured by inputs has improved considerably in the last decade…For instance, pupil-teacher ratios have fallen by nearly 20 per cent. However…inputs have not translated into improvements in learning outcomes, which **may be partly explained by the fact that education policy in the last decade has not prioritised learning outcomes.**'[74] (Emphasis is ours.)

The abandonment of learning as a goal is perhaps clearest in the cases where teachers are required to follow a curriculum irrespective of whether pupils are keeping up or learning anything. India's Right to Education Act obliges teachers to complete the curriculum every year, with no consideration or incentive to ensure that children are benefitting.[75]

5.1 Is consensus possible?

Literacy or numeracy are the most obvious and commonly-used measures of education systems. Yet these should be achieved by the end of primary school, and there is little agreement about what education beyond them is for.

The United Nations Educational, Scientific and Cultural Organization (UNESCO) developed the Education for All goals for 2015 recognising the need for more outcomes-focussed objectives.[76] However, even these predominantly deal with inputs, possibly because it is so difficult to agree on meaningful and realistic outcomes for education.

The Learning Metrics Task Force, based at The Brookings Institution, a not-for-profit public policy organisation based in Washington, DC, developed the indicators in Table 5 below, and, based on a wide consultation, recommended that they be tracked in all countries:

Table 5: Learning indicators for Global Tracking

Areas of measurement	Description of Indicators
Learning for All	Combine measures of completion and learning (reading proficiency at the end of primary school) into one indicator.
Age and Education Matter for Learning	Measure timely entry, progression, and completion of schooling, and population-based indicators to capture those who do not enter or leave school early.
Reading	Measure foundational skills by Grade 3 and proficiency by the end of primary school.
Numeracy	Measure basic skills by end of primary and proficiency by lower secondary school.
Ready to Learn	Measure acceptable levels of early learning and development across a subset of domains by the time a child enters primary school.
Citizen of the World	Measure among youth the demonstration of values and skills necessary for success in their communities, countries and the world.
Breadth of Learning Opportunities	Track exposure to learning opportunities across all seven domains of learning.

There have also been several wide consultations about education goals, the most prominent of which is the post-2015 global consultations. The UN System Task Team on the Post-2015 UN Development Agenda will manage the development of new post-2015 UN goals, and a set of 11 global thematic consultations and national consultations in over 60 countries is being conducted by the United Nations Development Group.[77] Prior to regional and global consultations, the 'World We Want' website was used to encourage broad engagement around the post-2015 goals.[78] To date, the consultations have highlighted some of the shortcomings of the MDGs and suggested alternatives. For example, because the MDGs focussed on national-level aggregated results, structural inequalities were overlooked. One solution to this is to disaggregate data by gender, ethnicity, religion, caste, age and geography.[79]

Clearly national governments, international donors, NGOs and practitioners and others vary in their priorities and systems, and often their goals for education. Hence little agreement has emerged to date. We have heard practitioners and researchers talk of various goals for education, which seem to fall into three broad categories – learning outcomes, labour market outcomes and social outcomes. We consider them in turn below.

5.1.1 Learning outcomes

Perhaps it's intrinsically valuable for a person to have facts and knowledge, but these will inevitably attenuate over time. Test results might be a good proximate outcome measure, indicating the likelihood of achieving longer-term goals, but this has yet to be demonstrated.[80]

> **'Education is what remains after one has forgotten what one has learned in school.'**
> *- Albert Einstein*[81]

In fact, the education system often behaves as though the goal were further education.

On the negative side, Dr Madhav Chavan, CEO of Pratham, believes that this is because schools were designed to feed the college system.[82] Hence they taught what colleges needed students to know, and exams were a quality control mechanism for the college system. Clearly if schools were designed to feed large employers or – as might be appropriate in India – provide skills for the many people who run small- and medium-sized enterprises, the curriculum might be quite different. This perhaps explains the marked discrepancy between what students are taught and what they need to know: for instance, in Europe, most teenagers are taught that copper sulphate is blue and goes white when heated, though clearly barely anybody needs to know that.

But on the positive side, for example the Harlem Children Zone (HCZ), a high school in New York City, considers its job to be getting students to complete a college education. This is not an end in itself but because HCZ believes that this opens the door to good jobs and lifelong success.[83] Hence it has moved away from simple learning outcomes and exam results. When it monitored former students, it found that many struggled with college life, particularly as many were the first in their family to go to college. In response, it created the College Success Office which supports students throughout higher education, including helping them to find funding and internships and manage their time well.[84] Geoff Canada, founder and Chief Executive of HCZ, says that 'the only benchmark of success is college graduation. That's the only one: how many kids you got in college, how many kids you got out.'[85]

5.1.2 Labour market outcomes

These are the effects of education on a person's future job and earning power. They include having a job, having a job with security and benefits (health, pension, holidays etc.), working in high-productivity sectors (e.g., services rather than agriculture), and lifetime income.[86]

Lifetime earnings: An education should increase a person's lifetime earnings, though it is difficult to measure this due to the length of study required.

Productivity: This is earnings per hour, and is important since clearly it is better to earn a reasonable wage in a reasonable number of hours, rather than by working endless hours in a lower-paid role.

Economic efficiency: Sometimes education has a signalling effect, used by employers in their selection. Business school education is an example, but secondary education can be too. On a macro level, this helps allocate people to jobs efficiently. Hence the value of an education establishment isn't just in the teaching but also in selecting students.

Studies of earnings and productivity are complicated for three reasons:
1. The labour market changes over time. For example, people in one cohort might get higher-paying jobs than their predecessors simply because well-paid high-tech jobs suddenly become available.[87] Hence it is hard to isolate the effect of the education.

2. Studies of labour outcomes need a range of timeframes, which impedes comparison between them. The effect of secondary education on labour market outcomes is easier to study than pre-school education because secondary students are nearer the labour market.

3. Labour market outcomes take a long time to materialise, making studies long and therefore expensive. But these problems are not insurmountable. For example, Professor Michael Kremer of Harvard University, who carried out an initial study into the impact of de-worming

primary school children on school attendance in Kenya, did a follow up study when the 'children' were 18-27 years old.[88] Another response is to use proximate outcomes, which have a causal relationship to desired long-term outcomes but manifest over a shorter timescale. For example, getting a degree and a well-paid job at age 23 might be proximate outcomes foretelling high lifetime earnings.[89] This technique is much more common in health than in education.

5.1.3 Social outcomes

Some commentators argue that a successful education system is defined by its social outcomes, such as citizens being less likely to commit crimes, less likely to need government support, and more likely to vote. Sometimes politicians have specific political goals for education: for example they might want to encourage people to think of themselves as, say, Kenyans rather than members of a particular tribe.[90]

While these outcomes indicate the huge role that education might play in the development of society, governance and political engagement, these outcomes have not hitherto been prioritised for research as they are difficult to measure.[91]

Emerging recommendation 1
Improve clarity about the goals of education (beyond basic literacy and numeracy). This will require hearing and considering the differing perspectives of stakeholders such as governments, employers, communities etc.

Emerging recommendation 2
Fund longer studies which track the effect of interventions right through to their intended effects (e.g., lifetime earnings). A possible alternative/adjunct is to explore using proximate outcomes to reduce the need for lengthy studies.

6 Production of evidence in health and education

> **'I solemnly affirm and believe, if a hundred or a thousand men of the same age, same temperament and habits, together with the same surroundings, were attacked at the same time by the same disease, that if one half followed the prescriptions of the doctors of the variety of those practising at the present day, and that the other half took no medicine but relied on Nature's instincts, I have no doubt as to which half would escape.'**
>
> *- Italian poet Francesco Petrarca, 1364[92]*

6.1 Types of evidence

There are essentially three types of evidence:

1. Qualitative evidence

This includes political economics, anthropology and sociology, and explores an intervention's context. For example it may detail the attitudes, beliefs and culture of a population served. Qualitative and contextual research is produced in both health and education, to understand why interventions work.

2. Baseline data

These provide a snapshot of the current situation, showing how large a problem is and where it is. Baseline data help policy-makers to identify and prioritise problems. They offer little insight into what has caused the situation, nor about how to improve it (though they are useful for case control studies: outlined below).

In health, they include data on incidence and mortality from particular conditions (e.g., how many people have flu and where they are). They may show the number and distribution of people with high blood pressure, they would not indicate what might ameliorate this.

In education, baseline data are collected on attendance, levels of numeracy and literacy, availability of equipment and so on. Pratham, an Indian NGO which created and runs the Annual Status of Education Results (ASER) (see box 7 on Aser) describes it as 'a big thermometer' measuring the 'educational temperature' at a point in time.[93]

3. Research into the effectiveness of interventions

This looks at how to improve conditions by testing putative solutions or interventions to a problem – to reduce the incidence of flu, or improve reading levels, for example. This research can be experimental (such as RCTs) or non-experimental (such as case control studies and analyses of historical patterns). It includes meta-analyses and SRs. It includes studies of individual interventions (e.g., the effect on teacher absence of putting cameras in classrooms) and comparing interventions (which might be through multi-arm trials, as illustrated in Figure 11 below). **There is much more research into the effectiveness of interventions in medicine than in education.**

6.2 The reliability of different types of evidence

The method matters

Studies of different types vary in their reliability. This matters because they can give quite different answers, some of which will be wrong.

For example, observational studies indicated that hormone replacement therapy reduces risk of death from heart attack. By contrast, randomised control trials (which are more rigorous and reliable) showed that it increases heart attacks by fully a third.[94]

The pattern holds elsewhere: in early evaluations, micro-loans to poor villagers in Northeastern Thailand looked pretty good. But these compared solely the outcomes (such as the amounts that households save, the time they spend working or the amount they spend on education) of people who took the loans with those of people who didn't. Those evaluations didn't take account of possible selection bias in the people who took the loans: perhaps only the richer people or better networked people wanted them or were allowed to have them. A careful study[95] which did correct for selection bias found that in fact the loans made no difference. The authors conclude that "naive' estimates significantly overestimate impact.'

Meta-research (research about research) finds that 80 percent of non-randomized studies (by far the most common type) turn out to be wrong.[96]

Hierarchy of evidence

The Centre for Evidence-based Medicine in Oxford University uses the hierarchy in Figure 11 to indicate the relative strengths of various types of evidence about causality:

**Figure 11:** The hierarchy of evidence[97]

The bottom two levels – **editorials, case studies and case reports** – are essentially observation and opinion. These are susceptible to various types of bias. For example, perhaps patients get better after seeing the doctor because they only see the doctor when at their lowest ebb after which nature heals them irrespective of whether they have seen a doctor.

Next come **case-control studies**, in which the background and behaviour of people who show a particular outcome (e.g., develop lung cancer) are compared to a control group who do not, in order to find factors which may be causes. For example, a case control study in 2002 compared the educational and employment history of teenage mothers in Belfast with those of a control group of teenage girls without children to see if there was a particular experience that might have contributed to the early pregnancies. The study found that teenage mothers were less likely than controls to have passed school examinations or be in full time education.[98] However, results of case-control studies may be unreliable because they are open to selection bias, and the data are by definition historic.

Cohort studies are longitudinal, tracking a group of people over time to see what factors correlate with certain outcomes e.g., the effect of class size on income levels. They are free of the problem of historic data because they follow groups of people for a long period. They are still open to selection bias (though can provide excellent data for examining 'natural experiments' such as the difference between children born just before and just after the Chernobyl disaster).

Better than all of these are **randomised control trials (RCTs)**, which reduce or eliminate selection bias. RCTs take a group of people and randomly divide them into a group which gets an intervention (e.g., books, a payment) and a group which doesn't (or which gets a different intervention). The sole difference between the groups is the intervention, which probably therefore accounts for different outcomes (e.g., reading levels).

However even well-conducted RCTs may produce conflicting results, and those results may be legitimate. A study on old, ill people may find a different result from one on young, healthy people; and any study is vulnerable to random effects, such as happening to select an atypical group of people. It's thus clear why the common notion that 'RCTs are the gold standard of evaluation' is unhelpful. A single RCT (or indeed single example of any other type of study) may be misleading by virtue of being done on an unusual population or getting a weird answer simply by statistical fluke.[99]

Figure 12: A multi-arm trial[100]
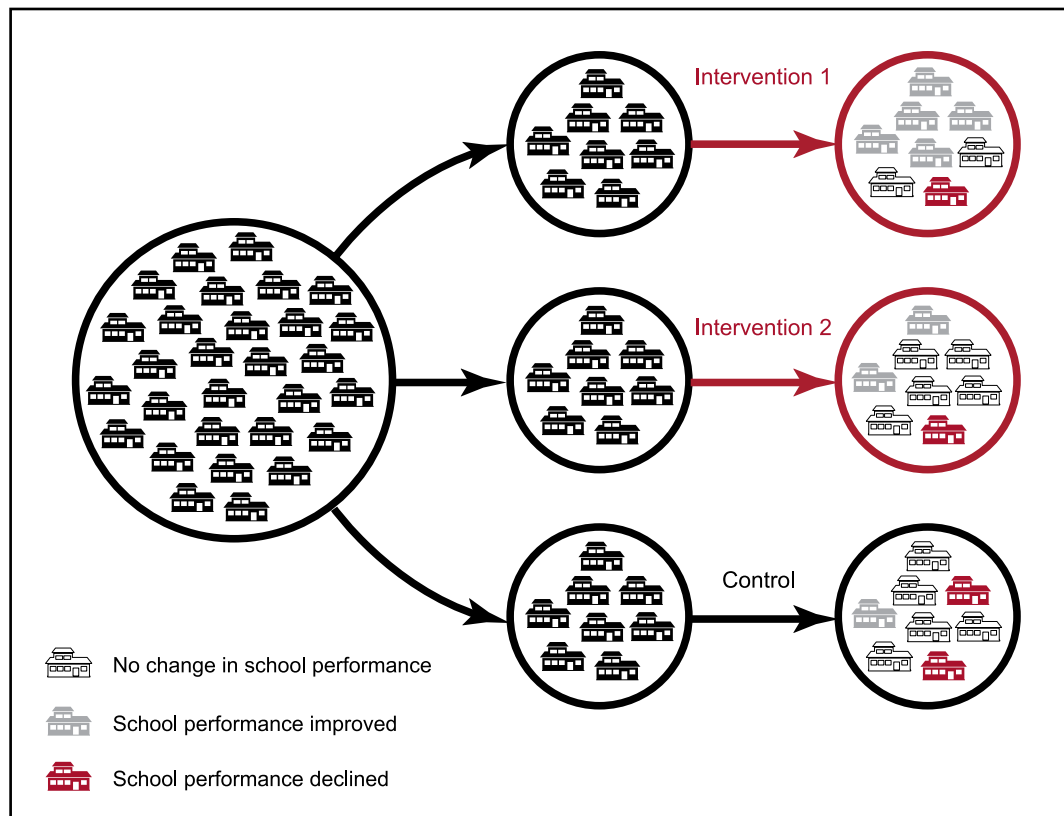
The best of all are **systematic reviews (SRs)** which synthesise the results of several studies in a particular area (including as many RCTs as are available) and therefore are least likely to be wrong. Meta-analyses are a subset of SRs which statistically combine RCTs as though combining their sample sizes.[101] They are very powerful, as they deal with conflicting results that are likely to arise from different RCTs. (They are also the most frequently-cited type of research in medicine.[102])

Many countries have a system for gathering information about the effects of drugs after they are released onto the market, such as adverse side-effects. The purpose of these systems is largely to catch problems (particularly serious ones) that may affect just a few people – too few to be captured pre-market RCTs because the experiments would need huge sample sizes to detect such effects. The UK's NHS has a Yellow Card scheme,[103] and the Center for Disease Control (CDC) in the US collects such reports. Sometimes data from these systems cause drugs to be removed from the market.

Box 2: Ethical concerns about RCTs

Some people object to RCTs by suggesting that it is unethical to withhold treatment or an intervention that has proved to be effective in the past from a control group. Indeed, if a treatment has been proven to be effective, this would be unethical.

We don't know

What if the impact of a treatment isn't known? Perhaps giving an untested intervention is unethical, since it may cause harm. Health has many examples of treatments considered to be effective until they were proven to be either ineffective or actively damaging. Blood-letting, leeches, trepanning and lobotomisation are some.

There are similar examples from social interventions. The Scared Straight programme sought to discourage young people in the US from committing crimes by taking them into prisons. For many years it was considered successful, but the evaluations failed to take into account that the children selected were already disproportionately unlikely to commit a crime. An RCT which corrected for this selection bias showed that the programme actually *increased* propensity to commit crimes.[104]

You have to choose somehow

Most organisations are oversubscribed, and hence must *somehow* choose who they serve. One benefit of choosing randomly is that this isolates the effects of their work, so they can see its effect.

Often, programmes are rolled out gradually: some regions get them before others. This makes it possible to compare performance in regions or locations that have it with those that don't. If we sequence the roll out randomly, then it's possible to do an RCT without withholding the intervention from anyone.

Another benefit is that randomising is 'fair' – perhaps fairer than other methods of selection, such as popularity or being first. Hence random allocation is often surprisingly popular with beneficiaries. IPA, which runs RCTs on many development interventions (and works closely with J-PAL), often uses a ballot in the village square. It finds that people often welcome a public and transparent process, rather than resources being allocated based on people's connections.

Comparing something with something, not something with nothing

RCTs don't always compare a programme with receiving nothing. Many RCTs compare interventions, or compare variations of an intervention (as per Fig 12). If we don't know which intervention is best, nobody loses out.

Dealing with ethical issues

Professor Chris Whitty, a clinical epidemiologist who is currently DFID's Chief Scientific Advisor and Director of Research and Evidence, emphasises that there will always be ethical concerns with carrying out experiments that involve people, but that the way to deal with them isn't to reject experimentation, but to establish a robust system to manage these concerns. Ethical committees are involved in most areas of health research, and development economists must also secure approval from ethics committees in universities before beginning trials.

6.3 Health: What evidence is produced and by whom?

6.3.1 Baseline data

In health, significant baseline data are collected on local, national and international levels. For example, the US National Health and Nutrition Examination Survey carried out by the National Center for Health Statistics (part of the Center for Disease Control, CDC), began in the 1960s to monitor the health and nutritional status of adults and children in the US. It examines a nationally-representative sample of 5,000 people each year, with 15 locations revisited each year.[105] Internationally, the WHO collects and publishes baseline data from its 194 member states.[106]

6.3.2 Intervention effectiveness research

6.4.2.1 RCTs

RCTs are well-established and uncontroversial in medicine. Tens of thousands are produced each year: over 26,000 in 2008, up from just 39 in 1965.[107] Carl Heneghan, Director of the Centre for Evidence-based Medicine at Oxford University, suggests that 'at current rates we can expect to see 50,000 RCTs published per year by 2018' (see Fig 13).[108]

Almost all clinical trials are RCTs. They are heavily regulated and are mainly funded by pharmaceutical companies, though may be conducted by academics. RCTs are also used in medicine to test many other activities, including alternative ways of teaching medical students, and methods of disseminating evidence to doctors, e.g., comparing the effect of a five minute podcast or a two page leaflet on prescribing behaviour.

> ## Box 3: An RCT, sir?
>
> This tale perhaps illustrates how experimentation is embedded in medical culture, and how RCTs are not invariably complicated, expensive or slow.
>
> When Dr Iain Chalmers, Coordinator of the James Lind Alliance, was given a knighthood for founding The Cochrane Collaboration, he wondered whether it would make people pay him more attention. So his outgoing letters were, by random allocation, signed either 'Iain Chalmers' or 'Sir Iain Chalmers'. The differing response rates are described in a paper tellingly called 'Yes sir, no sir, not much difference sir'.[109]
>
> Perhaps this tale in fact illustrates external validity. Perhaps a knighthood made no difference to response rates because everybody knew who Chalmers was: it might have made a difference to other people.

**Figure 13:** Number of RCTs produced per year based on PubMed Trend Data[110]
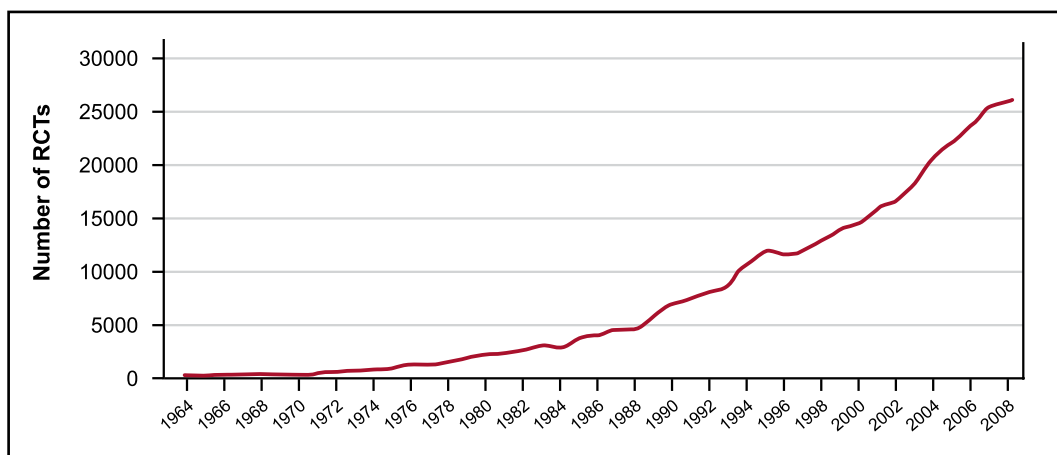
6.4.2.2 Systematic reviews

Systematic reviews synthesise evidence on a particular topic, such as vaccines to prevent flu.[111] Though one can 'systematically review' any type of evidence, medical SRs almost exclusively consider RCTs (the largest producer of SRs, The Cochrane Collaboration, only considers RCTs). Since SRs rely on having several RCTs, there are few in education as few RCTs exist. SRs and meta-analyses are normally conducted by academics.

Like RCTs, SRs can be done well or badly. The Measurement Tool to Assess Systematic Reviews (AMSTAR) rating system is a new tool to allow doctors to assess the methodological quality of SRs, which provides an additional level of analysis into evidence production.[112]

Figure 14: Number of systematic reviews produced per year in healthcare, 1990 to 2007[113]
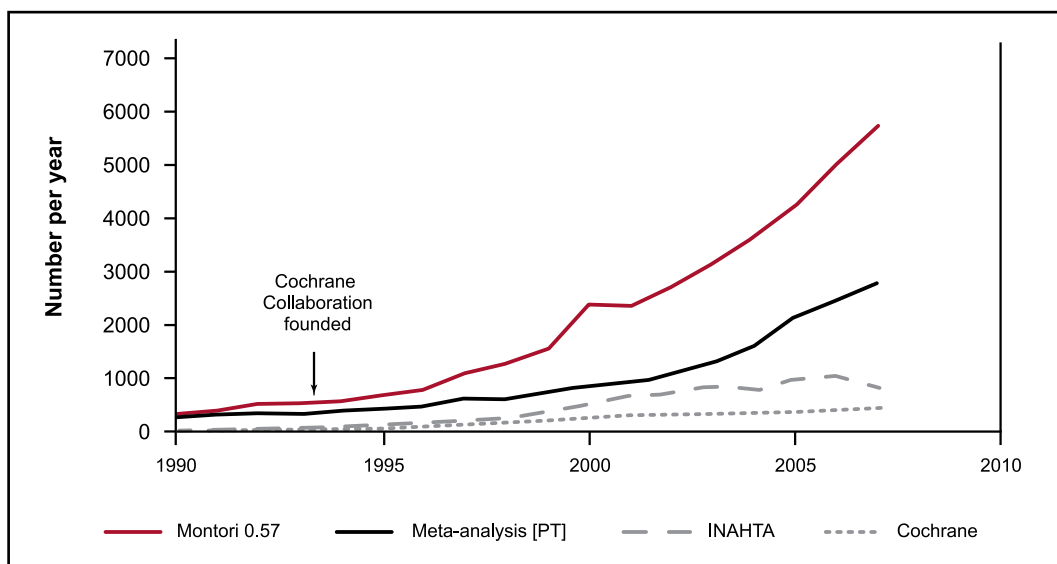
6.3.3 Research into the medical system and its evidence base

In health, the evidence system itself is researched, including the quality of evidence produced. For example, studies in health have examined publication bias (finding it rife),[114] the effect of various methods of randomisation in RCTs,[115] and the extent to which supposedly independent journals are more willing to publish papers sponsored by pharmaceutical companies than by independent academics.[116]

Meta-research (research about research) includes assessing the quality of research. One meta-research study found that for half of all recommended treatments in UK cardiology guidelines, the evidence is only 'Grade C'.[117] There are studies of the effect of trials which compare new interventions against nothing or an alternative known to be ineffective (which is common practice in trials funded the by pharmaceutical industry): an editorial based on such a study was bluntly entitled 'Blood on our hands: Seeing the evil in inappropriate comparators'.[118]

Teaching EBM is researched. For example, an SR looked into whether trainee doctors should study EBM in a standalone course or integrated into clinical practice.[119] It found that standalone teaching improved just knowledge, whereas the integrated teaching also improved skills, attitudes and behaviour.[120] It is striking that in health there are sufficiently many RCTs on the evidence system to conduct such an SR, in contrast to education.

We have found very little meta-research in education, and few studies of the quality of education research. We do not know of any meta-research into evidence produced by NGOs or other non-academics such as impact reports or monitoring and evaluation reports, despite them been collectively quite expensive. Without such quality control it is hard to know whether this evidence should be used.

> **Emerging recommendation 3:**
> Fund work to assess (and thence improve) the quality of education research by non-academics. (This will require making the research more accessible and clearer about research methods used: see recommendations below.)

6.4 Health: What evidence isn't produced and why not?

Medical devices and health management systems are much less researched than are clinical interventions.[121] Possible reasons for this include:

- Financial incentives and regulatory requirements: the pharmaceutical industry has obvious commercial and regulatory incentives to test its interventions. Incentives and requirements on devices and management are weaker.
- It is difficult to experiment with management options because 'units' (hospitals or governments) are too few and diverse to enable a robust sample.
- Medical researchers are often more interested in biochemistry than in management.
- Perhaps it is harder for research into devices and management to get published, though we have not explored this (see Box 4 on academic incentives).

Box 4: Academic incentives

One of the most significant drivers of producing and disseminating evidence is incentives for academics to publish research.

In health, academics and clinicians are incentivised to publish in prestigious journals (such as the *British Medical Journal (BMJ)* or *The Lancet*) to advance in their careers. In education, researchers are incentivised to publish in academic journals, despite no journal reliably reaching practitioners or policy-makers.

Prestigious journals are generally interested in publishing only the first, last and 'blockbuster' findings, rather than replication studies that repeat experiments in other circumstances.[122] For instance, the first study of de-worming children may be published, but a study of the same intervention in a diffe rent place won't be. The 'last' study might bring together numerous studies from many geographies and conclusively prove the case for or against an intervention.

Replication studies are vital, partly due to external validity (policy-makers need to know that an intervention will work *here*, not just that it worked *somewhere*), and partly because science is fundamentally about repeatable results: nobody's impressed if you got cold fusion in your bathroom last Tuesday if you can't do it again. Hence journals' uninterest in replication studies, which hinders their production, is a major problem.[123] [124] We look in more detail at how journals affect disseminating evidence in the next section of the report.

Economic journals seem to have a marked skew towards studies of the United States. A study of 76,000 articles published between 1985 and 2005 revealed that there were 'more papers focussed on the US than on Europe, Asia, Latin America the Middle East and Africa combined…The world's poorest countries are effectively ignored by the profession. From 1985 to 2005 Burundi was the subject of just four papers. The *American Economic Review*, the holy grail for many academics, published one paper on India, by some measures the world's third-largest economy, every two years.'[125] Clearly this is likely to affect the choice of topics studied by economists, for whom publication is imperative to advancement.

Furthermore, academic journals are interested in theory and experiments which advance or refute theory, rather than which might be useful in practice. Hence articles are not written to be easily understood or used by practitioners. Since academics are incentivised to gain tenure at a university, they have little incentive to study or work on implementation or translating research to help policy-makers or practitioners. However, some academics are willing to do this, such as Dean Karlan of Yale University, who is President of Innovations for Poverty Action (IPA), a not-for-profit using evidence-based interventions to address poverty.

It is reasonable to suppose that these incentives have two negative effects. First, that some evidence that is produced is not published, and second, research that could have been pursued is not investigated because it is unlikely to be picked up by journals.

In education, some evidence-producers such as IPA deliberately employ non-academics who can lead research precisely because they are not subject to the same incentives.

6.5 Health: How are research topics prioritised?

Possibly researchers just follow their peers when prioritising research topics. Daniel Kahneman, Nobel Prize-winning psychologist, notes this in other disciplines: 'the main reason that decision theorists study simple gambles is that this is what other decision theorists do'.[126]

Chalmers has found that researchers rarely look at systematic reviews before deciding what to study, leading to extensive duplication of research.[127] A study by Chalmers and Glasziou found that 'only 11 of 24 responding authors of trial reports that had been added to existing systematic reviews were even aware of the relevant reviews when they designed their new studies.'[128] Only *The Lancet* requires new research to be set in the context of systematic reviews into other relevant research.[129]

Duplication of trials delays the delivery of effective treatments and wastes research resources.[130] This was shown in Chalmers' study to be true of treatments for early breast cancer and in the (unnecessary) repeated demonstrations of the benefit of preventative antibiotics before some surgeries.[131]

Duplication of trials is exacerbated by a lack of transparency about studies currently underway. Although researchers are supposed to register trials when they begin, this is not always done, in part due to the concern that the idea will be poached before publication.[132] While some journals refuse to publish studies that aren't pre-registered, many still do.[133]

Chalmers' and Glasziou's study also showed that the public funding of research 'correlated only modestly with disease burden'[134] and patient needs; for example, in a survey of patients suffering from osteoarthritis of the knee, only 9 per cent of patients wanted more research into drug interventions, but in practice over 80 per cent of the RCTs carried out into interventions were drug evaluations.[135] Globally about 90 per cent of health funding goes to just 10 per cent of the disease burden.[136]

The James Lind Alliance was created (by Iain Chalmers) to improve alignment between patients' needs and researchers' priorities.

<div style="background:#e0e0e0;">

Box 5: The James Lind Alliance and patient involvement

The James Lind Alliance (JLA) is a UK not-for-profit organisation which seeks to ensure that medical research prioritises issues which concern patients – rather than those which concern academics or journals or nobody. Its Priority Setting Partnerships (PSPs) brings 'patients, carers and clinicians together to identify and prioritise the top 10 uncertainties, or 'unanswered questions', about the effects of treatments that they agree are most important.'[137]

PSPs are often initiated by patient groups, such as asthma or Parkinson's patients,[138] or by doctors or hospitals interested in an area underrepresented in health research. Acne is an example. In asthma, the JLA found that patients were keen to know if there is value in the breathing exercises they're asked to do, which researchers had overlooked.[139]

Coordinated by Chalmers, the JLA is still a tiny organisation, although it has recently been integrated into the NHS through the National Institute for Health Research Evaluation, Trials and Studies Coordinating Centre.[140] The UK Department of Health is now also beginning to require evidence of patient involvement in the design of research it funds.[141]

</div>

6.6 Qualitative research

Many quantitative studies in health involve qualitative research, to identify a problem or potential solution and understand context.[142] Health people recount how there used to be a 'tribal division' between quantitative and qualitative researchers (as there seems to be in education) and strong progress has been made by having the two groups collaborate.[143] For example, surgical gynaecologists Professor Arri Coomarasamy and Dr David Lissauer at Birmingham Women's Hospital are investigating the effectiveness of using pre-surgery prophylactic antibiotics to reduce infection in women having miscarriage surgery in Blantyre, Malawi.[144] Their RCT involves a medical anthropologist to understand the women's attitudes to surgery, antibiotics and miscarriage, to gain a deeper understanding of the potential barriers to wider implementation outside the trial. If women believe antibiotics to be useless or dangerous they are less likely to take them when a researcher is not present, which would make it difficult to extend the intervention from a small trial setting to a larger community or group.[145]

Cooperation between the 'tribes' has been driven in part by funders: the UK Medical Research Council (MRC), Wellcome Trust, DFID and the National Institute for Health Research (NIHR) now often require it.[146] Nonetheless, a barrier is that qualitative and quantitative research tends to be published in different journals, and journal requirements often influence researcher behaviour (see Box 4).[147]

6.7 Education: What evidence is produced and by whom?

'Convinced of the need to help people – perhaps on emotional rather than intellectual grounds – they feel social work is worthwhile whatever the chances of 'success' or 'improvement'. Not depending for its justification on its results; no failure can therefore discredit.'

- Paul Halmos, mathematician, 1965

There is scarce evidence about the effectiveness of interventions to improve outcomes. Even ubiquitous interventions, such as blackboards, have been studied only a handful of times. (See appendix 5). Whitty reports that most evidence in education is 'a description of the problem, in increasing detail, but not about testable solutions'.[148] Professor Roland Fryer, a Harvard economist, Clark Medal winner and founder of Harvard's EdLabs, concurs: 'People [in education] spend a lot of time trying to tease out causal relationships from observational data. Good luck with that'.[149]

6.7.1 Baseline data

Baseline data in education covers, for example, learning levels, school access, participation, and availability of resources such as teachers and schools.[150]

It is very patchy. Many countries still have no baseline data at all. According to the Education and Policy Data Center in Washington, DC: '**In any given year since 1999, national-level data on the number of out of school children are not available for nearly 40% of the countries listed in the UIS Data Centre**'. (The UIS is the UNESCO Institute for Statistics.) The report continues: 'A review of available [UIS] data shows a considerable amount of missing information, particularly for countries where the number of out of school children could potentially be quite high, given their recent history (i.e., Sierra Leone, post-secession Sudan and South Sudan, Haiti). For some countries, such as Bangladesh and the Democratic Republic of the Congo, the figures factored into the global estimate are not published, and the most recent available UIS figures are more than a decade old (1990-1995).'[151]

Who produces baseline data?

No single institution in education replicates the WHO in producing baseline data at a global level, although various organisations produce some baseline data. The UIS compiles baseline data from national governments in member countries, which are then published by the World Bank.[152] The UIS collects data on more than 30 indictors, generally input-focused, including:

- Children out of school
- Expenditure per student
- Youth literacy rate
- Primary completion rate
- Progression to secondary school
- Public spending on education
- Pupil-teacher ratio
- School enrolment
- Trained teachers in primary education[153]

UNESCO also produces the annual Education for All Global Monitoring Report which summarises progression towards the Education for All goals, as well as some broader topics in education research. The report includes contributions from a range of multilateral and bilateral donors.[154]

The Global Partnership for Education (GPE) brings together donor countries, private donors, NGOs and other groups to provide finance, strategic project support and monitoring to LEDCs to help them achieve a range of education goals. It also collates some baseline data.[155] The World Bank also reports annually on absenteeism in several public services (see Box 6 below).[156]

> ## Box 6: Missing in action: teacher and health worker absence in developing countries[157]
>
> An absenteeism study published in 2006 and led by researchers at the World Bank and Harvard University – based on unannounced visits to primary schools and health clinics in Bangladesh, Ecuador, India, Indonesia, Peru and Uganda – found that on average **35 per cent of health workers and approximately 19 per cent of teachers failed to show up for work**.
>
> The World Bank reported that 'beyond the widespread overall absence rates, provider absence was worse in poorer regions, and that higher-ranking and more powerful providers such as headmasters and doctors were absent more often than lower-ranking ones.'[158]

Other organisations gathering baseline data include:

- The **Programme for International Student Development (PISA)**, assesses reading, mathematics and science competencies of 15 year olds, as well as cross-curricular competences such as problem solving.[159] Over 70 countries have participated in PISA, which is coordinated by the Organisation for Economic Co-operation and Development (OECD). PISA has run every three years since 2000.

- The **Annual Status of Education Report (ASER)**, run by the NGO Pratham, provides 'reliable estimates of children's enrolment and basic learning levels for each district and state' in India.[160] It has been run every year since 2005, and other organisations are emulating it in other countries (See below).

- The **Trends in International Mathematics and Science Study (TIMSS)** has been conducted every four years since 1995. It measures learning levels in mathematics and science by children aged 8-10 years and 13-14 years (fourth and eighth grades) in more than 67 countries.[161]

- The **Progress in International Reading Literacy Study (PIRLS)** has measured reading comprehension by children aged 8-10 years (fourth grade) every five years since 2001.[162] It is collected by the International Association for the Evaluation of Educational Achievement (IEA), an 'independent, international cooperative of national research institutions and governmental research agencies'.[163]

- **Uwezo**, which measures competencies in literacy and numeracy among children aged 6-16 years old in Kenya, Tanzania and Uganda.[164] Similar initiatives also exist in Mali and Senegal.[165]

Box 7: The Annual Status of Education Report (ASER)

Invented and run by Pratham in India to inform its own education work, ASER is a huge operation, involving almost 30,000 volunteers. Unusually, they go house-to-house to conduct the survey, to include children who are not in school, as well as looking at behaviour and equipment in schools.

In each district, 30 villages are sampled. In each village, 20 randomly selected households are surveyed. This process touches approximately 700,000 children aged 3-16 years, in about 300,000 households nationally.[166] Remarkably, it only costs about $1 million to deliver.

ASER is now being used (in various forms) in Kenya, Uganda, Malawi, Senegal, Mexico, Mali and Pakistan, and Pratham is working on introducing it more widely still.

The UN Secretary General has called for a 'data revolution' in post-2015 development goals (which will replace the existing Millennium Development Goals). This is a particular challenge for education given the paucity of data, and highlights the importance of ASER-like baseline data in education.

Emerging recommendation 4:
Increase funding for producing better and more comprehensive baseline data in education in LEDCs.

6.7.2 Intervention effectiveness research

'Remarkably little is known about the impact of education policies on student outcomes in developing countries' [167] according to the authors of a systematic review of all research in the 20 years from 1990-2010 into the effectiveness of educational interventions. They found only 9,000 trials in total (as mentioned, health is now producing about three times that each year). Only 79 were considered to be of 'sufficient quality' to be included, and only 43 used a high quality statistical methodology.[168] †

It is shocking how little such ubiquitous interventions have been studied. Only three high quality studies examined the effect of blackboards on test scores, and no studies looked at the effect of teacher absenteeism or school meals on students' time in school, despite these being very popular interventions. The effect of building new schools – much beloved of donors – has only been studied three times.[169]

It is shocking how little such ubiquitous interventions have been studied. Only three high quality studies examined the effect of blackboards on test scores, and no studies looked at the effect of teacher absenteeism or school meals on students' time in school, despite these being very popular interventions. The effect of building new schools – much beloved of donors – has only been studied three times.[170]

---

† 'High quality' study design was defined as using difference in differences regression, regression discontinuity design or an RCT. 'Sufficient quality' also included studies with 'at least one general family background variable (e.g., parental schooling or household income) and school expenditure per pupil, or one family background variable, one teacher variable, and at least one additional school variable'

Professor Mike Clarke, former Director of the UK Cochrane Centre and former Co-Chair of the international Steering Group for The Cochrane Collaboration, suggests that this may be due to a lack of a culture of trials in the education sector.[171]

Box 8: How can we improve the evaluation of NGO-run interventions?

Most education funders require grantees to assess their work in some way. These processes are diverse, normally known collectively as 'monitoring and evaluation'. Since the choice of research method can dramatically influence the result, it is important that donors and practitioners rely only on evaluations which are reliable. However, unlike medicine, there are no agreed quality standards for research in education.

Consequently evaluations by NGOs in education (and elsewhere) vary in their rigour, **and may lead practitioners to bad decisions**. Indeed the sole study of the quality of charities' evaluations (which we have seen) is by the Paul Hamlyn Foundation, and found that 70 per cent of material it received from grantees was below what it considered to be 'good' quality.[172]

There are several – solvable – problems with research by NGOs and other non-academics. First, much isn't published. And much is published on organisations' own website which are hard to find. It could be more findable, given a central, easy-to-use repository and the will / incentives to publish. [The World Bank discovered that most documents published as PDFs on its website are barely read.[173] Most NGOs' M&E reports are published as PDFs on websites much less famous than the World Bank's.]

Second, research quality is variable. This is partly because the implementing organisations have an incentive to produce positive evaluations because this helps secure further funding. It's also partly because few implementing organisations are skilled in research methods. These factors combine dangerously because weak research methods tend to be better for supporting strong claims.[174] It is therefore perhaps fair to be suspicious about the quality of NGOs' evaluations. A first step to solving this is to make the research findable and clear – about the intervention used and the research method. This will enable meta-research to see where the problems are and to target support to improve it.

Thirdly, NGOs may be doing monitoring and evaluation as compliance with grant conditions, rather than to learn. As Clara Miller, President of FB Heron Foundation observes, 'Most philanthropy is a culture of bureaucracy, not a culture of discovery'.[175]

Hence our recommendations 5 and 6 below to make non-academic research more findable, and to show its quality such that work to improve quality can be targeted well.

Emerging recommendation 5:
Improve the ease of finding, assessing and using research by NGOs and other non-academics (including monitoring and evaluation). As a first step, create a repository for education research by NGOs, practitioners and any other entity, with requirements for clear description of the intervention used, the research method used, how it was used (e.g., how the sample was chosen), and the results. Funders could require entities they fund which conduct any monitoring and evaluation or other research to post it to the repository.

Emerging recommendation 6:
Improve the quality of research by NGOs and other non-academics. Recommendation 5 is are necessary for this, but it will also require training those research-producers, training funders, and increasing partnership between them and professional researchers.

6.7.2.1 RCTs

Production of RCTs into education is dominated by the Abdul Latif Jameel Poverty Action Lab (J-PAL). J-PAL is a network of development economists in many universities who use RCTs to explore interventions in various aspects of international development, including education, health, water, governance, the environment and agriculture. (Technically, studies are run by academics affiliated to J-PAL, rather than J-PAL itself.) J-PAL's affiliates have completed 60 RCTs in education, and have another 30 under way.[176]

Other major producers of RCTs and other evidence about intervention effectiveness include:

- The World Bank, which bases its research on national governments' priorities, along with its own research interests.[177]

- The Center for Global Development, an NGO in the US and UK, which has 47 published working papers, policy papers and other material on education.[178]

- The Centre for the Evaluation of Development Policy at the Institute of Fiscal Studies

- FHI 360, a not-for-profit based in the US that produces evidence into the effectiveness of interventions, particularly those funded by USAID. It also hosts the Education Policy and Data Center, a research and analysis unit that compiles the results of a range of research studies.[179]

There is a striking mismatch in the timescale of many RCTs and the stated goals of education, which are often around labour market outcomes and social stability. Most evaluate small scale interventions, and/or run for a short time period[180] and hence focus on near-term outcomes such as test scores.[181] This is not a criticism of the RCT method, but rather a feature of the funding of studies versus the reality of achieving educational outcomes.

Box 9: Never mind the interventions: it's great teachers we need to find

STIR (Schools and Teachers Innovating for Results) was founded on the observation that teachers are the biggest in-school determinant of a child's academic success. It sought to discover promising 'micro-innovations': low- or zero-cost interventions being used in classrooms, and to scale them up. Its model was as you'd expect: build a network to find these micro-innovations; vet them; rigorously test the most promising ones; and disseminate those which prove effective to schools, NGOs and government. Starting in India and extending to Uganda, it has had some success, such as finding Bindu's use of Bollywood songs to teach poetry or Jasbir's student letterbox to improve literacy.

STIR's model – of discovering successful interventions and scaling / advertising them – is intuitive and common. It's used in medicine (where the dissemination mechanisms are good, as discussed) and by IPA, J-PAL and others in international development. Yet STIR found that it was insufficient. Teacher absenteeism is rife; when teachers do show up, less than half their time is spent teaching; and few are motivated and interested in teaching. The problem, then, isn't exposing teachers to great interventions; it's getting them to actually do their job.

STIR found that the teachers who submit ideas are unusually motivated outliers. STIR now views its role as finding effective *teachers*, rather than effective *interventions*, and supporting them to influence those around them. It convenes these outlier teachers in 'teacher change-maker networks'. The teachers collaborate to implement new practices, and influence other teachers. Once these changemakers have experienced leading change, STIR connects them to a wide ecosystem of partners who can enable them to achieve more. These changemaker teachers are paid nothing for this extra work.

Each changemaker may be the sole teacher in their school actually interested in teaching, surrounded by colleagues who prefer to chat and drink tea. Perhaps the strongest effect of these networks is to normalise their ambition about teaching and learning. It also creates a cadre of teachers *who actually teach* who can be vectors for spreading effective interventions.

Thus STIR doesn't (just) find seeds which grow well, but rather nourishes the soil into which those seeds are planted. It's too early to tell whether STIR's model will work. But it stands to have profound influence because it challenges the scientific approach to development, that just discovering and publicising effective interventions will suffice.

6.8 Education: How are research topics prioritised?

'Of the 55 questions that we could research, how do we know that this one [in a research proposal] is a priority? Often we don't', complains a significant research funder.[182]

A criticism of education research is the lack of coherence between topics or geographies – witness the array of topics researched recently by J-PAL.

Figure 15: Recent education RCTs completed by J-PAL[183]

- Cash transfers for education in Morocco.
- Evaluating the effectiveness of online sexual education modules in Colombia.
- Evaluating the impact of moving to opportunity in the United States.
- Conditional community block grants in Indonesia.
- Boarding schools for secondary students from disadvantaged backgrounds in France.
- Textbooks and test scores in Kenya.
- Subsidizing remittances for education in El Salvador.
- A 20-year follow-up to an early childhood stimulation programme in Jamaica.
- The effects of education on voter turnout in the United States.
- Using radio and video as a means for financial education in Peru.

It's not clear that research topics are prioritised based on what practitioners need, rather than the interests and incentives of academics.[184] There is some 'match-making' between researchers and practitioners (the users of evidence) e.g., by J-PAL, though there seems to be rather limited involvement of policy-makers and practitioners in prioritising topics. A non-academic researcher who has attended some of these events reported that they weren't really consultation exercises. Rather, they were gathered researchers who were looking for NGOs who would be prepared to run a programme using their ideas to enable an evaluation. "Like recruiting lab-rats?", we asked. "Exactly" he replied.[185]

Figure 16: Not all academically-interesting questions are relevant to policy
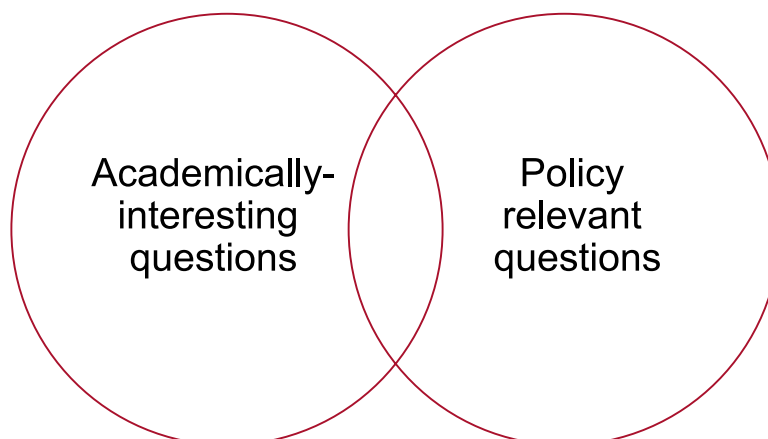
Academics are incentivised to look at questions that are academically interesting, which are not always policy-relevant. This seems to play out in practice: we heard several stories in which funders had been asked to fund studies for/on the NGOs that they support, with the academics conducting

the studies attempting to insert questions that are interesting to them, but not to the NGO or the funder (and for which the funder is asked to pay).

To our knowledge, there is no equivalent to the James Lind Alliance in education to ensure that education research is useful to practitioners and policy-makers. We found little evidence that the research spend matches the implementation spend in education.[186]

The domination by J-PAL of the RCT 'market' was raised by several funders and actors. One complaint was about J-PAL's 'initiatives', wherein it raises money for research in a particular topic (e.g., post-primary education) and invites research proposals. We heard complaints that this process is opaque and only open to J-PAL's affiliated academics.

The Brooking Institution's Center for Universal Education (CUE) has convened the Global Compact on Learning Research Task Force (which differs from the similarly-named but operationally distinct Learning Metrics Task Force) to help shape the research priorities in education.[187] Chaired by Dan Wagner from the University of Pennsylvania and comprising 16 experts from across the world, the Task Force aimed to develop a list of priority research topics, based on questions that remain unanswered related to learning in LEDCs. The Task Force acknowledges that while a focused research agenda is an 'urgent task', it is also 'complex and challenging'. 'Efforts to build this evidence base need to be coordinated, systematic, and sustained over time and more resources are needed to support this effort'.[188] (Interestingly, while this research was underway, Julie Nelson, Research Director of the (UK) National Foundation for Educational Research, wrote that '(UK) schools' needs and interests should have greater influence on what research gets commissioned.'[189])

> **Emerging recommendation 7:**
>
> Develop a mechanism in education similar to the James Lind Alliance, which could prioritise research topics with input from practitioners, policy-makers and other users of evidence.

6.8.1 Pre-registering research

Enabling - and obliging - researchers to pre-register trials is important for several reasons. First, it reduces the chance of two researchers investigating the same question, which obviously wastes resources. Without pre-registration, this is quite possible because researchers can only 'see' research once it is published, i.e., after it is finished, but not research underway is invisible. Second, it increases integrity and reliability because researchers pre-register the questions they will answer and hence cannot mine the data for findings. (Almost any data set will yield some correlations, just by random chance, and the purpose of research is not to *find* random correlations but to *exclude* them as possible causes.) And third, it prevents publication bias in which only favourable findings are published. This is a major (and fatal) problem in medicine, largely because of the commercial interests of pharmaceutical companies. Though it appears to be less dangerous in education, it should not be ignored.

Various registries exist for researchers to pre-register research. For example, the International Initiative for Impact Evaluation (3ie) has a registry on which any impact evaluation (i.e., which looks at causation) can be posted – not just RCTs or work by academics or economists. However, researchers rarely pre-register research – particularly research done by NGOs. Funders (both charitable foundations and academic research funders) could be highly influential here by making pre-registration a condition of funding.

6.8.2 Systematic reviews

Far fewer SRs are produced in education than in health, for two reasons. First, in most topics in education, there simply isn't enough evidence to synthesise. Second, SRs can only combine studies which look at the same intervention in similar contexts. This is rarer in education, since education interventions tend to be more context-specific.

SRs in education are produced by: The Campbell Collaboration; the Evidence for Policy and Practice Information and Co-ordinating Centre (EPPI-Centre) in the UK; 3ie; UN agencies (both in-house and outsourced to other research institutes); and academics, such as Patrick McEwan at Wellesley College.[190]

The Campbell Collaboration is an international network of researchers that produces SRs on the effects of social interventions, including crime and justice, education, international development and social welfare.[191] The Collaboration was founded in 2000 by researchers interested in social interventions, many of whom were involved with The Cochrane Collaboration.[192] However, whereas The Cochrane Collaboration has produced thousands of SRs, the Campbell Collaboration lists only **14 SRs on education interventions.**[193]

The EPPI-Centre is part of the Social Science Research Unit at the Institute of Education, University of London. It has produced 178 SRs since 1993 on topics including education and social policy, health promotion and public health, international health systems and development, and participative research and policy.[194] The EPPI-Centre's SR database contains **15 SRs on education in LEDCs**.[195]

3ie funds impact evaluations and SRs about international development interventions.[196] To date, it has funded 20 SRs, and its website holds over 200 SRs on the effectiveness of interventions in low and middle income countries. Currently only **12 SRs on education** are listed, again a miniscule number compared with those in health.

6.9 Two 'tribes' of researchers in education

One of our significant findings is the reported division between educationalists, who focus on non-causal aspects of education, and development economists, who research the causal relationship between interventions and outcomes. In many countries, economists seem to get preferential access to policy-makers and therefore their ideas and findings are more frequently adopted into policy than are those of educationalists. Understandably, this annoys educationalists who claim to understand education much more deeply.

These 'tribes' reportedly have little interaction – neither on prioritising research topics, conducting or interpreting research, or working with practitioners or policy-makers to use it.[197] Conceivably they genuinely cannot add value to each other's work, though this seems unlikely given the benefits which medicine has found in combining the approaches. A trivial example of the importance of context is cited by Howard White, Executive Director of 3ie which funds and conducts many RCTs in international development: 'you don't need an RCT to tell you the effect of opening the clinic another day a week if the boat there only goes once a week.'[198]

Funders can encourage this collaboration. Clearly this needs to be done carefully, and should perhaps focus on encouraging conversations, rather than incentivising fake collaborations. One

developed-world researcher told a salutary story about being forced to collaborate with an LEDC university, which simply charged him $5000 to put its stamp on his proposal and then contributed nothing to the research.[199]

> **Emerging recommendation 8:**
> Funders could encourage and facilitate more interaction between the two 'tribes' of researchers – educationalists and development economists. It seems likely that they could achieve more (and/or avoid redundant research) by collaborating better.

6.10 Qualitative research

Qualitative research is produced by anthropologists and educationalists interested in pedagogy and understanding educational contexts. Examples include 'Grounded Theory' which is based or 'grounded' in observations or data,[200] or using field research and questionnaires to observe participants in their natural setting.[201]

Qualitative research can suggest interventions that might improve outcomes, which can then be examined in trials.[202] It can also help to explain *why* an intervention worked, instead of just showing *that* it worked.[203] For example, a qualitative input on financial incentives might explain why financial incentives for teachers cease to improve educational outcomes after a certain point. Another example is analysis by Professors Bruce Fuller and Prema Clarke of Harvard University into the impact of cultural conditions on school effectiveness in developing countries.[204]

6.11 Non-experimental data

Interrogating historical data can show a great deal. Fryer has done this for US education. Observing that education standards have remained stagnant or declined (see below), one might hypothesise that this is due to badly qualified teachers. But he shows that teacher qualifications have in fact risen over the period:

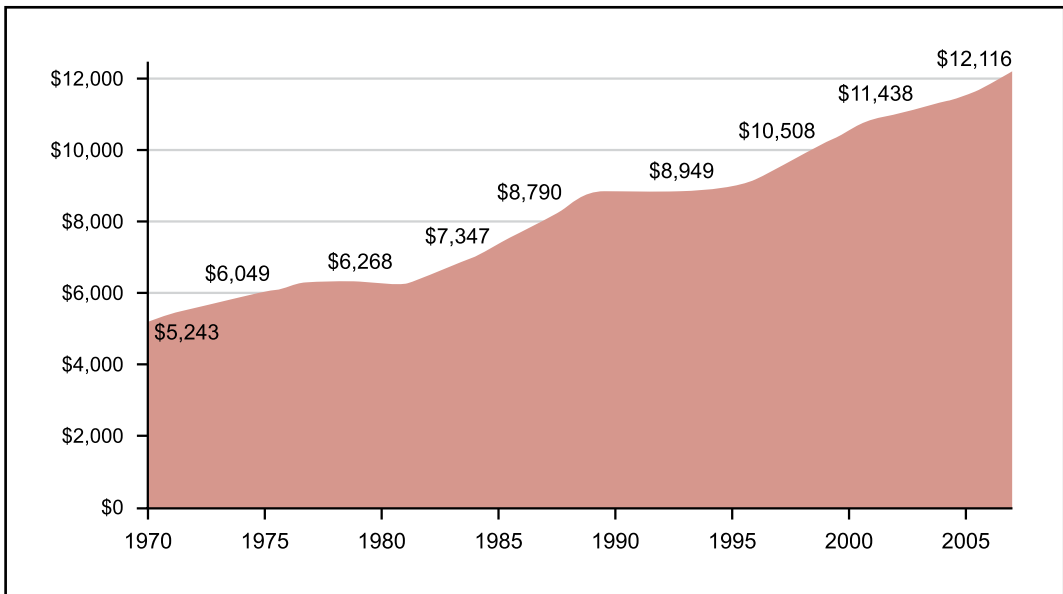**Figure 17:** Educational attainment for US students 1970 – 2008[205]
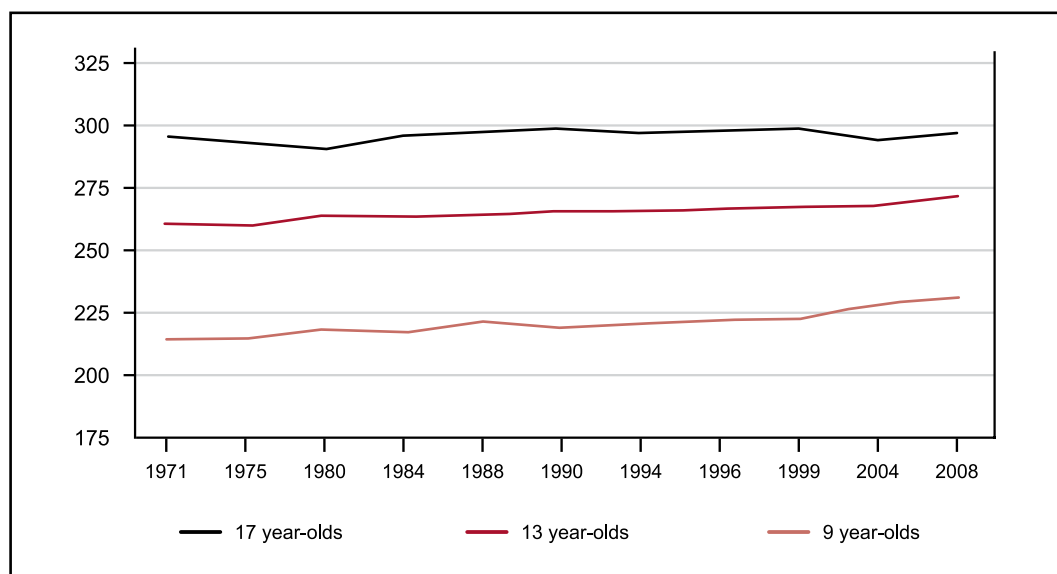
**Figure 18:** Percentage of US teachers with a Master's degree or higher 1961 – 2006[206]
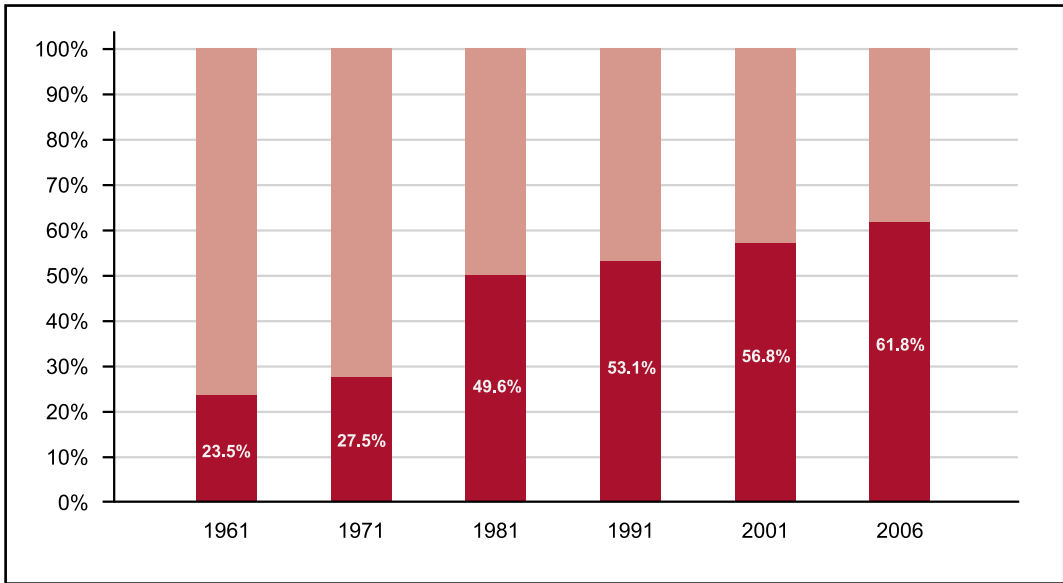
One might then suggest that the stagnation is due to falling expenditure per child, to which he shows that expenditure has in fact risen dramatically:

Figure 19: Total expenditure per pupil in the US 1970 – 2008 (USD)[207]

Alternatively it might be due to an increasing student to teacher ratios, but historical data show that this has fallen:

Figure 20: US student to teacher ratio, 1970 – 2008[208]
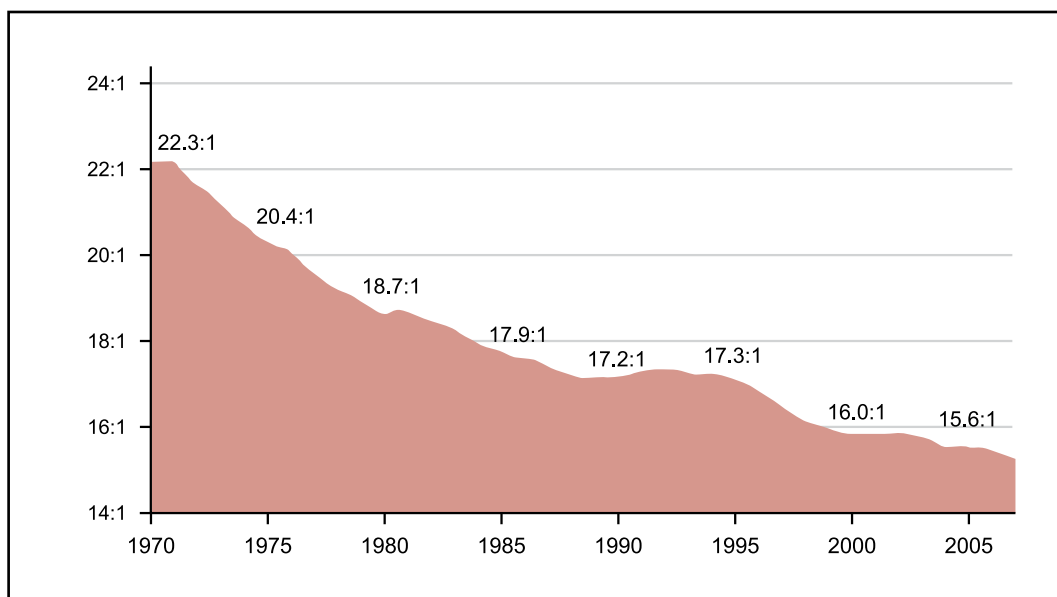
In this way the hypothesised causes of a 'condition' can be ruled out systematically. Clearly this analysis is powerful and cheap if historical data are available.

Cohort studies follow a group of individuals over time to assess the impact of circumstances and interventions. An example is Young Lives, a 15 year study based at Oxford University which follows a cohort of children in four low-income countries – Ethiopia, India (Andhra Pradesh), Peru and Vietnam. It covers topics including nutrition, health behaviours and education, along with the social, demographic and economic status of households.[209]

Young Lives has found some notable differences between education delivery systems. For example, in Vietnam teachers have a good understanding of the learning levels of their students, and there is a relatively small variation in pupils' attainment. In contrast, Indian teachers' assessments of pupils' ability are much less accurate, and learning levels are much more widely spread.[210] Caine Rolleston, Young Lives' Education Research Officer, suspects that the difference is due to history: India was encouraged by colonial powers to produce an elite and so the system tacitly encourages training 'winners', whereas the Communist history of Vietnam emphasised universal attainment.[211]

Box 10: The Khan Academy: A disruptive online intervention

The Khan Academy aspires to 'a free world-class education for anyone everywhere'.[212] It creates thousands of video tutorials on YouTube on a wide range of subjects and in many languages.

It also provides problem exercises for students. Because these are conducted online, the Khan Academy can see where and when a particular pupil is stuck or going fast, and hence provides a kind of continuous assessment of their learning. *The Economist* observed that 'You can follow the progress of each child – where she started, how she progressed, where she got stuck and 'unstuck'. You can also view the progress of the entire class. And you could aggregate the information of all the classes taught by one teacher, of an entire school or even district, with data covering a whole year.'[213]

Hence the Kahn Academy inverts much of the standard way of delivering education. It is individual rather than a batch process; and frees the teacher in class from being 'a sage on stage' to help students with exercises, which they otherwise would do alone at home.

Through online exercises, the Kahn Academy creates and captures baseline data on students' abilities, as well as evidence about the effectiveness of various lessons in helping them to progress.

6.12 Education: What isn't produced, and why not?

Most trials last a few years, if that, and hence focus on rapidly-observable outcomes such as test scores. As discussed, no relationship has yet been established between these short-term outcomes and eventual desired outcomes e.g., whether/how/when test scores correlate to labour market participation.

There are two possible remedies. One is longer trials, which observe the eventual outcome. The other is to research intermediate (or proximate) outcomes, and find intermediate 'markers' which reliably indicate that an eventual outcome will be achieved. Intermediate 'markers' are used extensively in health economics. For example, a researcher might be interested in creating outcome E, but only has evidence showing the probability that intervention A leads to outcome B. Health economists assemble data about probabilities and conditions necessary for getting from intervention B to outcome C, and thence to outcome D and so on.[214] Clearly funding and research capacity are required for either remedy.

6.13 Health: Who funds the production of evidence?

Health research is primarily funded by companies in the pharmaceutical and medical devices industries, national governments and charitable foundations. Total UK expenditure on health-related research from 2009-10 was approximately £8 billion, almost a third of all research and development spending that year.[215] Approximately half of this funding came from the private sector.[216] In the US, total spending on health research in 2010 was $140 billion, with industry investing $76 billion and the government $46 billion.[217]

On a global level, organisations such as the Global Fund to Fight Aids, Tuberculosis and Malaria (GFATM), the United Nations Children's Fund (UNICEF), the United Nations Development Programme (UNDP), the World Bank and the WHO also provide significant funding into health research. This funding is distributed by proactively commissioning studies by organisations such as the Special Programme for Research and Training in Tropical Diseases, hosted at the WHO, which put out regular requests for proposals.[218] The WHO is itself funded by member states and pharmaceutical companies, research institutes, banks, foundations and NGOs.[219]

6.13.1 Baseline data

Funding for baseline data are primarily provided by local, national and supranational governments and by not-for-profits researching specific conditions, such as the Association for Cancer Research in France.[220] International organisations such as the WHO also fund some data collection.

6.13.2 Intervention effectiveness research

Approximately 80 per cent of all clinical trials are funded by the pharmaceutical industry, conducted in-house or by universities.[221] Non-commercial funding comes from: national governments via health ministries; government-funded institutions such as the National Institutes of Health (NIH) in the US and the MRC in the UK; not-for-profit organisations including grant-making foundations such as the Wellcome Trust;[222] and supranational bodies such as the WHO and World Bank.

6.14 Education: Who funds the production of evidence?

Education research attracts only a fraction of the funding than health research receives.[223] In 2011 the UK government spent nearly 60 times as much on health research as it did on education research (see Fig 21).[224] Until 2013, DFID spent £75 million on health research, but less than £0.5 million on education research.[225] This explains not only the discrepancy in the amount of research available in education, but probably also the range of topics investigated. Precise figures for funding from the various sources proved elusive because research budgets are rarely itemised in public documents.

We're yet to understand *why* health receives so much more funding. One interviewee speculated that it is because a few landmark studies in health demonstrated outstanding results. It might be because healthcare affects adults, i.e., the people who pay for it, whereas education directly affects only children (an agency problem). Many health funders are founded with money made in health. For example, Henry Wellcome made his fortune in the pharmaceutical industry and founded the Wellcome Trust, one of the largest health funders in the world.[226] It is difficult to make serious money in education and so fewer funders supporting education have been created this way.

The funding that is provided comes from government departments of education, UNESCO, bilateral donors and charitable foundations. The World Bank is a major player – its annual research budget is approximately $52 million per annum[227] – yet perhaps even there, research isn't valued. An independent evaluation of its research activities between 1998 and 2005 found that 'Research is seldom part of an income-producing lending program, and it often fails to deliver the simple syllogisms that management wants to espouse in its advocacy of what it believes to be best development practice. The simple and compelling fact is that despite its centrality to the Bank's mission, research only accounts for 2.5 per cent of the Bank's administrative budget. By contrast, and as an example, the fraction devoted to supporting the Bank's executive board is more than twice as high.'[228]

Figure 21: UK government research spend on health and education in 2011[229]
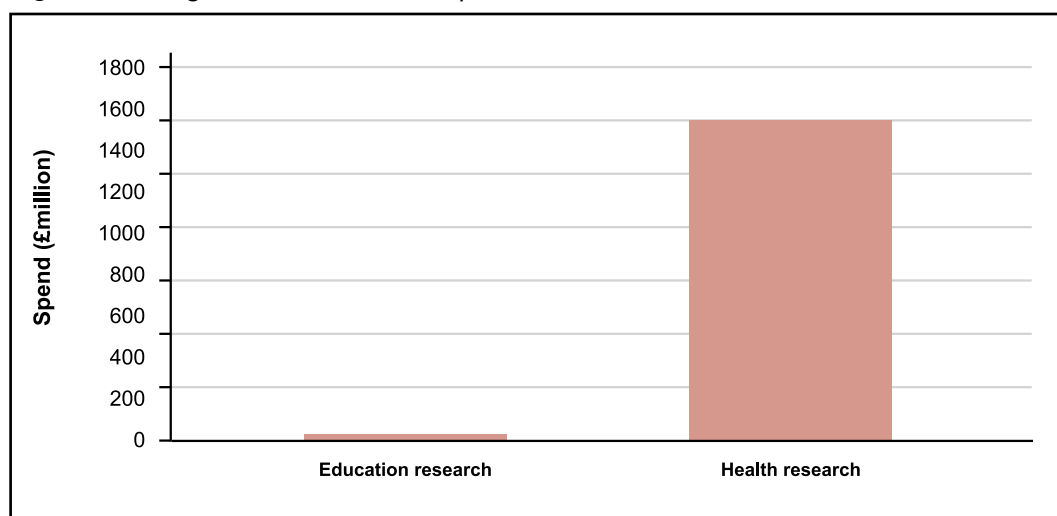
> **Emerging recommendation 9:**
> Increase funding for researching the effectiveness of interventions, since so little currently exists.

6.14.1 Baseline data

Funding for baseline data produced by the UIS (UNESCO's Institute of Statistics) has been predominantly provided by governments and their development agencies, including Australia, Canada Denmark, Finland, the Netherlands and Norway, plus UNESCO[230]. Pratham's work on ASER is funded philanthropically, e.g., by the Hewlett Foundation and local individuals, many of whom donate tiny amounts.[231] PISA is financed by the participants' national authorities, typically education ministries.[232]

6.14.2 Intervention effectiveness research

Significant funders here are the World Bank and national governments, through international development departments such as DFID and USAID.[233] For example, DFID has funded 14 SRs on education in LEDCs, as well as providing funding for both the Campbell Collaboration and EPPI-Centre.[234] Research funders are also involved, such as the UK Economic and Social Research Council.[235]

The GPE funds some academics researching intervention effectiveness. GPE is similar to the Global Fund to Fight Aids, Tuberculosis and Malaria (GFATM) although much smaller: the GFATM disbursed almost $3.5 billion in 2012,[236] versus GPE's $1.5 billion pledged by multilateral and bilateral donors between 2011 and 2014.[237] GPE is funded by bilateral donors, companies such as Microsoft and Pearson, and multilateral agencies including the World Bank.[238] 3ie funds (and conducts) studies and SRs into intervention effectiveness, including many RCTs.[239] Some charitable foundations (e.g., the Children's Investment Fund Foundation and the Hewlett Foundation) commission research into interventions relating to their programmes, such as de-worming.[240]

J-PAL's research also draws funding from Dell and Nike, foundations such as the Gates Foundation and Ford Foundation, and multilateral agencies including UNICEF and the UN Development Programme.[241]

6.15 Health: Why do people produce evidence?

6.15.1 Baseline data

Baseline data production has been driven in part by the priorities of national governments. For example, the UK government currently has a focus on dementia (domestically) which has released funding for research.[242] Comparative data, such as PISA, are used for international rankings, which can spur competition and highlight good practice to emulate.[243]

6.15.2 Intervention effectiveness research

As discussed, pharmaceutical companies drive much of the research into intervention effectiveness for commercial and regulatory reasons. Regulatory requirements drive the production of this type of evidence (see Box 11). In many regions and countries it is illegal to deliver medical treatments without approval by government regulator, such as the European Medicines Agency (EMA) in the EU[244], Food and Drug Administration (FDA) in the US[245] and the Therapeutic Goods Association (TGA) in Australia.[246] **This approval is based on evidence demonstrating the effectiveness of the intervention.**

Academics in health are often driven by the prestige of publishing in prominent journals (see Fig 21). As many of the producers of evidence are also practicing clinicians, intervention analysis is often also driven by professional interest in a specific healthcare subject (as was Cochrane's original interest in obstetrics).[247] The production of SRs is in part driven by clinicians who themselves need a new summary of research.[248] Cochrane Reviews often also highlight areas which need further coverage.[249] The intermingling of research and practice in health creates a supply of research time. One effect is that The Cochrane Collaboration does not offer funding to researchers (although in practice this research time can be limited to just a few hours a week, meaning that Cochrane Reviews can take a long time – e.g., 150 hours on the review might take a year).[250]

National governments drive the production of evidence by funding research into topics of particular political interest, while NGOs representing patient groups (e.g., arthritis or cancer patients) have in their mission to alleviate the incidence of disease, and so fund research.[251] There are few (if any) NGOs in education whose role is analogous to patient groups, such as parent groups, so this source of pressure to improve performance and source of funding just doesn't exist in education.

Box 11: Regulatory requirements and the production of evidence

Health

Medicine and medical procedures are heavily regulated: in many countries it is illegal to dispense a treatment that has not been approved by government. The quality of evidence demanded by these government agencies has been increasing,[252] with much of this regulation related to pharmaceuticals and medical devices. In Europe (and potentially further afield in future) these interventions are evaluated by Health Technology Assessments (HTAs).[253] Post-market regulation can include the review of clinical experiences and any reported failings or adverse reactions to a particular drug, device or intervention.

In the EU, the European Medicines Agency (EMA) must approve any new pharmaceutical before it goes to market. If a manufacturer seeks approval in more than one member state, it must apply for EU-wide permission.[254] The EU also regulates medical devices, ranging from low-risk items such as stethoscopes to high-risk interventions such as heart valves.[255] EU regulation of high-risk devices faced criticism recently when silicone breast implants that had been approved only in the EU and were later found to contain lower quality silicone than required by regulators.[256]

In the US, the Food and Drug Administration (FDA) regulates both pharmaceuticals and medical devices, and mandated the creation and use of clinicaltrials.gov, a database of clinical trials in the US.[257] The evidence required for the approval of medical devices varies, depending on the level of risk associated with a product. Low-risk items are subject only to 'general controls' such as sterility, while high-risk items require clinical trial evidence.[258]

HTAs are designed as a bridge between research and policy-making,[259] synthesising existing research, undertaking or commissioning trials, and collecting additional data. They summarise the available information on a particular health intervention in a systematic and transparent way to inform policy-makers and organisations such as the National Institute for Health and Care Excellence (NICE) in the UK.[260] An HTA can also recommend additional research to generate more evidence on particular health interventions.

The first HTA agency was created in Sweden in the 1980s, and there are a number of HTAs throughout Europe.[261] International organisations, such as the Organisation for Economic Co-operation and Development (OECD) and the WHO, are currently evaluating the introduction of similar assessments around the world.

Education

By contrast, intervention in education faces few regulatory hurdles. There are no real parallels to pre-market regulation based on evidence. Instead, curricula (the closest analogue to treatments in medicine) are influenced by individuals, political or other special interest groups, current educational trends and recent events, or funding organisations that require the inclusion of particular topics.[262] For example, in Vietnam, the Ministry of Education and Training has obliged children studying in schools funded by foreign investment to obtain a 'full diploma in Marxist-Lenin philosophy, Marxist Lenin political economy, scientific socialism, history of the Communist party of Vietnam and ideology of Ho Chi Minh'.[263]

Reform of curricula and educational goals (a possible parallel to post-market regulation in health) is rarely based on an assessment of 'what works' in classrooms, and again is often influenced by political forces or particular leaders or practitioners.[264] Testing new curricular design is complicated by incomplete or inaccurate reporting of results, from a local to central government level.[265] Supervision of educational interventions in LEDCs is difficult because of poor infrastructure and transportation, the lack of sufficiently qualified supervisors and constraints on funding.[266]

Private schools in LEDCs pose a further regulatory challenge. They have grown significant in recent years, making up over 50% of all schools in some countries (such as Bangladesh and Guatemala)[267] and varying markedly in quality. In any African countries, many private schools are 'clandestine', operating without a government licence.[268] Attempts to regulate the private sector include restricting access to nationally-recognised exams to formally accredited schools. However, this type of regulation has been criticised for being outdated and subject to the whims of accrediting officials.[269] In some cases, the need for accreditation has limited the effectiveness of private schools, as they are forced to follow poorly-developed national curricula if they are to be recognised by government.[270]

In India, commentators have called for the establishment of an Educational Regulatory Authority to prevent the exploitation of parents and teachers by those investing private funds into schools,[271] as well as strengthening the governance and monitoring of Indian public schools more generally.[272] Recent research, consisting of interviews with school managers, teachers, parents and senior government officials in Hyderabad, Andhra Pradesh, revealed that at the state level regulation theoretically covers all aspects of a school's work. However, due to bribery and corruption, and a lack of inspectors, none of the recognised schools met more than two of these regulations.[273]

6.16 Education: Why do people produce evidence?

The incentives for producing evidence about education are quite different from those in medicine. There is little demand from education practitioners, much less funding, less prestige, little if any regulatory requirement, and research and practice are normally quite separate. The drivers that do exist are discussed below.

6.16.1 Baseline data

National governments are one driver of production, as they sometimes use poor performance in baseline data studies as a trigger to initiate new policies.[274] International goals and targets around education, such as the MDGs, are another driver to produce baseline data.

6.16.2 Intervention effectiveness research

In some cases, education NGOs and programme managers drive the production of evidence, as they are interested in understanding how to increase the effectiveness of their work. Pratham was keen to improve educational performance in India, and so worked with J-PAL on RCTs into its educational interventions.

Pratham is somewhat unusual in this respect, because much 'monitoring and evaluation' of education and development programmes is essentially compliance. It aims to 'prove' that the project has

been run correctly, rather than to learn to improve which is much more commonly the goal in health. Nonetheless, increasingly many implementing organisations are initiating research into the effectiveness of their interventions, creating a growing research market.

Private research providers, such as those contracted by Pratham and EurekaChild (an Indian NGO), are likely to have commercial incentives to produce evidence. However, perhaps these incentives make it more likely that their evaluations of clients' programmes will be positive.

As with baseline data, the production of intervention effectiveness data is driven in part by the policy priorities of national and international governments and international organisations such as the World Bank, which provide funding into studies relating to these priorities.[275] Another possible driver could be governments and foundations. If they decided to fund only education interventions with a strong evidence base, NGOs would be incentivised to properly evaluate their programmes.

6.17 Health: What are the barriers to producing evidence?

As discussed, academics are incentivised to produce the **first and last studies** on a topic that prestigious journals are reportedly interested in publishing, discouraging the production of replication studies.

Cognitive biases also impede production – for example, the biggest RCT in the world (on de-worming in India) was withheld for five years allegedly because the results were surprising – essentially a case of confirmation bias.[276]

SRs are only reliable when the studies being synthesised are similar. This is rarely the case, hence this acts as a barrier to their production.[277] (Cognitive biases are discussed more in Chapter 8.)

Evidently some subjects are perceived as so taboo that publishing on them risks academics losing tenure, despite the tenure system being intended to provide 'editorial freedom' to researchers. For example, it recently came to light that research which showed that the impressive performance of long distance runners in Kenya might be related to genetic factors was suppressed by an academic who feared losing tenure for reasons of racism.[278]

6.18 Education: What are the barriers to producing evidence?

6.18.1 Baseline data

Baseline data production is inhibited by a lack of funding.[279] It can be subject to production bias. In some states of India, baseline scores were so bad that producers (state governments) did not want to repeat the study.[280]

It might be possible to produce a false result if pupils are tested too soon on the information they are supposed to know, as it takes time to adequately absorb new information. This might reduce the reliability of evidence.[281]

6.18.2 Intervention effectiveness research

Funding is a significant barrier to producing evidence.

The number of producers is very limited, in marked contrast to health.[282] This may in part be due to the limited intermingling of practitioners and researchers: demand for evidence is not created and therefore less evidence is produced. Several people commented that the 'market' and debate is dominated by J-PAL and that they would like to have alternative providers and voices. (In fairness, because J-PAL is a network of accomplished academics active in this field, it's hard to see how it could succeed without having as members most leading voices.)

Lack of prestige may impede research into education. For instance, in contrast to medicine there is no Nobel Prize for education.[283]

There is **limited commercial potential** of education evidence, in contrast to the significant commercial opportunities available to pharmaceutical companies.

There are a lack of **career incentives** for practitioners to use evidence: doing research will not make it easier to become a head teacher or a senior policy-maker.

Long-duration studies face particular barriers because they do not produce results fast enough to be useful to programme managers or policy-makers. Hence they struggle to attract funding. Sometimes they also struggle to engage organisations running interventions which might be studied, given that trials create some disruption for implementers.[284]

Teaching unions in both LEDCs and MEDCs have reportedly hindered studies of interventions. Some unions seem to resist moves to measure teacher performance: many of the arguments used are similar to those used in health some decades ago, such as that performance is fundamentally unmeasurable.[285]

Concerns about the ethics or cost of RCTs may be a factor. As discussed, most of these concerns are in fact misplaced, though remain common nonetheless.

The low generalisability of findings in education limits the usefulness of study findings. If educational interventions are more specific to local contexts than those of health, education research will be less useful when put into new contexts than health research. This can reduce the number of people who can benefit from a piece of education research, which reduces the research's value. However, rather than seeing this as an argument *against* education research, Goldacre's view is that 'it's an illustration of why we need to make trials part of the everyday routine, so that we can replicate trials, in different settings, instead of blindly assuming we can use results from other countries. It doesn't mean that we can never trust the results of a trial.'[286]

Small effect sizes are often seen in intervention effectiveness studies, and are unpopular with journals. The lack of academic incentives to produce and publish trials with small effect sizes acts as a barrier to producing evidence (see Box 4).

6.19 Health: How are rational decisions made between interventions?

Given limited resources, some governments have developed mechanisms to prioritise interventions. Health economics is 'the application of economic theory, models and empirical techniques to the analysis of decision-making by individuals, healthcare providers and governments with respect to health and healthcare.' [287]

These mechanisms enable comparisons between alternative courses of action (treatment) in terms of their costs and consequences. They include:

- In **cost-minimisation analysis** the effectiveness of the two interventions must be proven to be equivalent. The 'cost-effective' intervention is simply the one which costs less to achieve the same outcome. [288]

- In **cost-benefit analysis**, costs and benefits are both valued in cash terms whereas **cost-effectiveness analysis** measures outcomes in 'natural units', such as symptom-free days or life-years gained. [289]

- **Cost-utility analysis** measures length and quality of life, represented most frequently as the quality-adjusted life year (QALY). [290]

Box 12: QALYs – comparing health outcomes

Suppose we have two patients. They have different conditions which require different treatments with different costs. We cannot afford to treat both so must choose which patient will benefit most. The treatment for Patient A will make his life pretty nasty but will extend it a lot, whereas the treatment for Patient B will make her life reasonable but extend it relatively little. Health economists assess the length of life which each treatment adds, recognising also the resulting quality of life. Quality-adjusted life years (QALYs) 'measure a person's length of life weighted by a valuation of their health and related quality of life.' [291] Each year of life is assigned a score of quality, based on a wide range of criteria. A year of full health scores one, poor health scores less than one, death counts as zero, and some dreadful conditions score below zero.

QALYs were invented in the UK's NHS, which effectively has a single budget for the nation's health, and therefore requires decisions to be made between widely differing treatments. In effect, QALYs squash the numerous differences between such options into a single number which allows comparison. [292]

QALYs are helpful because although patients, conditions, treatments and costs vary hugely, the various treatments share the same goal – giving patients quality and quantity of life. QALYs have their detractors, not least because people disagree with the weightings of diverse variables. In the absence of an alternative, QALYs are widely used and helpful for medical practitioners and donors.

6.20 Education: How are rational decisions made between interventions?

Making comparisons between interventions is essential for policy-makers who are almost invariably deciding between courses of action. Yet studying comparative costs and benefits of education interventions is much less established in education than in health. For instance, there were a third fewer publications on education economics than on health economics from 1960 to 2011.[293] Education has no analogy of QALYs. However, J-PAL and others sometimes give the outcomes of various interventions as (for example) the additional years of education[294] or additional days of teacher attendance which they create in order to enable comparison.[295]

It is technically possible to consider the education interventions from a purely financial 'return on investment' perspective, e.g., the increase in tax revenue from a more economically productive population versus the cost of delivering education. However, Shawn Powers, Education Initiative Manager at J-PAL, suggests that this is too simplistic, and hence analysts rarely use it to compare interventions.[296]

Figure 22: Number of publications on health economics and education economics (1960-2011) [297]
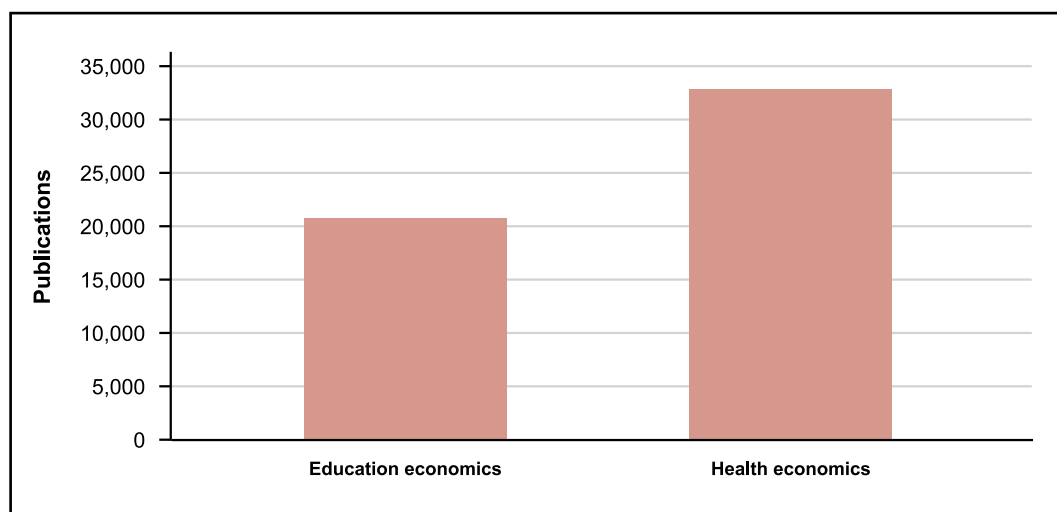
The effects of education interventions are often reported as changes in the standard deviations of results e.g., the equivalent standard deviation increase in test scores. Though precise statistically, it's rather unhelpful for policy-makers and decision-makers, being rarely understood. Worse, it doesn't allow interventions to be compared (should we fund more blackboards or more teachers?) A recent paper in the journal *Science* discussed the effect of numerous education interventions in terms of standard deviation changes, to the criticism that this doesn't help decision-makers who need to choose between them (See graph below).[298]

Figure 23: Impacts on test scores from randomised trials of primary school programmes in the developing world[299]

6.21 Conclusion

It is currently genuinely hard – perhaps even impossible – for education policy to be evidence-based. High quality evidence is far too scarce, and the evidence which does exist doesn't provide guidance about allocating resources. Hence priorities are understanding the current status of education by collecting baseline data, and studying what interventions could improve outcomes. This evidence will need to be of sufficient quality to ensure the results are reliable.

7 Dissemination of evidence in health and education

If evidence is to be incorporated into practice, effective dissemination is crucial. Effective dissemination comprises both 'push', wherein the research finds the user (e.g., through journals and conferences), and 'pull', wherein users can speedily locate the articles, guidelines or research they need. Neither of these is currently strong in education. The upshot is that it's difficult to find evidence about education, and even harder to see whether it's reliable. Since, as Thaler says, 'if you want to encourage some activity, make it easy' , we recommend improving this.

The users of evidence in health are clinicians, health service managers, patients, policy-makers, international organisations such as the WHO, and NGOs, funders, companies active in the health sector and academics. In education they are school principals and teachers, policy-makers, funders, programme managers including NGO and social enterprise staff, teacher trainers, parents and communities, international bodies such as UNESCO, private companies that provide equipment to schools, academics and economists. (Listed In Chapter 8.3, Table 6 and Table 7)

7.1 What are the main differences between dissemination in health and education?

The most striking difference is the paucity of dissemination channels in education. None reliably reaches practitioners, presumably in part because there is so little evidence to disseminate. It is also likely that infrastructure problems in LEDCs, such as poor internet connections, an infrequent postal service, and limited transport also hinder disseminating evidence through many of the channels used in health.

The dearth of evidence itself deters funding for dissemination – there is little point spending on dissemination if there is little to disseminate. A second explanation is the involvement of the pharmaceutical industry in funding medical journals (which in fact appear to warp what ostensibly independent journals will publish[300]), for which there is no equivalent in education.

Journals in education are aimed primarily at economists and other researchers, which led one senior education funder to comment that 'the audiences for economics papers are other economists'.

In terms of users 'pulling' information, this is much easier in health than in education because the information architecture is better – it's easier to find information. Articles are published in journals which are findable, and they have standard, searchable formats. However, closed access, paywalled, subscription-only health journals and databases impede this 'pull'.

7.2 Which organisations disseminate evidence in education?

The World Bank, UNESCO and the GPE Fund are all major funders of dissemination in education, and also disseminate the evidence that they produce.

The GPE, launched by the World Bank in collaboration with other donors in 2002 has explicitly prioritised dissemination, stating that 'the dissemination and cross-pollination of education sector knowledge lies at the heart of the Global Partnership for Education's work program.'[301] Its Global and Regional Activities programme is designed to disseminate information on three topics: learning outcomes; out of school children; and education financing.[302] Databases and libraries are also funded by other foundations: for example the International Institute for Educational Planning (IIEP) is funded by the Hewlett Foundation.[303]

The World Bank is an important disseminator (as well as producer, funder and user) of evidence. It uses conferences, its website, regular reports (such as the World Development Report) and face-to-face meetings with policy-makers and programme managers.[304]

7.3 What are the main dissemination channels?

7.3.1 Journals

7.3.1.1 Health

Journals are the main way of 'pushing' primary academic medical research to users, and include specialist journals such as the *Journal of Medical Genetics*[305], and general journals such as the *BMJ*.[306] Journals can be 'closed,' 'free' or 'open access' (see box below), and both on- and off-line, which variously affects the ability of users to access the evidence.[307]

Funding: Some journal publishers are commercial; others, such as medical societies, are not-for-profit. Sources of revenue include subscriptions: universities complain because they have to pay subscription fees while their academics generate and peer-review content, unpaid. Subscription rates vary, but as an indication the print version of *The Lancet* costs an institution about £1,000 per year.[308] Journals are incentivised to accept articles from pharmaceutical companies to increase their income from reprints[309]. Journals (both open access and closed access) also receive advertising revenue.

Journals such as *PLOS Medicine* are open access. Here, fees come from authors rather than from subscribers or readers: authors pay a fee to cover peer review (which is important for the credibility of the study), journal production, and online hosting and archiving. *PLOS* and other open access publishers offer fee waivers to academics in countries with lower incomes, to increase participation.[310]

An important **driver** of dissemination is the prestige of publishing in well-regarded journals. However, there are fewer incentives for academics to disseminate information through channels that reach mainstream media, policy-makers, or in a way that is understandable for 'lay' users.

Barriers to disseminating evidence through journals include publication bias and the particular difficulty in getting replication studies published despite initiatives such as the Reproducibility Initiative, organised by *PLOS One*.[311] The risk of legal challenges to published research is another barrier: Goldacre and science writer Simon Singh have both been sued for libel for comments in mainstream media critical of health treatments (homeopathy[312] and chiropractic[313] respectively). Dr Peter Wilmshurst was sued for libel for comments made at a medical conference about a trial in which he was involved.[314]

Box 13: Open access, free and closed access journals

Open Access was defined by the Budapest Open Access Initiative as:
'By "open access" to [peer-reviewed research literature], we mean its free availability on the public internet, permitting any users to read, download, copy, distribute, print, search, or link to the full texts of these articles, crawl them for indexing, pass them as data to software, or use them for any other lawful purpose, without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself.'[315]

Authors looking to publish on an open access basis can use the 'gold route', where they (either personally or via a research grant) pay a fee to a journal to make their work free at the point of access, and available for reusing and reproducing (as in the case of *PLOS Medicine*[316]). Alternatively, authors can use the 'green route', where they have permission to self-archive papers on institutional websites at the time of submission to a publisher.[317]

Many journals will offer some degree of open access, as this is often required by funders.[318]

Free journals make articles available without cost. The journal retains copyright. Publications can either make journal articles free to read immediately or after a period of time.[319] Free access does not mean open access.

Closed journals charge users for content, e.g., through subscriptions or paywalls. Copyright remains with the publisher, and authors are not free to publish their own work on other websites.[320]

7.3.1.2 Education

Education journals do not seem to be effective at disseminating findings to users. Firstly, there are no go-to journals for education[321]: one senior funder told us that he's as likely to find education results in *The Economist* as anywhere. Second, there is not a strong culture or norm amongst practitioners of reading journals. And third, economists investigating education normally use journals that are unlikely to be read by practitioners or policy-makers in education or even educationalists (e.g., *Science* or *Econometrica*[322]). These are not aimed at or read by practitioners or policy-makers, and the articles are written in an academic language (sometimes the language of academic economics) which is reportedly difficult for non-specialists. Journals do not translate material for non-academic users.

Other **barriers** to dissemination through journals are similar to health, and include publication bias,[323] closed access journals and the difficulty of publishing studies showing small effects. There are also few incentives for academics to release data in a timeframe that might be useful for practitioners. Few funders require dissemination plans in advance of funding, although some are introducing this policy.[324]

> **Emerging recommendation 10:**
> Create better 'push' channels for disseminating education research to its intended users, such as journals, conferences.

<div style="background-color:#b01c2e; color:white; padding:1em;">

Emerging recommendation 11:
Increase the incentives on researchers (both academics and practitioners, e.g., NGOs) to share their findings with policy-makers and the field. Funders could make this a condition of funding**.**

</div>

7.3.2 Databases and websites

7.3.2.1 Health

Most databases are now online, many of which are funded by national or federal governments. In the US, the National Library of Medicine offers access to a range of databases including ClinicalTrials. gov and PubMed.[325] PubMed Central has more than 23 million citations and abstracts from biomedical literature articles from over 5,600 worldwide journals in about 40 languages journals covering topics across health[326]. It is funded by the US government to provide open access to a wide range of intervention effectiveness data.[327]

7.3.2.2 Education

Databases are an important 'pull' mechanism in education, although as there is less to disseminate, there are fewer databases. The information in these databases is not as rigorously systematised as in health. This might be because the evidence is less standardised: on some databases it includes meeting minutes, journal articles and NGO reports.

Major databases include the Research and Evidence Library, produced by the Center for Education Innovations, which holds resources related to global education topics such as innovations in non-state education.[328] The Education Resources Information Center (ERIC) includes bibliographic records (citations, abstracts, and other pertinent data) for more than 1.4 million items indexed since 1966. It has permission to display more than 400,000 full-text materials in PDF format, although these materials are generally conference papers and reports, rather than journal articles and books.[329] The IIEP, part of UNESCO, has an online library, holding a large selection of key publications on educational planning,[330] while J-PAL has an online registry of all its RCTs.[331] The GPE also disseminates evidence to other grantees (including practitioners and NGOs), as well as more generally via its website.[332] Baseline data are published on the UIS and World Bank websites.[333]

7.3.3 Conferences and lectures

7.3.3.1 Health

Clinicians, health professionals, relevant health commissioners and policy-makers can receive up-to-date information by attending conferences, which often form part of a clinician's CPD requirements. These conferences may be organised by a professional group or Royal College (in the UK), or by a national body e.g., the Center for Disease Control (CDC) in the US.[334]

7.3.3.2 Education

Conferences are one of the principal mechanisms for disseminating evidence in education. The World Bank holds a range of conferences, which often discuss evidence that it and national governments have produced.[335] IPA and J-PAL also run conferences for policy-makers, e.g., a 2012 conference with the Ghana education ministry entitled 'Evidence-based education: Policy-making and reform in Africa'.[336] The Brookings Institution has an annual research conference, and also organises one-day events such as New Ideas to Scale Up and Finance Global Education.[337] Bilateral agencies and NGOs organise conferences to disseminate information to each other and to policy-makers, which might also influence participants' decision-making on funding and further topics for research.[338] The GPE disseminates the results of its own studies to its grantees through technical

workshops, peer-learning events, and more generally through its website as part of its Global and Regional Activities programme.[339]

Comparative International and Education Society conferences are held annually around the world, bringing together practitioners, policy-makers and other actors to discuss education on a regional, national and international level.

7.3.4 Professional groups and membership societies

7.3.4.1 Health

Clinicians have access to professional groups or membership societies, such as the Royal College of Surgeons in the UK, or the Royal Australian College of General Practitioners.[340] These societies provide updates on evidence with newsletters, lectures, workshops and conferences. They also play a crucial role in translating evidence into clinical practice guidelines (CPGs) on various health topics. Membership of a Royal College or society is sometimes compulsory.[341]

7.3.4.2 Education

Academics, practitioners and students receive trial results through professional groups or membership societies, such as the Comparative and International Education Society (CIES) in the US. Generally, these organisations are less established than those in health.[342] Evidence is disseminated to members through newsletters, websites, conferences, networking events, and standing and ad hoc committees, as well as through journals produced by some membership groups.

7.3.5 Training and continuing professional development (CPD)

7.3.5.1 Health

Doctors are taught at medical school to produce, interpret and use evidence on the effectiveness of interventions, and are expected to keep up to date. A systematic review into CPD found that effective mechanisms for improving clinician performance include short courses, reminders, patient-mediated interventions, outreach visits and opinion leaders. The SR found that 70 per cent of the 160 CPD interventions reviewed improved physician performance.[343]

Doctors' medical education also includes grand rounds, where physicians present interesting clinical cases and recent research findings.[344] These can work well. However, like much in health, they are distorted by the pharmaceutical industry, often being sponsored by pharmaceutical companies, and hence predominantly discussing research favourable to those companies. Goldacre challenges this practice, asking why doctors need a 'commercial break' in the middle of their day.[345]

Medical education is generally funded either by national governments or by students (as in the US). Pharmaceutical companies are reportedly highly involved in medical education (including CPD), including face-to-face meetings between students and pharmaceutical representatives, sponsored publications and continuing medical education courses. Such interactions with pharmaceutical sales representatives have been shown to change prescribing behaviours.[346]

7.3.5.2 Education

The World Bank has identified that at least 25 per cent of primary school teachers in 22 LEDCs are not qualified.[347] In some instances, the requirements for teacher qualification are very low: for example, as part of the recruitment process for public schools in Mexico, teachers are given 120 questions but are only required to get 34 of them right to find a job.[348]

Practitioners receive some information about new research through pre-service and in-service teacher training.[349] Given that such a significant proportion of teachers in LEDCs have no training at all, this is inevitably a weaker dissemination mechanism than training in medicine.

Pre-service training is reportedly very theoretical and ideological.[350] In-service training is often driven by funders, such as Children's Investment Fund Foundation (CIFF) in Zimbabwe, although it is delivered by governments. This training is easier to adapt and update than pre-service education, although the usefulness of the content depends on regulation: a current donor suggests that most governments do not regulate this training.[351] We have not come across requirements for CPD in education in LEDCs.

Even teachers who are trained are rarely taught in either their pre- or in-service training to understand or use evidence.[352] There are no mechanisms that are analogous to grand rounds, where teachers could share experiences of interesting or difficult cases with other practitioners – although this might be relatively easy to introduce. Similarly, policy-makers in education are rarely trained or supported to understand, interpret or use evidence.

There is also little intermingling between practitioners and researchers in education. For instance, there are no 'teaching schools' in the way that there are teaching hospitals.

7.3.6 Prizes

7.3.6.1 Health

Prizes can be considered a dissemination mechanism because they highlight particular research findings. They also drive production of research that is likely to qualify for awards. The best known prize in medicine is the Nobel Prize in Physiology and Medicine, which is awarded to up to three people a year.[353]

In the UK, the Royal Society of Medicine offers prizes for members and non-members at all stages of their career, with prizes ranging from small financial awards to travel bursaries and bottles of champagne.[354] The Lasker Awards, established in 1945 by the Albert and Mary Lasker Foundation, offer prizes for advances in medical research, public service and special achievement, with 86 previous winners going on to win a Nobel Prize. The Hideyo Noguchi Africa Prize, funded by the Government of Japan, is for medical research to combat infectious and other diseases in Africa. At Y100 million ($976,000), the prize is substantial. providing an additional financial incentive for researchers to disseminate trial results.

7.3.6.2 Education

Prizes are also available in education, although to a lesser extent than in health.[355] The World Innovation Summit for Education (WISE) offers a prize of $500,000 for individuals who have demonstrated innovation in education.[356] The Grand Challenges for Development Initiative, led by USAID and funded by USAID, World Vision, Australian Agency for International Development (AusAID) and the US Department of Education, is designed to encourage organisations to publicise their evidence-based solutions to four 'Grand Challenges'. One of these challenges is entitled All Children Reading, and submissions are disseminated on the All Children Reading website, where they are likely to be seen by programme managers, policy-makers and practitioners.[357]

The Schools and Teachers Innovating for Results (STIR) Education Initiative (see Box 9) does not offer a prize as such, but is a mechanism for finding sustainable initiatives led by teachers, and

scaling them up. Currently only active in India, it plans to begin in East Africa fairly soon. STIR chose 25 of the best ideas to pilot and test their scalability.[358] These ideas may be adopted outside STIR's bailiwick simply through the profile and attention which STIR's involvement creates.

7.3.7 Change agents/face-to-face dissemination
7.3.7.1 Health
'Change agents' – individuals who are able to 'modify the social environment' and encourage people around them to develop the skills they need to change[359] – can effectively disseminate information to practitioners in face-to-face meetings and conversations. Sales representatives for pharmaceutical companies are focussed on developing relationships with clinicians to promote specific products, and hence can be regarded as change agents. Strikingly, three out of four of the WHO's stated primary audiences for a 2007 publication entitled 'A guide for fostering change to scale effective health services' were 'change coordination teams', 'change agents' and 'change agent's change teams'.[360]

7.3.7.2 Education
In some instances governments and policy-makers consult with academics and researchers in education to understand current evidence, as in the case of Muralidharan in India and Kremer in Kenya.[361] Staff at J-PAL and IPA also meet with governments and decision-makers to explain their results, because many of their trials are context-specific, and require detailed explanation and interpretation.[362]

The Hewlett Foundation previously hired a 'communicator' in East Africa to disseminate the results of trials that it has funded.[363] Some users of evidence, such as national programme managers (e.g., Pratham in India), discuss evidence with other practitioners.[364]

7.3.8 Media
7.3.8.1 Health
Mainstream media now play a growing role in sharing evidence, particularly as patients and the general public become more accustomed to receiving information in this way. The media will sometimes sensationalise and misinterpret evidence, as happened in the UK over the alleged connection between the MMR (measles, mumps and rubella) vaccine and autism, which had no basis in sound evidence.[365]

The pharmaceutical industry markets its products in the mainstream media, as well as in medical journals. For example, a journal article in June 2013 revealed how medical writers are employed by pharmaceutical companies to produce materials for consumer-facing magazines. In one example, the writer was paid to write material claiming that low testosterone has one effective treatment – produced by the client company.[366]

7.3.8.2 Education
National and international media disseminate evidence in education if it is deemed to be of broad public interest. For example, *The Economist* recently wrote about the impact of the Khan Academy, and the Indian press have an ongoing interest in ASER results.[367] Programmes undertaken by NGOs, and potentially the research that underlies it, are also highlighted by the media.[368]

As in health, there is a focus on sensationalised stories and findings, probably because the incentives for media outlets to disseminate evidence (e.g., gaining audience share) do not always lead to objectively communicating results.

Box 14 The power of peer to peer communication

Health

Word of mouth is better at informing UK General Practitioners (GPs) about innovations than journals, conferences or any other information channel, according to a recent survey by Nesta, a UK innovations agency.[369]
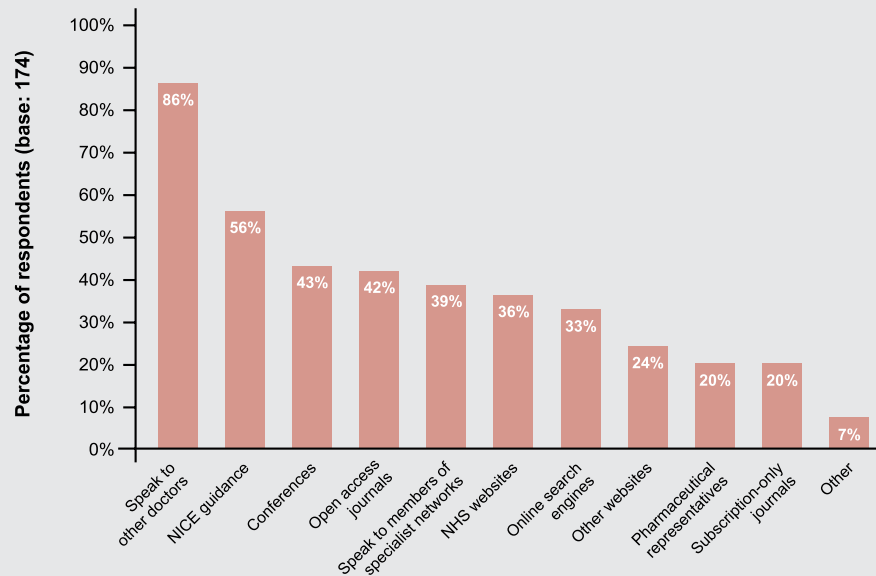
Strikingly, the number of GPs who (claim to) get the information they are seeking by speaking to other doctors is more than twice the number who go to open access journals, and that is more than twice the number who use subscription-only journals. This perhaps indicates a shift away from the dominance of closed access journals.

Furthermore the study found no correlation between patient satisfaction (as measured by the UK GP Patient Survey) and a GP's early adoption of innovation.[370] Perhaps this is unsurprising: patient satisfaction is influenced by factors such as 'bedside manner', waiting times and appointment schedules, as well as the treatment's effectiveness. This contrasts with the growth in patients' interest in evidence, as it becomes easier for them to access it themselves online. This analysis requires objective data on input measures, such as GPs' prescribing behaviour. This is not yet possible in education due to the lack of data about what really happens in classrooms.

Education

The importance of peer to peer dissemination of evidence is also evident in education. For example, the TES Teaching Resources site has 3.2 million registered online users from 276 countries and claims to be 'the world's largest online network of teachers'.[371] However, these resources have not been rigorously tested as medical interventions must. If a GP learns about an innovation from a colleague, he or she can 'pull' from the medical literature studies about its effectiveness. In education, with no similar evidence (or at least not readily available), peer to peer dissemination can be unreliable or even dangerous.

7.3.9 Information Architecture

7.3.9.1 Health

Good quality information architecture, such as PubMed's searchable database of articles and abstracts, ensures that users can find evidence. Improved 'searchability' on websites such as that of the *BMJ* and the NIH also improves the information architecture of the evidence system, driving dissemination.[372] The Cochrane Collaboration provides free access to new or updated reviews after 12 months of publication[373] as of February 2013: it was also reportedly one of the first bodies to use an electronic format for its reviews, which enabled immediate access.[374]

Increasingly, funders of health research require articles to be published on an open access basis.[375] †

The Max Planck Society, Howard Hughes Medical Institute and the Wellcome Trust have recently launched a new open access journal for life and biomedical sciences research called *eLife*. This is in addition to a large number of established high-quality open access journals, such as *PLOS Medicine*.[376]

The WHO website offers a free, searchable database of baseline evidence, with graphic and narrative analysis – making its baseline evidence easily understood not only by clinicians, but also by policy-makers and other non-clinicians.[377]

Important problems with information architecture (noted by Paul Simpson, Senior Editor of *PLOS Medicine*) include:

- Access: content is held behind paywalls and/or journals that provide conditional access to evidence, such as restricting access to papers for six months after publication[378]
- Licences that don't enable reuse of content, even for researchers to publish or distribute their own research
- Often the data behind journal articles are often not readily available to other researchers.
- Limitations to search engines and databases. For example, there is a bias to English language publications in PubMed, and several studies have shown that the main bibliographic databases have varied and incomplete coverage of material that is not written in English.[379]

7.3.9.2 Education

The information architecture in education is rather weaker.

For academic research, there are registries of RCTs, such as the American Economic Association's which is open to all economists.[380] Open access journals are reportedly rare in the social sciences.[381] Journal abstracts are reportedly much less structured in education and development economics than in medicine, making it harder to locate studies on specific topics.

There is no system at all for research completed by NGOs, practitioners or other non-academics. There is nowhere central for NGOs to deposit their monitoring and evaluation research (The World Bank is creating a store of all studies with a robust counterfactual, but most NGO-generated research doesn't have this so wouldn't be eligible.) The effect of this is that brilliant discoveries by NGOs or other non-academics may not spread because there are no reliable dissemination channels.

---

† Examples include
http://www.researchinfonet.org/publish/finch/
http://www.rcuk.ac.uk/research/outputs/
http://www.wellcome.ac.uk/About-us/Policy/Spotlight-issues/Open-access/Guides/

Some research is not published at all, perhaps especially 'monitoring', which is essentially compliance. Some is published on organisations' own websites or documents. It is published in whatever format the organisation chooses.

This means that it's very hard to find NGO-generated research (either to assess its quality, or its findings, or even just to map what's going on in order to avoid duplication) or synthesise (e.g., data published in PDFs must often be re-entered manually if they are to be used). It becomes 'grey literature'.[382]

An interesting approach to prevent this has been adopted by FB Heron Foundation. It operates solely domestically in the US but requires all entities it funds to put their data into a public data-warehouse so that anybody can see it and any funder can use it: no reporting data is collected or required solely for FB Heron. "We don't want any data captive within the walls of the foundation" says Clara Miller, the foundation's president.[383]

> **Emerging recommendation 12:**
> Make education research easier for users to 'pull' (find). This relates to recommendations 5 and 6, of publishing it in a findable place, and with standard searchable sections (as medical journal articles have standardised abstracts).

7.3.10 Translation of evidence
7.3.10.1 Health
Evidence needs to be 'translated' from the jargon-heavy language of academia into a useable format for most users.[384] This is essentially a usage device to encourage and enable practitioners to use evidence. In health, CPGs are produced by national and international medical institutions or governmental bodies with the aim of synthesising the evidence on health interventions into practical procedures for clinicians. CPGs are discussed in detail in Chapter 8.6.1, where we discuss overcoming barriers to using evidence.

Similarly, SRs are processed into 'user friendly summaries', produced by the Australasian Cochrane Centre Policy Liaison Initiative and Evidence Aid amongst others.[385]

7.3.10.2 Education
There is little translation of evidence in education, which we discuss further in the 'Use' section of this report.

7.4 Conclusion

Disseminating evidence in education is currently difficult. There is little research to disseminate, few incentives for researchers to disseminate it in a useful form, and insufficient funding for dissemination. It is hard for practitioners in education to find information on a 'pull' basis because results are often not publicised and/or not published in a way that search terms can find them. However, it seems to us that in principle it would be relatively simple to improve the information architecture in education. One could augment the 'push' mechanisms such as conferences, specialised journals and grand rounds, and make 'pull' easier by centralising and standardising reports by non-academics.

8 Use of evidence in health and education

8.1 Why is it important to look at how evidence is used?

The aim of an evidence system is to affect the behaviour of practitioners. This requires that evidence be produced, and we've seen the serious lack of evidence about education. But this isn't sufficient: there is also an important task to get people to use it.

As mentioned in the introduction, Richard Thaler makes two observations about encouraging evidence-based practice:

- 'You can't make evidence-based policy decisions without evidence.
- If you want to encourage some activity, make it easy.' [386]

Chapter 6 on production of evidence looked at the first issue. This section will address how health and education evidence systems can make it easy to find and use evidence.

8.2 What are the main differences between use of evidence in health and education?

Perhaps the reason that evidence is used more in health than in education is that health has better overcome the barriers that inhibit using evidence. MIT professors Abhijit Banerjee and Esther Duflo identified three such barriers in their award-winning book "Poor Economics": **inertia** (in which category we include behavioural and cognitive biases, discussed later), **ideology and ignorance**.[387] Health uses various devices to help users (not just clinicians, but also policy-makers, funders etc.) to overcome these barriers:

1. CPGs, which condense the findings from the latest studies into a form which is useful to busy practitioners.
2. Delineating diseases and conditions into a binary 'yes' or 'no' form, to help clinicians make treatment decisions.
3. Creating checklists, such as for surgical procedures. These are essentially a 'make it easy' nudge.

So far, education doesn't have 'practice guidelines'. The nearest thing is a curriculum, but this doesn't tell practitioners *how* to teach to achieve the best outcome based on evidence: it just lays out the material to be covered. Users are not provided with checklists, and educational 'conditions' are not generally delineated into binary forms, for example, pupils who do or don't need extra help in a subject.

8.3 Who uses evidence, and how?

Both health and education cater to a range of users of evidence, who vary in their incentives and motivations to use evidence, and the barriers they face in doing so:

Table 6: Users of evidence in health

User of evidence	Example issues on which they might use evidence
Practitioners e.g., doctors, nurses	Clinical or care interventions.
Health service managers	Distributing hospital resources amongst clinical departments.
Local policy-makers	Identifying appropriate locations for hospitals and primary care facilities.
National policy-makers	Allocating spending between diseases based in part on incidence and mortality data.
Funders of health research	Identifying interventions which effectively reduce the incidence or mortality of a particular disease.
International organisations e.g., the WHO	Monitoring levels of infectious diseases and defining priorities.
Industries associated with healthcare e.g., pharmaceutical companies	Making decisions about drug development.
Academics and researchers	Producing SRs.

Table 7: Users of evidence in education

Audience/User of evidence	Example issues on which they might use evidence
Practitioners e.g., teachers, head teachers Programme managers in non-governmental organisations (NGOs) delivering educational interventions	Interventions to improve classroom learning, e.g., teachers in Shanghai and Singapore participate in 'journal clubs', where research is assessed and discussed.[388]
Teacher trainers	What to teach e.g., evidence about teaching methods. How to teach teachers, e.g., evidence about what influences their behaviour.
Local policy-makers	How to incentivise teachers to reduce absenteeism; the value of providing free school lunches.
National policy-makers e.g., education ministries	Whether to have small, local schools or larger, regional schools.
International organisations e.g., the United Nations Children's Fund (UNICEF)	Prioritising countries and issues for focus, based on the baseline data, e.g., around literacy rates.[389]
Funders e.g., the World Bank	Balance of spending between countries (based on baseline data), and between interventions and the stipulations attached to grants and loans (based on evidence of their effectiveness).
Industries supplying equipment e.g., computers, books	Making decisions about types of equipment to supply.
Academics and economists	Use of evidence as the basis for future studies.

8.3.1 Education

Few potential users of evidence in education (teachers, schools, policy-makers, programme managers etc.) use evidence effectively. For example, **some of the largest funders of education programmes in LEDCs have indicated that they do not review evidence before making their decisions**. Similarly programme managers of some large programme said that they use their own findings on the effectiveness of certain interventions, but do not seek evidence from elsewhere.[390]

In contrast to health, there is no educational agency analogous to NICE in the UK, which requires practitioners to use evidence-driven interventions. The What Works Clearing House in the US has begun to assemble data on intervention effectiveness, as has the Education Endowment Foundation (EEF) in the UK, but neither has regulatory authority as NICE does.[391 392]

Baseline data could and should be used in several ways. They can help funders, policy-makers and practitioners prioritise issues. They can highlight gaps in provision (such as materials and equipment) by districts, regions and countries.

Evidence on intervention effectiveness obviously can help funders and practitioners prioritise parts of the education system on which to focus, and which interventions to use.

8.4 What are the incentives for practitioners to use evidence?

8.4.1 Health

Apart from the obvious desire to serve patients well, performance monitoring is an important incentivise for doctors and healthcare providers. Monitoring and accountability mechanisms in the UK include regulation by organisations such as NICE, and 'naming and shaming' by organisations such as Dr Foster.[393] For example, the John Radcliffe Hospital in Oxford was investigated by the regulator in November 2005 after it transpired that the number of patients who died between April 2002 and March 2005 after their first coronary artery bypass graft was more than double the national average.[394][395] A second incentive to follow the evidence comes from patients' increasing ability to look up evidence online and their willingness to complain to regulators and government about standards of care, and perhaps prosecute for clinical negligence.[396]

There are reportedly four elements that must be in place to ensure successful performance monitoring: good data infrastructure; trust in the process by clinicians and patients; training in using performance monitoring systems; and a small number of performance measures to track.[397]

8.4.2 Education

Given that very little evidence is produced or disseminated effectively in education, it is very hard for practitioners to use evidence. Some NGOs are incentivised by funders to use evidence, although as stated many funders themselves do not use evidence in decision-making.[398] Perhaps one incentive for policy-makers and national governments to use baseline evidence is the negative publicity around performance. For example, Brazil's very low PISA scores in 2009 led to a renewed government focus on using evidence about effective interventions to raise learning levels.

8.5 What are the barriers to using evidence?

The barriers of inertia, ideology and ignorance cited by Banerjee and Duflo apply to health and education.[399]

8.5.1 Inertia

It is difficult for anyone to change their behaviour even if they are rationally convinced of the reasons to do so, because of cognitive biases. Kahneman discusses a range of cognitive biases that limit the use of evidence. These arise because individuals operate primarily in what Kahneman describes as their System 1: they think fast, fail to analyse situations properly, and make judgements using rules of thumb. These lead to predictable errors ('cognitive biases'). Activating their System 2, thinking which is slower and more accurate, is difficult and normally physically somewhat painful.[400] The resulting cognitive biases include:

Confirmation bias makes us all more likely to believe evidence that re-enforces our existing views. We doubt information which conflicts with our existing view – often so much so that we literally don't register or remember it at all. The cognitive dissonance that it creates is painful, so we ignore it.[401] Education programme managers have reported that teachers are likely to advocate 'trusting their gut'[402] or have a theory about what works in their classroom and trust the data that fit with this theory.[403] For example, many people think that girls skip school because of difficulty managing their periods, and therefore many programmes build latrines and so on. In fact, the one rigorous study

into this issue (by J-PAL) indicates that a lack of appropriate facilities for girls to use during menstruation has only a small impact on absenteeism.[404]

In **availability bias**, 'personal experiences, pictures and vivid examples are more available [and hence more likely to influence behaviour] than…mere words or statistics'.[405] This bias is also reflected in comments by programme managers that practitioners often look to case studies and anecdotes as a basis for their practice, instead of data from intervention effectiveness studies. Given the lack of evidence, this of course might be unavoidable at the moment.[406]

In the **illusion of skill**, people assume they are more skilled than they really are. For example, in a study, clinicians were asked for their confidence in their diagnoses at the time they made them, and these diagnoses were later compared to autopsy results. The study found that clinicians who were 'completely certain' of the diagnosis were wrong 40 per cent of the time.[407]

Other biases include **risk aversion** and the impact of '**framing**' (how questions and situations are described) on decision-making.[408]

Applying 'averages' to individuals

Some practitioners are unwilling to accept the findings of research on the basis that 'average' results do not apply to their patients. A study into the public perceptions of doctors in the US found that 'decisions about how to care for patients are often made on the basis of local custom, conjecture, and the personal experience of individual physicians, and invasive operations can diffuse into widespread use on the basis of no hard evidence.'[409]

Rate of adoption depends upon ease of adoption

Adoption is slow, even in medicine. Senior doctors in the UK believe that it takes on average about 16 years for a proven innovation to be widely adopted.[410]

Gawande also suggests that two factors influence the speed at which evidence-based innovations are adopted: first, demonstrating significant and observable results; and second making life easier for the practitioner.[411] He compares anaesthesia, which was rapidly adopted by clinicians, to early sterilisation techniques, which took decades to be adopted. Anaesthesia made surgery easier for doctors (a screaming patient on the operating table is unpleasant for the surgeon), and offered immediate and easily observed results. By contrast, sterilisation with carbolic acid made life harder for doctors without readily perceptible benefit.[412]

Most educational innovations fail either or both criteria. On the first, educational attainment is rarely visible, particularly in a world with few baseline data and where the outcomes (i.e., definition of attainment) are disputed anyway. On the second, interventions rarely make life easier for practitioners. Furthermore, changing performance is hard: teachers would need to be re-skilled, teacher training college syllabi would need to be changed and the exam system would need to be redesigned.[413]

Sustaining changes

Beyond innovation, there is difficulty in maintaining new behaviours. Gawande discusses the 'rule of seven touches', which says that people need seven reminders of appropriate behaviour before they can maintain it.[414]

Legal challenges

In health, the increasing number of legal challenges to medical interventions discourages using the most up-to-date research if the implication is to depart from conventional practice. In some countries, compliance with common practice is a strong legal defence against claims of negligence, irrespective of the evidence.[415]

A sense of practitioner ownership

In education, there is little input from practitioners into the production of evidence, partly due to the separation of practice and research in education. As a result, many practitioners do not feel a sense of ownership over the research findings, and so are less likely to use them.[416] For example, the evaluation of the World Bank's research between 1998 and 2005 highlighted several flaws in its processes.[417] The report criticised the fact that research topics were often not relevant to programme staff, that researchers were not perceived as credible by staff and that users of research were often unclear about how to implement research findings.[418]

In education, one palpable example of inertia is **absenteeism. This is a significant barrier to using evidence – simply because many teachers are not present in the classroom to teach at all.** While practitioners might not use evidence in an ideal way, some are operating in difficult circumstances, with 60 per cent of teachers absent and practitioners only being paid sporadically.[419] Reducing teacher absenteeism is an obvious priority for research, since teachers can't teach if they're not in the room.

> **Emerging Recommendation 13:**
> (As per recommendation 8): Involve users in prioritising research topics, to ensure that research is relevant. This would secure buy-in from practitioners, making it more likely that they use the research, and reduce wastage**.**

8.5.2 Ignorance

8.5.2.1 Health

Despite the availability of structured abstracts, databases and systematic reviews (all of which are crucial to using evidence in medicine), clinicians still feel a lack of time and competence to search for, identify, read, assess and apply the evidence for each patient.[420] This barrier might apply particularly to users who require a broad overview of the evidence base, such as GPs or policy-makers.

Challenges in processing the amount of available evidence was already an issue in the 1990s, when a study found that 'research produces important evidence about the benefits and harms of healthcare but there is a large gap between what the evidence shows and the care that most patients receive. The gap is broadened by the extensive processing that evidence requires before it can be used.'[421]

Clinicians also face a trade-off between making the wrong decision based on low-quality evidence or waiting for more evidence by postponing a health intervention. Some SRs are inconclusive (as the evidence may genuinely be mixed), which prevents practitioners using them.

8.5.2.2 Education

Many users of evidence cannot integrate research into their practice because they don't know that the evidence exists. This ignorance arises partly from the lack of 'push' dissemination mechanisms, and partly because the users struggle to 'pull' because it is only published on organisations' own

websites and in unreadable or incomparable formats. Furthermore, practitioners may be unsure of which research to trust because they do not know how to assess its quality and rigour: some programme managers reported this.

Ideally, research users would either be able to see the quality of research by being trained to assess it themselves, or by being provided with independent assessments of its quality in the way that NICE, CPGs and the EEF Toolkit provides. As a first step, we recommend bringing together the research done by academics, NGOs, practitioners and any other entity, such that anybody qualified could assess its quality and/or conduct systematic reviews to see the answers. This increases the visibility of quality for users 'pulling' evidence.

> **Emerging recommendations 14:**
> Produce guidance and training for education funders about the various types of evidence and their differing strengths and weaknesses.

8.5.3 Ideology

In some cases the ideological perspectives of governments or teacher trainers (such as political beliefs around caste in India) might prevent practitioners from accepting research findings.

Practitioners in education are reportedly ideologically less inclined to use evidence than people in health,[422] and there is some aversion to measurement among teachers and teaching unions.[423] For example, a teaching union in Kenya took a dispute over a study from J-PAL to the Kenyan High Court because it showed that contract teachers performed better than salaried teachers, which ran contrary to the interests of the union.[424] Teachers in Mexico went on strike over their 'right' to pass their jobs to their children and resistance to performance appraisals.[425]

> # Box 15: Is teaching a last resort in LEDCs?
>
> In his study on teacher provision in Anglophone Africa, Professor Aidan Mulkeen, Vice-President Academic and Deputy President at the National University of Ireland Maynooth, found that the academic performance required to enter teacher training is low, and in some cases declining.[426] The profession's poor reputation means that teacher training tends to attract school leavers who failed to get places on other higher education courses.[427] Mulkeen also notes that teaching courses particularly struggled to attract students with qualifications in mathematics and science.[428]

In some countries, the legislative calendar can impede changing the educational curriculum even if evidence suggests that changes are needed: some countries' constitutions only allow for the curriculum to be changed at certain times of the year. In other instances party political pressures trump the evidence when politicians make decisions about education practice.[429] Other barriers to using evidence include the fact that policy-makers rarely have a background in education or educational research, and so arguably do not have the depth of knowledge to use evidence in an appropriate way.

8.6 How has health overcome these barriers?

8.6.1 Clinical Practice Guidelines (CPGs)

CPGs have been introduced in health to guide decisions about diagnosis, treatment, and management of specific issues. These are aids to using evidence: they spare practitioners from continually referring to research, and they are deliberately concise and attractively designed.[430]

CPGs are produced at national and international levels by medical institutions, professional bodies and governmental bodies. For example, the US National Guideline Clearinghouse currently has 2498 guideline summaries on its database, many of which have been produced by specialist academies such as the American Academy of Orthopaedic Surgeons.[431] The WHO currently holds approximately 130 guidelines on its website.[432] Guidelines are funded by national governments (as in the case of the UK and US) and professional colleges. Where guidelines from different bodies are contradictory (as in the case of UK guidelines on drinking during pregnancy), clinicians will look at the most recent CPG, the most thorough CPG or at their institution's 'practice terms' which might provide an indication as to which CPG to follow.[433]

Most of the institutions responsible for developing CPGs disseminate them in clearinghouses and guideline libraries. The two largest libraries are the National Guidelines Clearinghouse and the international Guidelines International Network (which runs the International Guideline Library), based in the US. CPGs are also published in journals, released on relevant organisations' websites and included in medical databases. Similarly NICE produces regular 'evidence updates' on many conditions and areas of research.[434]

CPGs relate to the area of 'translational research', which has two meanings. First, the 'translation of new knowledge, mechanisms, and techniques generated by advances in research into new approaches for prevention, diagnosis, and treatment of disease'[435] and second, translating the research into practical changes to healthcare for patients.[436]

Where clinicians need to make decisions about interventions before guidelines exist (the 'care gap' between evidence and medical practice described above), a two-step process was proposed by an expert working group for tuberculosis that convened for a meeting at the Gates Foundation. Technical recommendations based on low-quality evidence that can be produced quickly can be first released to practitioners, followed by recommendations based on rigorous studies that are completed after scale-up.[437]

NICE has produced a handbook outlining its process for creating CPGs, which involves: (a) defining of the scope of the CPG; (b) selecting the members of a development group who will be producing the CPG; (c) identifying the evidence through literature searching; (d) reviewing the evidence; (e) assessing cost-effectiveness; (f) writing the CPG; and (g) finalising, publishing and offering implementation support for the CPG.[438] The contents page of a CPG produced by the UK Royal College of Nursing for the 'assessment and prevention of falls in older people' illustrates the type of information usually included in a CPG.[439] Notably, the guideline includes information about how it was developed, what is and is not included, the process for dissemination and a timescale for update and review, which allows for new evidence to be incorporated into the CPG.[440]

Other national governments and health services often adopt CPGs from more established institutions.[441] For example, The Adapte Collaboration was originally founded with the aim of adapting

French CPGs on the treatment of cancer for Quebec. The organisation now works as an international collaboration of researchers and guideline developers who promote the effective adaptation of international guidelines to other localities.[442]

Figure 24: Contents page of Royal College of Nursing CPG for the assessment and prevention of falls in older people, 2004[443]

Contents

NICE produces a range of 'clinical pathways', relating to their guidelines, which give a simple visual guide on the most appropriate way to diagnose a condition and choose a clinical intervention.

Figure 25: Example of a NICE clinical pathway: Self-limiting respiratory tract infections – antibiotic prescribing overview[444]
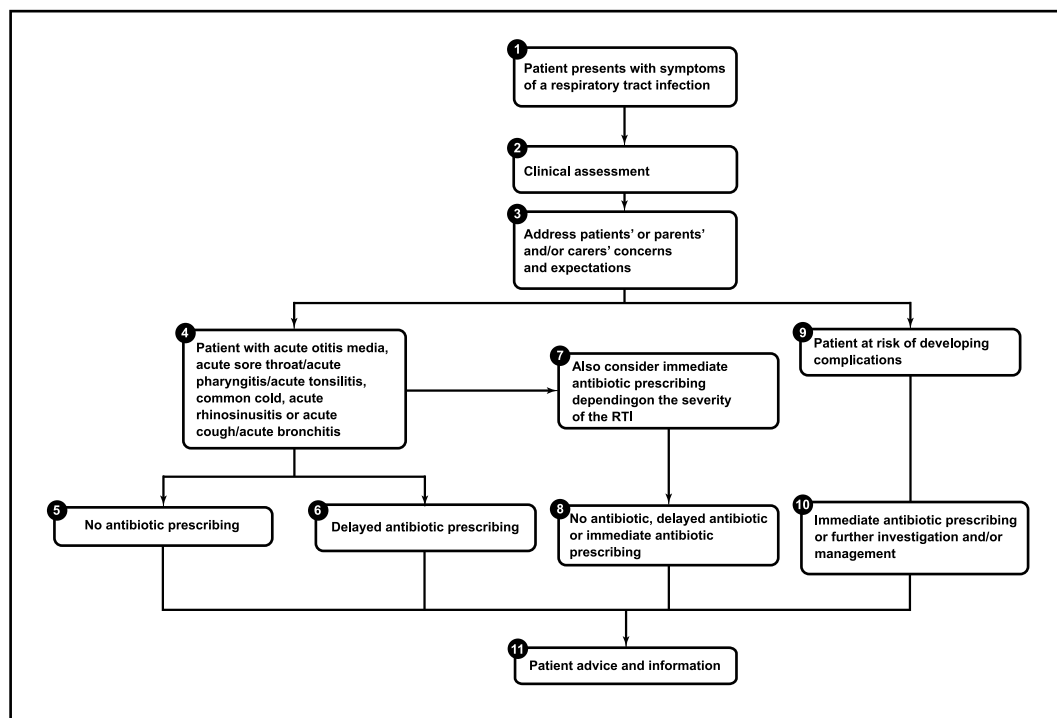
The US sees significant public resistance to the idea of clinicians using guidelines based on research, for fear that they will prevent care tailor to the individual. Many patients trust their doctor's judgement, irrespective of the evidence.[445] Many patients distrust published evidence on the basis that it reflects an 'average patient' rather than them, and/or is used to ration or withhold treatment.[446] The latter issue is sensitive, as illustrated by former vice-presidential candidate Sarah Palin's coining of the phrase 'death panel' in relation to NICE, because it bans some interventions on the basis of safety or cost-effectiveness.

Goldacre suggests that education research findings are rarely adapted or interpreted for users, stating that 'while individual studies are written up in very technical documents, in obscure academic journals, these are rarely read by teachers. And rightly so – most doctors rarely bother to read technical academic journals either. The *British Medical Journal* has brief summaries of important new research from around the world; and there is a thriving market of people offering accessible summary information on new 'what works' research to doctors, nurses, and other healthcare professionals'. [447]

While, for example, J-PAL's policy summaries, designed to provide implementation guidance on the basis of its trials, provide some guidance to policy-makers, programme managers and practitioners, there are still no comprehensive educational 'practice guidelines'. Education also does not have an equivalent repository to the What Works Clearing House in the US, which uses an expert panel to decide what evidence to accept.

Educational 'practice guidelines' may be easier to develop for some users of evidence than others. Funders, programme managers, NGOs and policy-makers might all benefit from broad-reaching guidelines supporting decisions on, for example, interventions to improve literacy or reduce absenteeism. It may be more challenging to draw up guidelines for teachers that can be used in multiple countries or contexts, given differing regulation and legal frameworks. However, just as in health, one way to overcome this might be to establish groups of researchers, policy-makers and practitioners at a national or local level to adapt international guidelines for their specific context.

Guidance and/or training on types of evidence and what constitutes high quality evidence might be useful for funders, NGOs and practitioners. This is a low cost and easy way to improve these groups' ability to commission, produce and interpret evidence.

8.6.2 Checklists

Checklists are easy-to-understand guides for practitioners, designed to standardise practice and limit human error as much as possible.[448]

Gawande argues that checklists help practitioners to overcome 'errors of ineptitude', which happen because people have not applied what has been learnt to a given situation. This is in contrast to 'errors of ignorance', which are mistakes due to a lack of information.[449] He points to a range of examples that show how simple tasks are often not completed, even by highly intelligent people with extensive training. For example, 'in 2005, Columbus Children's Hospital [in the US] found that more than one-third of its appendectomy patients failed to get the right antibiotic at the right time'.[450] When tracking the performance of six basic surgery safety steps in eight hospitals, 'on average, hospitals missed one of them in a startling two-thirds of patients'.[451]

Gawande and Dan Boorman of Boeing (an aeronautics company which uses checklists extensively to limit errors) have since co-developed a surgical safety checklist for the WHO (see Figure 26).[452] A study into the effectiveness of the checklist showed a significant reduction in mortality and morbidity compared to the control group.[453] Gawande is now leading the development of further checklists at the WHO Patient Safety programme.[454]

Other examples of checklists' effectiveness in both LEDCs and MEDCs are listed by Gawande. Columbus Children's Hospital created a 'Cleared for Take-Off' checklist for surgery, and 'after three months, 89 per cent of appendicitis patients got the right antibiotic at the right time. After ten months, 100 per cent did.'[455] Poor households in Karachi were given free soap and directed to use it in six situations, such as preparing food and wiping a child. In houses that used the checklist, diarrhoea fell by 52 per cent compared to the control group, pneumonia fell by 48 per cent, and impetigo (a bacterial skin infection) fell by 35 per cent. This was despite issues around illiteracy, poverty, overcrowding and drinking and washing with contaminated water. It emerged that many people already had soap, but the checklists made its use more systematic.[456]

Figure 26: WHO surgical safety checklist[457]

As far as we know, no checklists are used in education. However, given their effectiveness not only in medicine but also in other industries, such as construction and aeronautics, they could be an important way of integrating evidence into education. Gawande and Boorman developed a 'checklist for checklists' (see Figure 27), which could be used in education to help develop checklists for processes such as teaching children to read.

Emerging recommendation 15:
Investigate mechanisms to make it easier for practitioners to consume, understand and use research. These might include 'practice guidelines' (analogous to doctors' clinical practice guidelines). They might also include checklists. Though terribly simple, checklists seem to be highly effective.[458]

Figure 27: A checklist for checklists[459]
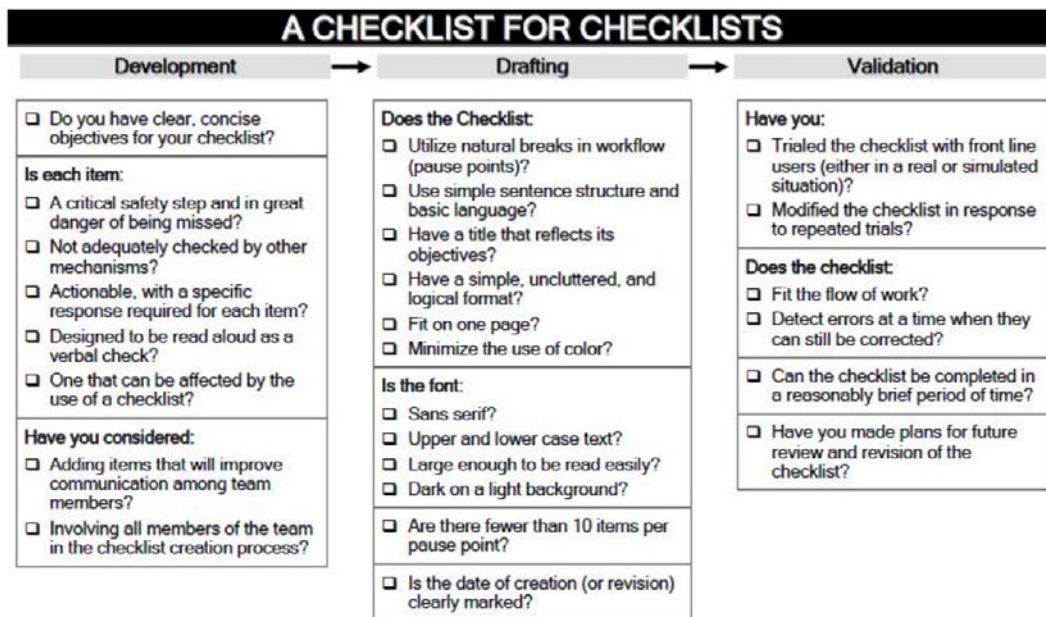
8.6.3 Delineating binary conditions to help practitioners choose appropriate treatment

Guidelines and other recommendations are also produced in health to provide a binary definition of what it means to have a condition. For example, severe hypertension (high blood pressure) is defined as a condition where 'systolic blood pressure is 180 mmHg or higher, or clinic diastolic blood pressure is 110 mmHg or higher'.[460] This makes it easier for doctors to make diagnoses and hence prescribe treatments.[461] Without this 'binary-fication' there would be less consistent use of interventions, as diagnosis would be subject to the personal opinion of practitioners.

Figure 28: Definition of severe hypertension in health[462]
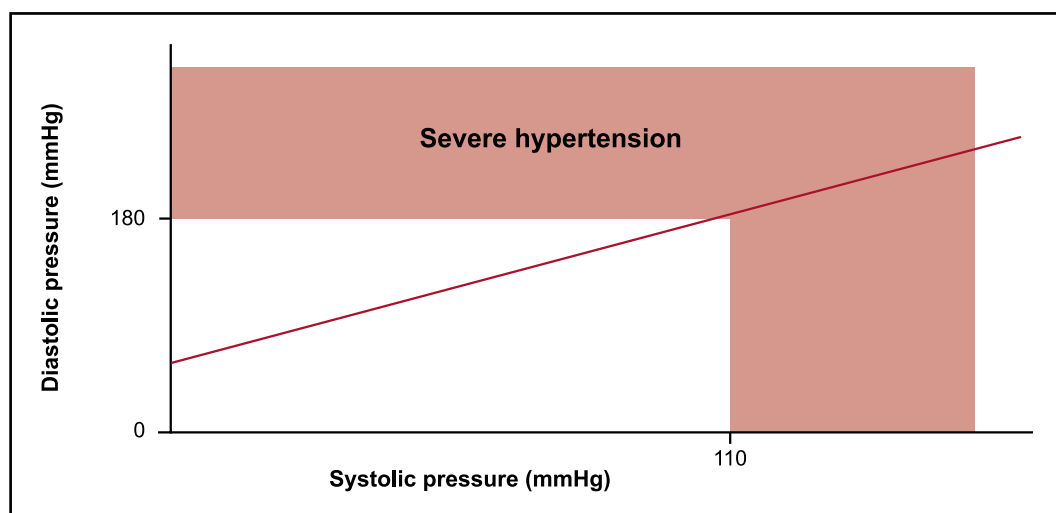
As far as we know there has been no 'binary-fication' of educational standards to help users of evidence identify when certain interventions should be used.[463] One example might be to provide a cut-off point in attainment, below which a school should be closed.

> **Emerging recommendation 16:**
> Explore delineating binary 'conditions' in education (e.g., either a child needs remedial maths tuition or they don't) to help teachers make decisions (analogous to how doctors are aided in diagnosing conditions by binary delineation: either you have diabetes or you don't).

8.7 Conclusion

As in dissemination, it is very difficult for practitioners and others in education to use evidence when there is so little available. The lack of intermingling between users and producers of evidence, and some practitioners' inertia, ignorance and ideological resistance to using evidence also limits the use of the little research that does exist.

The cognitive and practical barriers faced by education practitioners are encountered in many sectors, including health. Education can benefit from the tools that have been created to improve the diffusion of new evidence and aid behaviour change, such as checklists and guidelines. It is likely that as EBM and other evidence-based sectors progress, these tools will be refined and enhanced, particularly as our understanding of cognitive processes improves.

Appendices

Appendix 1: List of interviewees

- Amrita Ahuja, Research Fellow, Sustainability Science Program, Harvard University
- Colin Almeleh, CIFF
- Uzma Anzar, Senior Research Manager, University Research Co., LLC
- Oscar Arcos, CIFF
- Rukmini Banerji, Director of ASER Centre, Pratham
- Sir Iain Chalmers, Co-founder, The Cochrane Collaboration, the James Lind Alliance, the James Lind Library
- Madhav Chavan, CEO, Pratham
- Mike Clarke, former Director of the UK Cochrane Centre, former Co-Chair of the International Steering Group for The Cochrane Collaboration, now Director of the All–Ireland Hub for Trials Methodology Research
- Arri Coomarasamy, Professor of Gynaecology, School of Clinical and Experimental Medicine, Birmingham University
- Katherine Cowan, Senior Adviser, James Lind Alliance
- Lisa Deyo, Project Director, University Research Co., LLC
- Annie Duflo, Executive Director, IPA
- Colin Felsman, Senior Program Associate, Results for Development Institute
- Max Gasteen, Education Research Adviser, DFID
- Paul Glewwe, Professor of Applied Economics at the University of Minnesota and Director of the Center for International Food and Agricultural Policy
- Jonathan Grant, then President of RAND Europe
- Anna Hakobyan, CIFF
- Lucy Heady, Performance Measurement and Effectiveness, CIFF
- Megan Heggarty, Co-ordinator, International Education Funders' Group
- Robin Horn, Director of Education, CIFF
- Sharath Jeevan, Director, STIR
- Kim Kerr, International Development and Education Consultant
- Jessica Kiessel, Chief Program Officer, Innovations for Poverty Action
- Elizabeth King, Director of Education Human Development Network, World Bank
- Peter Laughan, Director, Firelight Foundation
- Michelle Levene, London School of Hygiene and Tropical Medicine
- David Lissauer, Clinical Lecturer in Obstetrics and Gynaecology, School of Clinical and Experimental Medicine, Birmingham University
- Mieke Lopes Cardozo, Lecturer, Faculty of Social and Behavioural Sciences, University of Amsterdam
- Geoff Murphy, Director, Economic Development Program, Wellspring Advisors
- Alex Murray, Chief Superintendent of West Midlands Police and Founder of the Society of Evidence Based Policing
- Karthik Muralidharan, Assistant Professor of Economics, University of California, San Diego
- Milena Novy Marx, Visiting Assistant Professor of Political Science, University of Rochester
- Sandy Oliver, Deputy Director, Social Science Research Unit, Institute of Education, University of London
- Eric Patashnik, Professor of Public Policy and Politics, University of Virginia

- Harry Patrinos, Lead Education Economist, World Bank
- Shawn Powers, Education Initiative Manager, J–PAL
- Caine Rolleston, Education Research Officer, University of Oxford
- Balaji Sampath, Founder, EurekaChild
- Justin Sandefur, Research Fellow, Center for Global Development
- William Savedoff, Senior Fellow, Center for Global Development
- Dana Schmidt, Program Officer, Global Development and Population Program, Hewlett Foundation
- Richard Sedlmayr, Philanthropic Advisor, Wellspring Advisors
- Buddy Shah, Founding Partner, ID Insight
- Paul Simpson, Senior Editor, *PLOS Medicine*
- Kathleen Stokes, Policy Advisor, Nesta
- James Townsend, Programme Director, STIR Education Initiative
- Rebecca Winthrop, Director of the Center for Universal Education, Brookings Institution
- Howard White, Executive Director, 3ie
- Chris Whitty, Chief Scientific Advisor and Director, Research and Evidence Division, DFID and Professor of International Health, London School of Hygiene and Tropical Medicine
- Dierdre Williams, Senior Program Officer, Education Support Program, Open Society Foundations

We apologise if we have inadvertently omitted anybody from this list.

Appendix 2: Glossary of abbreviations

3ie	International Initiative for Impact Evaluation
AMSTAR	A Measurement Tool to Assess Systematic Reviews
ASER	Annual Status of Education Report
AusAID	Australian Agency for International Development
BMJ	British Medical Journal
CDC	Center for Disease Control (US)
CEA	Cost effectiveness analysis
CEBM	Centre for Evidence-based Medicine
CIES	Comparative and International Education Society
CPD	Continuing professional development
CPG	Clinical Practice Guideline
CUE	Center for Universal Education
CIFF	Children's Investment Fund Foundation
DFID	Department for International Development
EBM	Evidence-based medicine
EBP	Evidence-based policing
EEF	Education Endowment Foundation
EMA	European Medicines Agency
EPPI-Centre	Evidence for Policy and Practice Information and Co-ordinating Centre
ERIC	Education Resources Information Center
FDA	Food and Drug Administration
GFATM	Global Fund to Fight Aids, Tuberculosis and Malaria
GP	General Practitioner
GPE	Global Partnership for Education
HCZ	Harlem Children Zone
IIEP	International Institute for Educational Planning
IPA	Innovations for Poverty Action
J-PAL	Abdul Latif Jameel Poverty Action Lab
LEDC	Less economically developed country
LSHTM	London School of Hygiene and Tropical Medicine
MDG	Millennium Development Goal
MRC	Medical Research Council
NGO	Non-governmental organisation
NHS	National Health Service (UK)
NICE	National Institute for Health and Care Excellence (UK)
NIH	National Institutes of Health (US)
NIHR	National Institute for Health Research (UK)
OECD	Organisation for Economic Co-operation and Development
PIRLS	Progress in International Reading Literacy Study
PISA	Programme for International Student Development
PSP	Priority Setting Partnership
QALY	Quality-adjusted life year
RCT	Randomised control trial
SR	Systematic review
STIR	Schools and Teachers Innovating for Results
TIMSS	Trends in International Mathematics and Science Study

UNDP	United Nations Development Programme
USAID	US Agency for International Development
UNESCO	United Nations Educational, Scientific and Cultural Organization
UNICEF	United Nations Children's Fund
UIS	UNESCO Institute for Statistics
USAID	US Agency for International Development
WHO	World Health Organization
WISE	World Innovation Summit for Education

Appendix 3: Areas which may yield further insight

Based on the findings from stage one of this project, we suspect that answers to the following questions would provide further insights into improving education by improving its evidence system. We are open to your feedback on this list.

Health
- Do the evidence systems in other parts of health – nursing, physiotherapy, talking therapies, public health, public information / behaviour change campaigns – offer lessons for education?
- Is there anything to be learnt from the way that evidence is used in public as opposed to private health contexts in LEDCs and MEDCs?
- What can studies and evidence about the medical education system teach us about education systems in LEDCs? Health has studied the impact of various teaching techniques and practices, which might be of use to education.
- How exactly did health make the transition to EBM?
- How is qualitative evidence produced, disseminated and used?

Education
- Is there anything in pedagogy that relates to the concepts in health of 'effectiveness' in contrast to 'efficaciousness'?
- How is qualitative evidence produced, disseminated and used?
- How is the research by educationalists produced, disseminated and used? Does that evidence system differ meaningfully from that used by development economists?
- Where does the implementation spend go in relation to the research spend in education?
- Could 'active dissemination' e.g., bringing someone into staffroom or policy meetings to disseminate findings, be more effective than passive dissemination via journals/publications and online channels?
- How do users of education evidence access information about education, and can these mechanisms be used to more appropriately disseminate evidence?
- What have users of evidence in education found helpful and not helpful in terms of receiving and interpreting research?
- How can the culture of education practitioners and users be adjusted to encourage using evidence?

Appendix 4: Some documented problems with evidence in medicine

Publication bias

This is when research which is published does not accurately reflect research that is conducted. Without access to all of the evidence produced, it is virtually impossible for clinicians to make well-informed decisions on the most appropriate interventions to pursue.[464]

Causes include:

(a) Journals' disproportionate interest in trials showing positive or 'blockbuster' results.

(b) Deliberate withholding by pharmaceutical companies of unfavourable results. Goldacre writes that 'around half of all trials for the treatments being used today have gone unpublished; and… trials with positive results are twice as likely to be disseminated'.[465] A study published in July 2013 showed that of the 8,907 trials registered on clinicaltrials.gov over three years, for over half there was no evidence of any results.[466] A study of several hundred cancer trials also found that four out of five trials showed no results at 12 months.[467]

(c) Journals' bias towards publishing results that show large effect sizes.

(d) Journals being more willing to publish trials funded by pharmaceutical companies than by independent researchers, because of the revenue produced from reprints used by sales representatives as marketing material.[468]

(e) Research being withheld by producers as it is counter-intuitive or surprising due to confirmation bias.

Some attempts have been made to overcome publication bias. In 2010 the Nordic Cochrane Collaboration convinced the European Medicines Agency (EMA) to allow the Collaboration to include unpublished data in all of their SRs; this change eventually led to the EMA recommending that the Federal Drug Administration (FDA) in the US and other drug regulators do likewise.[469] The EMA also released over 1.9 million pages of information on clinical studies, due to an EU Ombudsman ruling of maladministration against it (although this ceased in May 2013 due to a legal suit against the EMA by two pharmaceutical companies).[470] The AllTrials campaign seeks to end the withholding of trial data by pharmaceutical companies, and to date over 60,000 people have signed its petition and more than 400 organisations have publicly expressed their support for the cause, including the MRC, NICE and the Wellcome Trust.[471] In March 2013 GlaxoSmithKline, one of the largest pharmaceutical companies in the world, announced its support for the campaign and committed to publishing formal study reports on all of its trials within a year of completion.[472]

False/unrepeatable research

The pressures on researchers to produce new and blockbuster results, driven by the need to secure funding and publish in well-regarded journals, has reportedly led to false results that do not stand up to repeat testing.[473]

In 2005, Professor John Ioannidis, now at Stanford University School of Medicine, showed that the findings of 80 per cent of non-randomised trials and 25 per cent of randomised trials are probably false, because researchers' conscious or subconscious biases influenced the design and/or the interpretation of a study.[474]

Ioannidis found that 'a research finding is less likely to be true when the studies conducted in a field are smaller; when effect sizes are smaller; when there is a greater number and lesser pre-selection of

tested relationships; where there is greater flexibility in designs, definitions, outcomes, and analytical modes; when there is greater financial and other interest and prejudice; and when more teams are involved in a scientific field in chase of statistical significance.' [475]

This finding was borne out by a follow-up analysis, which selected 49 of the most highly regarded findings from 1992 to 2005 (based on their being published in the most-cited journals and being most widely cited in those journals). These studies included well known and accepted treatments such as using hormone-replacement therapy for menopausal women, and using daily low-dose aspirin to control blood pressure and prevent heart attacks and strokes. Ioannidis found that almost half of these studies (41 per cent) were found to be wrong or exaggerated when re-tested.[476]

Varying quality of peer reviews

Many open access medical journals (discussed in more detail in Chapter 8 on the dissemination of evidence) claim to offer a comprehensive peer review system. However, an experiment carried out in August 2013 by John Bohannon, journalist and visiting scholar at Harvard University, revealed that they vary significantly in the quality of peer review. He sent a bogus article full of scientific and ethical irregularities to 304 journals. It was accepted by 157 of them. Bohannon commented that 'Of the 106 journals that discernibly performed any review, 70 per cent ultimately accepted the paper. Most reviews focused exclusively on the paper's layout, formatting, and language.'[477]

Regulatory capture

Goldacre argues in his book "Bad Pharma" that there is a 'revolving door' between the pharmaceutical industry and various regulators, including the FDA in the US, the Medicines and Healthcare products Regulatory Agency (MHRA) in the UK and the EMA. This probably contributes to the weak enforcement of regulation of research by pharmaceutical companies, such as pre-registering trials and publishing all trial results.[478]

Incomparable outcomes

SRs rely on combining the results of multiple trials. Hence it's a problem that many trials define their own outcome measures which confounds comparison and excludes them from SRs.

For example, in 1992 a group of international health professionals interested in rheumatology came together at the first 'Outcome Measures in Rheumatology' conference in the Netherlands. They compared the outcome measures of rheumatoid arthritis used in European trials with those of North American trials. The group noted that each regional group used different endpoints and outcome measures, which made it extremely difficult to compare and combine results in meta-analyses.[479]

A response is the COMET (Core Outcome Measures in Effectiveness Trials) Initiative. COMET seeks to develop 'Core Outcome Sets' for all trials in a specific area of health research, which are agreed minimum sets of outcomes or outcome measures for various areas of study.[480] Agreement on core outcome sets has now been reached in the fields of rheumatoid arthritis, osteoarthritis, osteoporosis, and psoriasis/psoriatic arthritis.[481] Steps have also been taken to find core outcome measures for studies into the treatment of asthma.[482] This idea could be easily incorporated into the education evidence system.

Is the World Health Organization (WHO) following its own guidelines?

The WHO, which produces recommendations and CPGs used in health systems across the world, does not properly use evidence when writing them, according to a 2007 paper in *The Lancet*.[483]

The study showed that 'processes usually rely heavily on experts in a particular specialty, rather than representatives of those who will have to live with the recommendations or on experts in particular methodological areas. There is a tendency to get people around the table and get consensus – everything they do has a scientific part and a political part. This usually means you go to the lowest common denominator or the view of a 'strong' person.'[484]

Appendix 5: Total number of quality studies per input 1990-2010 (adapted Glewwe et al, 2012)[485]

Input	Quality of Study					
	Studies of sufficient quality to be included	Studies of high quality	RCTs	Studies of sufficient quality	Studies of high quality to be included	RCTs
School infrastructure	**Studies measuring effect on test scores**			**Studies measuring effect on students' time in school**		
Textbooks/workbooks	21	8	2	4	4	2
Desks/tables/chairs	8	4	0	0	0	0
Computers/ electronic games	8	6	5	0	0	0
Blackboard/flip chart	6	3	1	0	0	0
Electricity	6	3	0	0	0	0
Roof/wall/floor	4	4	0	2	2	2
Building new schools	0	0	0	3	3	0
Teacher and principal characteristics	**Studies measuring effect on test scores**			**Studies measuring effect on students' time in school**		
Teacher education level	24	6	0	4	4	0
Teacher experience	20	9	0	5	4	0
Female teachers	11	2	0	0	0	0
Teacher training (in service)	11	3	0	2	2	0
Teacher knowledge (test)	9	5	0	0	0	0
Principal education	2	0	0	0	0	0
School organisation	**Studies measuring effect on test scores**			**Studies measuring effect on students' time in school**		
Pupil-teacher ratio	29	14	1	3	3	0
Teacher absenteeism	5	2	0	0	0	0
School provides meals	4	2	1	0	0	1
Salaried teacher	3	0	0	0	0	0
Contract teacher	2	2	1	0	0	0
Cost of attending	2	0	0	4	4	0
Expenditure per pupil	2	0	0	0	0	0

References

1   Cochrane, A.L. (1999) *Effectiveness and Efficiency: Random Reflections on Health Services.* CRC Press. p. 19.

2   Glewwe, P. et al. (2011) 'School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010' *Center for International Food and Agricultural Policy*, p. 45.

3   Brookings (2012) *Research Symposium on Learning* [Online] http://www.brookings.edu/~/media/events/2012/12/06%20learning%20developing%20world/1206%20research%20symposium%20agenda.pdf [Accessed: 11.02.14].

4   Van Fleet, J.W., Watkins, K., and Greubel L. (2012) *Africa Learning Barometer*, a collaboration between the Brookings Center for Universal Education and This Is Africa, a publication of the Financial Times [Online] http://www.brookings.edu/research/interactives/africa-learning-barometer. [Accessed: 21.08.13].

5   Pratham (2012) *Annual Status of Education Report.*

6   UNESCO (2012) *EFA Global Monitoring Report.*

7   Sackett, D.L. et al. (1996) 'Evidence based medicine: what it is and what it isn't', *British Medical Journal*, 312(7023), pp. 71–2.

8   Adapted from Shepherd, J. (2007) 'The production and management of evidence for public service reform', *Evidence and Policy*, 3(2), pp. 231-251 [Online].

9   *Ibid.*

10   Gawande, A. (2011) *The Checklist Manifesto.* Profile Books.

11   Conversations anonymised as requested.

12   United Nations. *Millennium Development Goals & Beyond 2015: Goal 2: Achieve Universal Primary Education* [Online] http://www.un.org/millenniumgoals/education.shtml, [Accessed: 05.05.14].

13   Glewwe, P. et al. (2011) National Bureau of Economic Research, Working Paper no. 17554, *'School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010'*, Center for International Food and Agricultural Policy, January 2012, p. 3.

14   Haynes, L. et al. *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials,* Cabinet Office [Online] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/62529/TLA-1906126.pdf [Accessed: 02.02.14].

15   London School of Economics. (2013) *Advocates of RCTs in education should look more closely at the differences between medical research and education research* [Online] http://blogs.lse.ac.uk/politicsandpolicy/archives/32605 [Accessed: 06.02.14].

16   Dercon, S., UKCDS Town Hall Meeting on Education Systems, 23.01.2014.

17   London School of Economics. (2013) *Advocates of RCTs in education should look more closely at the differences between medical research and education research* [Online] http://blogs.lse.ac.uk/politicsandpolicy/archives/32605 [Accessed: 06.02.14].

18   Handel, A. et al. 'High reprint orders in medical journals and pharmaceutical industry funding: case-control study', *British Medical Journal*, 2012; 344: e4212.

19   Cochrane, A.L. (1999) *Effectiveness and Efficiency: Random Reflections on Health Services*, CRC Press. p. 87.

20   Armstrong, G.L., Conn, L.A., Pinner, R.W. (1999) 'Trends in infectious disease mortality in the United States during the 20th century', *Journal of the American Medical Association.* 281, pp. 61–6 [Online] http://www.cdc.gov/mmwr/PDF/wk/mm4829.pdf [Accessed: 06.02.14].

21   World Health Organization (2013) *Poliomyelitis global annual reported cases and Pol3 coverage, 1980-2012* [Online] http://www.who.int/immunization_monitoring/diseases/poliomyelitis/en/index.html [Accessed: 21.11.13].

22   Australian Vaccination Network (2012), *Reducing the Incidence of Grand Claims* [Online] http://nocompulsoryvaccination.com/2012/12/12/reducing-the-incidence-of-grand-claims/ [Accessed: 21.11.13].

23   Nanda, P. K. (2014) 'Education quality worsens under UPA: ASER' [Online] http://www.livemint.com/Politics/db5Bfb8qSTFhO70WEafBpJ/Indias-education-quality-has-failed-to-improve-finds-ASER.html [Accessed: 11.02.14].

24   Pratham (2012) *Annual Status of Education Report.*

25   The World Bank. *Education* [Online] http://data.worldbank.org/topic/education. [Accessed 07.07.13].

26   Friedrich, M.J. (2013) 'The Cochrane Collaboration Turns 20', *Journal of the American Medical Association*, 309(18), pp. 1881.

27   Goldacre, B. (2013) *Building Evidence into Education* [Online] http://media.education.gov.uk/assets/files/pdf/b/ben%20goldacre%20paper.pdf [Accessed 11.02.14].

28   Gawande, A. (2007) *Better: A Surgeon's Notes on Performance*. Macmillan.

29   Gawande, A. (2013) 'Slow Ideas: Some innovations spread fast. How do you speed the ones that don't?' *New Yorker Magazine*.

30   *Ibid.*

31   Cochrane, A.L. (1999) *Effectiveness and Efficiency: Random reflections on Health Services.* CRC Press. p. 9.

32   Friedrich, M.J. (2013) 'The Cochrane Collaboration Turns 20', *Journal of the American Medical Association*, 309(18), pp. 1881.

33   Harford, T. (2011), TED Talks, *Trial, Error and the God Complex* [Online] http://www.ted.com/talks/tim_harford.html [Accessed: 11.02.14].

34   Dan Gardner. (2013) [Online] http://www.dangardner.ca/articles [Accessed: 11.02.14].

35   Harford, T. (2011), TED Talks, *Trial, Error and the God Complex* [Online] http://www.ted.com/talks/tim_harford.html [Accessed: 11.02.14].

36   Naylor, C.D. (1995) 'Grey zones of clinical practice: some limits to evidence based medicine', *The Lancet,* 345, pp. 840-842.

37   Chalmers, I. et al. (2008) 'The James Lind Library: explaining and illustrating the evolution of fair tests of medical treatments' *The journal of the Royal College of Physicians of Edinburgh* 38, pp. 259–64 [Online] http://www.jameslindlibrary.org/pdf/jll-article.pdf [Accessed: 04.03.14].

38   Goodman, K.W. (2002) *Ethics and Evidence-Based Medicine: Fallibility and Responsibility in Clinical Science*. Cambridge University Press. Foundations and history of evidence-based practice, pp. 1-10.

39   Smith, R., Rennie, D. (2014) 'Evidence-Based Medicine - An Oral History' *The Journal of the American Medical Association*, 311(4), pp. 365-367. [Online] http://jama.jamanetwork.com/article.aspx?articleid=1817042 [Accessed 04.03.14].

40   *Ibid.*

41   *Ibid.*

42   Davidovitch, N., Filc, D. (2006) 'Reconstructing data: Evidence-Based Medicine and Evidence-Based Public Health in context' *Dynamis*, 26, pp. 287-306.

43   Yoshioka, A. (1998). 'Use of randomisation in the Medical Research Council's clinical trial of streptomycin in pulmonary tuberculosis in the 1940s'. *British Medical Journal.* 317(7167), pp. 1220–1223.

44   Davidovitch, N., Filc, D. (2006) 'Reconstructing data: Evidence-Based Medicine and Evidence-Based Public Health in context' *Dynamis*, 26, pp. 287-306.

45   *Ibid.*

46   Howick, J.H. (2011) *The Philosophy of Evidence-based Medicine*. BMJ Books. p. 15.

47   Evidence-Based Medicine Working Group (1992) 'Evidence-based medicine. A new approach to teaching the practice of medicine'. *Journal of the American Medical Association* 268(17), pp. 2420–5.

48   Sackett, D.L. et al. (1996) 'Evidence based medicine: what it is and what it isn't'. *British Medical Journal.* 312(7023), pp. 71–2.

49   The Cochrane Collaboration, *About Us* [Online] http://www.cochrane.org/about-us. [Accessed: 01.10.13].

50   The Cochrane Collaboration, *Evidence-based healthcare and systematic reviews* [Online] http://www.cochrane.org/about-us/evidence-based-health-care [Accessed 13.09.12].

51   Centre for Evidence Based Medicine, *What is EBM?* [Online] http://www.cebm.net/?o=1014 [Accessed: 13.09.12].

52   Eysenck, H.J. (1995) 'Meta-analysis of best-evidence synthesis?', *Journal of Evaluation in Clinical Practice*, 1(1), pp. 29-36.

53   Donald, A. (2014) *Evidence-Based Medicine: Key Concepts* [Online] http://www.medscape.com/viewarticle/430709 [Accessed: 04.03.14] .

54   Claridge, J.A., Fabian, T.C. (2005) 'History and development of evidence-based medicine' *World Journal of Surgery,* 29(5), pp. 547-53.

55   *Ibid.*

56   Donald, A. (2014) *Evidence-Based Medicine: Key Concepts* [Online] http://www.medscape.com/viewarticle/430709 [Accessed: 04.03.14] .

57   Gray, J.A. (1997) *Evidence-based Healthcare*. New York: Churchill Livingstone.

58   Bastian, H., Glasziou, P., Chalmers, I. (2010) 'Seventy-Five Trials and Eleven Systematic Reviews a Day: How Will We Ever Keep Up?', *Public Library of Science*, 7(9), pp. 1-6.

59   Department of Clinical Epidemiology and Biostatistics, McMaster University. (1981) 'How to read clinical journals. 1-Why to read them and how to start reading them critically.' Canadian Medical Association Journal. 124, pp. 555-558 and Sackett, D.L. et al. (1991) Clinical epidemiology, a basic science for clinical medicine. Little, Brown and Company. Second Edition. Boston/Toronto.

60   Guyatt, G. et al. 'Users' Guides to the Medical Literature: A Manual for Evidence-Based Clinical Practice,' 2nd Edition. Foreward [Online] http://jamaevidence.com/resource/520 [Accessed: 01.10.13].

61   Goldacre, B. (2013) *Bad Pharma*. Fourth Estate.

62   Murray, A. (2013) Chief Superintendent, West Midlands Police. Interview with Fiennes, C., 19.09.13.

63   *Ibid.*

64   *Ibid.*

65   College of Policing. *What Works Centre for Crime Reduction* [Online] http://www.college.police.uk/en/20399.htm

66   Neyroud, P. (2012) *Offender Desistance Policing: Operation Turning Point Experiment in Birmingham, UK* [Online] http://www.crim.cam.ac.uk/events/conferences/ebp/2012/operationturningpoint.pdf [Accessed: 16.12.13].

67   Murray, A. (2013) Chief Superintendent, West Midlands Police. Interview with Fiennes, C., 19.09.13.

68   *Ibid.*

69   *Ibid.*

70   Carl J. Jensen III, Ph.D. (2013) 'Consuming and Applying Research Evidence-Based Policing'. *The Police Chief magazine* [Online] http://www.policechiefmagazine.org/magazine/index.cfm?fuseaction=display_arch&article_id=815&issue_id=22006 [Accessed: 16.12.13].

71   *Ibid.*

72   *Ibid.*

73   World Health Organization (2000) *The World Health Report - Health Systems: Improving Performance*
     [Online] http://www.who.int/whr/2000/en/ [Accessed: 11.02.14].

74   Muralidharan, K. (2012) *Priorities for primary education policy in India's 12th Five-year plan*. India Policy
     Forum 2013.

75   Sampath, B. (2013) EurekaChild, Interview with Fiennes, C., 19.06.13.

76   UNESCO. *Education for All Goals* [Online] http://www.unesco.org/new/en/education/themes/leading-the-
     international-agenda/education-for-all/efa-goals/ [Accessed: 11.02.14].

77   United Nations Development Policy and Analysis Division *Preparing for the development agenda
     beyond 2015*, UN [Online] http://www.un.org/en/development/desa/policy/untaskteam_undf/index.shtml
     [Accessed: 16.02.14]

78   United Nations, *The world we want 2015* [Online] http://www.worldwewant2015.org/ [Accessed 16.02.14].

79   United Nations Development Group (2013) *The global conversation begins: Emerging views for a new
     development agenda*.

80   Sampath, B. (2013) EurekaChild, Interview with Fiennes, C., 19.06.13.

81   Einstein, A. Cited at [Online] http://www.woodwardenglish.com/einstein-quote-about-education/
     [Accessed: 11.02.14].

82   Chavan, M. (2014) Speaking at the Skoll World Forum, Oxford. 04.14.

83   Morino, M. (2011) *Leap of Reason: Managing to Outcomes in an Era of Scarcity*. Venture Philanthropy
     Partners. p. 2.

84   Harlem Children's Zone *College Success Office*, HNZ College [Online] http://www.hcz.org/index.php/
     programs/college [Accessed: 11.02.14].

85   Morino, M. (2011) *Leap of Reason: Managing to Outcomes in an Era of Scarcity*. Venture Philanthropy
     Partners.

86   Powers, S. (2013) MIT, Interview with Fiennes, C., 31.05.13.

87   *Ibid.*

88   Baird, S. et al. (2011) *Worms at Work: Long-run Impacts of Child Health Gains* [Online] http://dev3.cepr.
     org/meets/wkcn/7/786/papers/MiguelChildHealth.pdf [Accessed: 11.02.14].

89   Whitty, C. (2013) UK Department for International Development and London School of Hygiene and
     Tropical Medicine, Interview with Fiennes, C., London, 30.07.13.

90   Glewwe, P. (2013) University of Minnesota, Interview with Fiennes, C., 07.07.13.

91   Powers, S. (2013) MIT, Interview with Fiennes, C., 31.05.13.

92   Health Watch. (2009) *The HealthWatch Award 2009: Sir Iain Chalmers* [Online] http://www.healthwatch-
     uk.org/awardwinners/chalmers.htm [Accessed: 11.02.14].

93   Pratham. (2013) Interview with Fiennes, C., 07.13.

94   Goldacre, B. (2013) *Bad Pharma*. Fourth Estate. Afterword: what happened next?

95   Coleman, B., (1999) 'The impact of group lending in Northeast Thailand', Journal of Development
     Economics, 60 pp. 105–141.

96   Freedman, D., (2010) *Lies, Damned Lies, and Medical Science*, The Atlantic [Online] http://www.theatlantic.
     com/magazine/archive/2010/11/lies-damned-lies-and-medical-science/308269/?single_page=true

97   Centre for Evidence-based Medicine (2011), OCEBM levels of evidence working group. The Oxford 2011
     levels of evidence. Oxford: Oxford 2011. Accessed from Multiple Sclerosis Trust (2012), *How the MS
     Trust keeps professionals up-to-date with research* [Online] http://www.mstrust.org.uk/professionals/
     information/wayahead/articles/16032012_08.jsp [Accessed 14.02.14].

98   Bradley, T., Cupples, M.E., Irvine, H. (2002) 'A case control study of a deprivation triangle: teenage
     motherhood, poor educational achievement and unemployment', I*nternational Journal of Adolescent
     Medicine and Health,* 14(2), pp. 117-23 [Online] http://www.ncbi.nlm.nih.gov/pubmed/12467182
     [Accessed: 11.12.13].

99   Concato, J. (2004) 'Observational *Versus* Experimental Studies: What's the Evidence for a Hierarchy?'
     NeuroRx 1(3) pp. 341–347.

100 Haynes, L. et al. *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials,* Cabinet
     Office [Online] https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/62529/
     TLA-1906126.pdf [Accessed: 02.02.14].

101 Clarke, M. (2013) Queen's University, Interview with Fiennes, C., 07.13.

102 Chalmers, I., Glasziou, P. (2009) 'Avoidable waste in the production and reporting of research evidence'.
     The Lancet, 374.

103 National Health Service [Online] http://www.nhs.uk/chq/Pages/1962.aspx?CategoryID=73&SubCategoryI
     D=101 [Accessed 15.5.2014]

104 Goldacre, B. (2013) *Building Evidence into Education.*

105 Centers for Disease Control and Prevention (2014) *About the National Health and Nutrition Examination
     Survey* [Online] http://www.cdc.gov/nchs/nhanes/about_nhanes.htm [Accessed: 11.02.14].

106 World Health Organization, *Data and Statistics* [Online] http://www.who.int/research/en/ [Accessed
     11.09.13].

107 Heneghan, C. (2010) *How many randomized trials are published each year?* [Online] http://blogs.
     trusttheevidence.net/carl-heneghan/how-many-randomized-trials-are-published-each-year [Accessed:
     11.10.13].

108 *Ibid.*

109 Clarke, M., Clarke, L., Clarke, T. (2007) 'Yes Sir, no Sir, not much difference Sir'*, Journal of the Royal Society of Medicine*, 100(12), pp. 571-2.

110 Heneghan, C. (2010) *How many randomized trials are published each year?* [Online] http://blogs. trusttheevidence.net/carl-heneghan/how-many-randomized-trials-are-published-each-year [Accessed: 11.10.13].

111 Jefferson, T. et al. (2013) *Vaccines to prevent influenza in healthy adults* [Online] http://summaries. cochrane.org/CD001269/vaccines-to-prevent-influenza-in-healthy-adults

112 National Collaborating Centre for Methods and Tools (2011) *AMSTAR: assessing methodological quality of systematic reviews*. Hamilton, ON: McMaster University [Online] http://www.nccmt.ca/registry/view/ eng/97.html

113 Bastian, H., Glasziou, P., Chalmers, I. (2010) Seventy-Five Trials and Eleven Systematic Reviews a Day: How Will We Ever Keep Up? *PLOS Medicine* 7(9).

114 Goldacre, B. (2013) 'Are clinical trial data shared sufficiently today? No.', *British Medical Journal* [Online] http://www.bmj.com/content/347/bmj.f1880 [Accessed 11.09.13].

115 Goldacre, B. (2009) *Bad Science*. Harper Perennial. Methods of randomisation, p. 49-50.

116 British Medical Journal (2012) 'High reprint orders in medical journals and pharmaceutical industry funding: case-control study' *British Medical Journal.* 344 [Online] http://dx.doi.org/10.1136/bmj.e4212

117 Goldacre, B. (2013) *Bad Pharma*. Fourth Estate. Afterword: What happened next?

118 *Ibid.*

119 Coomarasamy, A., Khan, K.S. (2004) What is the evidence that postgraduate teaching in evidence based medicine changes anything? A systematic review. British Medical Journal. p. 329.

120 *Ibid.*

121 Levene, M. (2013) Interview with Fiennes, C.

122 Clarke, M. (2013) Queen's University, Interview with Fiennes, C., 07.13.

123 Glewwe, P. (2013) University of Minnesota, Interview with Fiennes, C., 07.07.13.

124 Simpson, P. (2013) Senior Editor of *PLOS Medicine*, Interview with Fiennes, C., 26.07.13.

125 *The Economist* (2014) using source "Us and them: The Geography of Academic Research", *Journal of Development Economics*.

126 Kahneman, D. (2012) *Thinking, Fast and Slow*. Penguin. p 270.

127 Chalmers, I., Dickersin, K., Chalmers, T.C., (1992) 'Getting to grips with Archie Cochrane's agenda', *British Medical Journal*. 305(6857), pp 786–788.

128 Chalmers, I., Glasziou, P. (2009) 'Avoidable waste in the production and reporting of research evidence', *The Lancet*, 374.

129 *Ibid.*

130 Chalmers, I., Dickersin, K., Chalmers, T.C. (1992) 'Getting to grips with Archie Cochrane's agenda', *British Medical Journal*. 305(6857), pp. 786–788.

131 *Ibid.*

132 Clarke, M. (2013) Queen's University, Interview with Fiennes, C., 07.13.

133 *Ibid.*

134 Chalmers, I., Glasziou, P. (2009) 'Avoidable waste in the production and reporting of research evidence'. *The Lancet,* 374.

135 *Ibid.*

136 Goldacre, B. (2008) *Bad Science*. Fourth Estate.

137 The James Lind Alliance. *Welcome to The James Lind Alliance* [Online] http://www.lindalliance.org/ [Assessed: 10.13].

138 The James Lind Alliance. *Priority Setting Partnerships* [Online] http://www.lindalliance.org/Priority%20 Setting%20Members.asp [Accessed: 22.10.13].

139 Cowan, K. (2013) Senior Advisor to the James Lind Alliance, Interview with Fiennes, C., 07.13.

140 National Institute for Health Research. *Evaluation, Trials and Studies Coordinating Centre* [Online] http:// www.netscc.ac.uk/news/item/08042013.asp [Accessed: 17.10.13].

141 Cowan, K. (2013) Senior Advisor to the James Lind Alliance, Interview with Fiennes, C., 07.13.

142 Coomarasamy, A., Lissauer, D. (2013) Professors from Birmingham Women's Hospital, Interview with Fiennes, C., 19.08.13.

143 *Ibid.*

144 Current Controlled Trials, *The AIMS (Antibiotics in Miscarriage Surgery) Trial* [Online] http://www. controlled-trials.com/ISRCTN97143849/ [Accessed: 08.01.14].

145 Coomarasamy, A., Lissauer, D. (2013) Professors from Birmingham Women's Hospital, Interview with Fiennes, C., 19.08.13.

146 *Ibid.*

147 *Ibid.*

148 Whitty, C. (2013) UK Department for International Development and London School of Hygiene and Tropical Medicine, Interview with Fiennes, C., 30.07.13.

149 Center for Effective Philanthropy. *Roland Fryer, speaking at the Center for Effective Philanthropy conference*. Detroit, 21-22 May 2013.

150 Education and Policy Data Center, *About EPDC Data* [Online] http://www.epdc.org/data/datanotes [Accessed: 11.09.13].

151 Omoeva, C. et al. (2013) 'Out of School Children: Data Challenges In Measuring Access to Education'. *Education Policy and Data Center* [Online] http://www.epdc.org/sites/default/files/documents/Out_of_ school_children_data_challenges.pdf, [Accessed: 11.09.13].

152 World Bank. *Data: Education* [Online] http://data.worldbank.org/topic/education [Accessed: 07.07.13].

153 UNESCO. *Data Centre* [Online] http://stats.uis.unesco.org/unesco/TableViewer/document. aspx?ReportId=143&IF_Language=eng [Accessed: 26.11.13].

154 UNESCO, *Education for All Global Monitoring Report* [Online] http://www.unesco.org/new/en/education/ themes/leading-the-international-agenda/efareport/reports/2013/ [Accessed 16.02.14].

155 Global Partnership for Education. *Data* [Online] https://www.globalpartnership.org/results/data/ [Accessed: 20.09.13].

156 World Bank (2006) *Teaching and Doctors: Missing in Action*.

157 Chaudhury, N. et al. 'Missing in Action: Teacher and Health Worker Absence in Developing Countries.' *Journal of Economic Perspectives*, 20(1), pp. 91-116 [Online] http://lingli.ccer.edu.cn/ahe2012/week9/ chaudhury2006.pdf [Accessed: 11.02.14].

158 World Bank (2006) 'Missing in Action: Teacher and Medical Provider Absence in Developing Countries.' [Online] http://econ.worldbank.org/WBSITE/EXTERNAL/EXTDEC/EXTRESEARCH/0,,contentMDK:208248 81~pagePK:64165401~piPK:64165026~theSitePK:469382,00.html

159 Organisation for Economic Co-operation and Development. *Programme for International Student Assessment (PISA)* [Online] http://www.oecd.org/pisa/ [Accessed: 11.09.13].

160 ASER Centre. *Current Research and Assessment Activities in Education* [Online] http://www.asercentre. org/p/134.html?p=119 [Accessed: 11.09.13].

161 Trends in International Mathematics and Science Study. *Countries Participating in TIMSS 2011* [Online] http://timss.bc.edu/timss2011/countries.html [Accessed: 11.09.13].

162 Trends in International Mathematics and Science Study & Progress in International Reading Literacy Study. *Latest Publications from TIMSS and PIRLS* [Online] http://timssandpirls.bc.edu/index.html [Accessed: 11.09.13].

163 International Association for the Evaluation of Educational Achievement, *About Us* [Online] http://www. iea.nl/about_us.html [Accessed: 11.09.13].

164 Uwezo. *About Us* [Online] http://www.uwezo.net/about-us/ [Accessed: 20.09.13].

165 Powers, S. (2013) MIT. Interview with Fiennes, C., 31.05.13.

166 ASER Centre. *Overview* [Online] http://www.asercentre.org/Survey/Basic/Pack/Sampling/History/p/54. html [Accessed: 30.11.13].

167 Glewwe, P. et al. (2011) 'School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010' *Center for International Food and Agricultural Policy*.

168 Glewwe, P. et al. (2011) 'School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010' *Center for International Food and Agricultural Policy*, p. 3.

169 *Ibid.*

170 *Ibid.*

171 Clarke, M. (2013) Queen's University. Interview with Fiennes, C., 07.13.

172 Paul Hamlyn Foundation *Assessing Impact* [Online] http://issuu.com/paulhamlynfoundation/docs/ assessing_impact?e=1112503/2404563 [Accessed 16.03.14].

173 Washington Post. *The solutions to all our problems may be buried in PDFs that nobody reads* [Online] www.washingtonpost.com/blogs/wonkblog/wp/2014/05/08/the-solutions-to-all-our-problems-may-be- buried-in-pdfs-that-nobody-reads/

174 *Evaluation in government* (2013) London: National Audit Office.

175 Miller, C. (2014) conversation with Fiennes, C., 04.14.

176 Powers, S. (2013) MIT. Interview with Fiennes, C., 31.05.13.

177 Glewwe, P. (2013) University of Minnesota. Interview with Fiennes, C., 07.07.13.

178 Center for Global Development. *Publications* [Online] http://international.cgdev.org/section/publications [Accessed: 29.11.13].

179 FSI 360, Research, monitoring and evaluation: Education: Global [Online] http://www.fhi360.org/ education-global/research-monitoring-and-evaluation [Accessed 16.02.14].

180 Whitty, C. (2013) UK Department for International Development and London School of Hygiene and Tropical Medicine, Interview with Fiennes, C., 30.07.13.

181 Powers, S. (2013) MIT. Interview with Fiennes, C., 31.05.13.

182 Anonymous (2013) Interview with Giving Evidence.

183 Abdul Latif Jameel Poverty Action Lab. *Evaluations* [Online] http://www.povertyactionlab.org/search/ apachesolr_search?filters=type%3Aevaluation%20sm_cck_field_themes%3A2&view=&viewall=all [Accessed: 26.11.13].

184 Glewwe, P. (2013) University of Minnesota. Interview with Fiennes, C., 07.07.13.

185 Conversation with non-academic researcher who asked to remain anonymous, 03.14.

186 Children's Investment Fund Foundation / Hewlett Foundation (2013) Interview of staff with Fiennes, C., 23.07.13 & 24.07.13.

187 Brookings Institution, *Setting a Research Agenda for Learning: Research Task Force and Consultative Process* [Online] http://www.brookings.edu/about/centers/universal-education/global-compact-on- learning/research-task-force

188 *Ibid.*

189   Nelson, J., (2014), *Evidence in education – a lever for professional autonomy* – London: Alliance4 Useful
         Evidence [Online] http://www.alliance4usefulevidence.org/evidence-in-education/ [Accessed: 19.02.14].

190   Children's Investment Fund Foundation / Hewlett Foundation (2013) Interview of staff with Fiennes, C.,
         23.07.13 & 24.07.13.

191   The Campbell Collaboration. *About Us* [Online] http://www.campbellcollaboration.org/about_us/index.
         php [Accessed: 13.09.13].

192   *Ibid.*

193   The Campbell Collaboration. *Advanced Search* [Online] http://www.campbellcollaboration.org/lib/index.
         php?go=browseandadvanced_search=1andsearch_data%5B0%5D%5Bquery%5D=andsearch_data%5
         B0%5D%5Bcriteria%5D=titleandcoordinating_groups%5Beducation%5D=educationandpublished_type
         %5Breview%5D=reviewandpublished_from=andpublished_to= [Accessed: 13.09.13].

194   EPPI-Centre. *About the EPPI-Centre* [Online] https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=63
         [Accessed: 13.09.13].

195   International Initiative for Impact Evaluation. *Systematic Reviews* [Online] http://www.3ieimpact.org/en/
         evidence/systematic-reviews/ [Accessed: 10.12.13].

196   International Initiative for Impact Evaluation. *About 3ie* [Online] http://www.3ieimpact.org/en/about/
         [Accessed: 10.12.13].

197   Whitty, C. (2013) UK Department for International Development and London School of Hygiene and
         Tropical Medicine, Interview with Fiennes, C., 30.07.13.

198   White, H. (2013) 3ie. Speaking at London School of Hygiene and Tropical Medicine 3ie Public Lecture,
         15.04.13.

199   Anonymous (2013) Interview with Giving Evidence.

200   Grounded Theory Institute. *What is Grounded Theory?* [Online] http://www.groundedtheory.com/what-is-
         gt.aspx [Accessed: 12.13].

201   Trochim, W.M. *The Research Methods Knowledge Base*, 2nd Edition [Online] http://www.
         socialresearchmethods.net/kb/ (version current as of 01.11.00). [Accessed: 12.13].

202   Goldacre, B. (2013) *Building Evidence into Education.*

203   *Ibid.*

204   Fuller, B., Clarke, P. (1994) 'Raising School Effects While Ignoring Culture? Local Conditions and the
         Influence of Classroom Tools, Rules, and Pedagogy', *Review of Educational Research*, 64(1), pp. 119-
         157.

205   Fryer, R., Harvard University. The Education Innovation Laboratory [Online] http://www.edlabs.harvard.
         edu/research-publications

206   *Ibid.*

207   *Ibid.*

208   *Ibid.*

209   Barnett, I. et al. (2013) 'The Young Lives Study', *International Journal of Epidemiology* 5.5
         Publisher Oxford University Press.

210   Rolleston, C (2013) Speaking at a UKCDS meeting. London. 08.13.

211   *Ibid.*

212   Khan Academy. *About* [Online] https://www.khanacademy.org/about [Accessed: 01.14].

213   *The Economist*. (2011) *Flipping the Classroom* [Online] http://www.economist.com/node/21529062

214   Coomarasamy, A., Lissauer, D. (2013) Doctors at Birmingham Women's Hospital. Interview with Fiennes,
         C., 19.08.13.

215   UK Clinical Research Collaboration (2012) *UK Health Research Analysis* [Online] http://www.ukcrc.org/
         researchcoordination/healthresearchanalysis/ukanalysis/ [Accessed: 20.09.13].

216   *Ibid.*

217   Research America (2010) 'Health Research and Health Care Spending in the U.S.' [Online] http://www.
         researchamerica.org/uploads/healthdollar10.pdf [Accessed: 20.09.13].

218   TDR/World Health Organization. *About Us* [Online] http://www.who.int/tdr/about/en/ [Accessed:
         20.09.13].

219   World Health Organization (2013) *Voluntary contributions by fund and by contributor for the year ended 31
         December 2012* [Online] http://www.who.int/about/resources_planning/A66_29add1-en.pdf

220   Foundation Arc Pour La Recherche Sur Le Cancer. *ARC Foundation Leopold Griffuel Award* (English
         Version) [Online] http://www.recherche-cancer.net/espace-recherche/palg/english.html [Accessed: 01.14].

221   Brezis, M. (2008) Big Pharma and Healthcare: Unsolvable Conflict of Interests between Private Enterprise
         and Public Health. *The Israel Journal of Psychiatry and Related Sciences*, 45(2), pp. 83-9.

222   UK Clinical Research Collaboration (2012) UK Health Research Analysis [Online] http://www.ukcrc.org/
         researchcoordination/healthresearchanalysis/ukanalysis/ [Accessed: 20.09.13].

223   DFID, ESRC and UKCDS workshop, 21.06.13.

224   London School of Economics (2013) Advocates of RCTs in Education Should Look More Closely at
         the Differences between Medical Research and Education Research [Online] http://blogs.lse.ac.uk/
         politicsandpolicy/archives/32605 [Accessed: 03.12.13].

225   Dercon, S., UKCDS Town Hall Meeting on Education Systems, 23.01.14.

226   Wellcome Trust. *History of Henry Wellcome* [Online] http://www.wellcome.ac.uk/about-us/history/index.
         htm [Accessed: 08.01.14].

227   World Bank (2012) Research at Work: Assessing the Influence of World Bank Research [Online] http://
      www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/2012/04/18/000386194_201204
      18003646/Rendered/PDF/680860BR0SecM20Official0Use0Only090.pdf

228   Banerjee, A. et al. (2006) *An Evaluation of World Bank Research, 1998-2005* [Online] http://siteresources.
      worldbank.org/DEC/Resources/84797-1109362238001/726454-1164121166494/RESEARCH-
      EVALUATION-2006-Main-Report.pdf [Accessed: 10.11.13].

229   London School of Economics (2013) Advocates of RCTs in Education Should Look More Closely at
      the Differences between Medical Research and Education Research [Online] http://blogs.lse.ac.uk/
      politicsandpolicy/archives/32605 [Accessed: 03.12.13].

230   United Nations Educational, Scientific and Cultural Organization. *Our Donors* [Online] http://www.uis.
      unesco.org/AboutUIS/Pages/uis-donors.aspx [Accessed: 22.08.13].

231   ASER Centre. *Funding* [Online] http://www.asercentre.org/p/51.html [Accessed: 27.09.13].

232   OECD. *About PISA* [Online] http://www.oecd.org/pisa/aboutpisa/ [Accessed: 27.09.13].

233   Heady, L. (2013) CIFF. Interview with Fiennes, C., 26.06.13.

234   Department for International Development. *Systematic Reviews* [Online] http://r4d.dfid.gov.uk/
      SystematicReviews.aspx#Education [Accessed: 21.08.13].

235   Economic and Social Research Council. *Funding opportunities* [Online] http://www.esrc.ac.uk/funding-
      and-guidance/funding-opportunities/ [Accessed: 22.10.13].

236   The Global Fund. *Funding and Spending* [Online] http://www.theglobalfund.org/en/about/
      fundingspending/ [Accessed: 22.10.13].

237   Global Partnership for Education. *Pledging Conference – November 2011* [Online] http://www.
      globalpartnership.org/finance-and-funding/replenishment/2011-pledging-conference/ [Accessed: 22.10.13].

238   Global Partnership for Education. *Replenishing our Financial Resources* [Online] http://www.
      globalpartnership.org/finance-and-funding/replenishment/ [Accessed: 22.10.13].

239   Martinez, S., Naudeau, S., Pereira, V. (2012) 'The Promise of Preschool in Africa: A Randomised Impact
      Evaluation of Early Childhood Development in Rural Mozambique'. *International Initiative for Impact
      Evaluation* [Online] http://www.3ieimpact.org/media/filer/2013/04/11/3ie_mozambique_ie001.pdf

240   Children's Investment Fund Foundation / Hewlett Foundation (2013) Interview of staff with Fiennes, C.,
      23.07.13 & 24.07.13.

241   Abdul Latif Jameel Poverty Action Lab. *Partners* [Online] http://www.povertyactionlab.org/search/
      apachesolr_search?filters=type:partner [Accessed: 21.08.13].

242   Department of Health. *Improving Research* [Online] http://dementiachallenge.dh.gov.uk/category/areas-
      for-action/research/ [Accessed: 22.10.13].

243   World Health Organization. *Welcome to the Global Health Observatory* [Online] http://apps.who.int/gho/
      data/view.main [Accessed: 22.10.13].

244   European Commission. *Legal framework governing medicinal products for human use in the EU* [Online]
      http://ec.europa.eu/health/human-use/legal-framework/index_en.htm [Accessed: 21.01.14].

245   US Food and Drug Administration [Online] http://www.fda.gov/default.htm

246   Australian Government Department of Health. *How therapeutic goods are regulated in Australia* [Online]
      http://www.tga.gov.au/industry/basics-regulation.htm#.Ut5zrhDFLZ4 [Accessed: 21.01.14].

247   Friedrich, M.J. (2013) 'The Cochrane Collaboration Turns 20', *Journal of the American Medical
      Association*, 309(18), pp 1881.

248   Clarke, M. (2013) Queen's University. Interview with Fiennes, C., 07.13.

249   Friedrich, M.J. (2013) 'The Cochrane Collaboration Turns 20', *Journal of the American Medical
      Association*, 309(18), pp 1881.

250   Clarke, M. (2013) Queen's University. Email conversation with Fiennes, C., 17.12.13.

251   UK Clinical Research Collaboration (2012) *UK Health Research Analysis 2009/10*.

252   Chalmers, I. et al. (2008) 'The James Lind Library: explaining and illustrating the evolution of fair tests of
      medical treatments', *The journal of the Royal College of Physicians of Edinburgh* 38, pp. 259–64 [Online]
      http://www.jameslindlibrary.org/pdf/jll-article.pdf [Accessed: 04.03.14].

253   Velasco, A. et al. (2008) '**H**ealth Technology Assessment and Health Policy-Making in Europe: Current
      status, challenges and potential', *European Observatory on Health Systems and Policies,* WHO,
      Copenhagen [Online] http://www.euro.who.int/__data/assets/pdf_file/0003/90426/E91922.pdf [Accessed
      10.03.14].

254   Kohler, M. (2011) 'Regulatory pathways in the European Union', *MAbs* 3(3), pp. 241–242 [Online] https://
      www.landesbioscience.com/journals/mabs/article/15474/

255   Kramer, D.B. et al. (2012) 'Regulation of Medical Devices in the United States and European Union'
      The New England Journal of Medicine 366 pp. 848-855 [Online] http://www.nejm.org/doi/full/10.1056/
      NEJMhle1113918

256   Donawa, M. and Gray, R. (2012) 'The breast implant scandal and European medical device regulation'
      European Industrial Pharmacy 12 pp. 4-6.

257   Institute of Medicine (2010*)* 'Transforming Clinical Research in the United States: Challenges and
      Opportunities: Workshop Summary', *The National Academies Press ,*Washington, DC [Online] http://www.
      ncbi.nlm.nih.gov/books/NBK50886/

258   Kramer, D.B. et al. (2012) 'Regulation of Medical Devices in the United States and European Union'
      The New England Journal of Medicine 366 pp.848-855 [Online] http://www.nejm.org/doi/full/10.1056/
      NEJMhle1113918

259   Velasco, A. et al. (2008) **'**Health Technology Assessment and Health Policy-Making in Europe: Cur-
       rent status, challenges and potential', *European Observatory on Health Systems and Policies – WHO,
       Copenhagen* [Online] http://www.euro.who.int/__data/assets/pdf_file/0003/90426/E91922.pdf [Accessed
       10.03.14].

260   *Ibid.*

261   *Ibid.*

262   Cowell, R. (1993) 'The Design and Development of Secondary Education Curriculum' The World Bank
       [Online] http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/1993/09/01/00000
       9265_3980624090527/Rendered/PDF/multi_page.pdf

263   The evolving regulatory context

264   Cowell, R. (1993) 'The Design and Development of Secondary Education Curriculum' The World Bank
       [Online] http://www-wds.worldbank.org/external/default/WDSContentServer/WDSP/IB/1993/09/01/00000
       9265_3980624090527/Rendered/PDF/multi_page.pdf

265   *Ibid.*

266   *Ibid.*

267   Fielden, J. and LaRocque, N. (2008) 'The Evolving Regulatory Context for Private
       Education in Emerging Economies: Discussion Paper', *International Finance Group, World
       Bank Group*, Washington, DC [Online] http://siteresources.worldbank.org/EDUCATION/
       Resources/278200-1099079877269/547664-1099079934475/547667-1135281523948/EWPS14_Private_
       Education.pdf

268   *Ibid.*

269   *Ibid.*

270   *Ibid.*

271   Kaushik, A. (2013) Ab Initio Consulting, Panel discussion with Shekhar Shah at 2013 School Choice
       National Conference, Delhi 20.12.13 [Online] http://prayatna.typepad.com/education/2013/12/3-things-
       india-should-do-to-improve-learning-outcomes-by-2025.html

272   Muralidharan, K. (2013), University of California at San Diego, , Panel discussion with Shekhar Shah
       at 2013 School Choice National Conference, Delhi 20.12.13 [Online] http://prayatna.typepad.com/
       education/2013/12/3-things-india-should-do-to-improve-learning-outcomes-by-2025.html

273   Tooley, J. and Dixon, P. (2005) 'An inspector calls: the regulation of 'budget' private schools in
       Hyderabad, Andhra Pradesh, India', *International Journal of Educational Development* 25, pp. 269-
       285 (abstract) [Online] http://www.researchgate.net/publication/222646057_An_inspector_calls_the_
       regulation_of_budget_private_schools_in_Hyderabad_Andhra_Pradesh_India

274   Children's Investment Fund Foundation / Hewlett Foundation (2013) Interview of staff with Fiennes, C.,
       23.07.13 & 24.07.13.

275   *Ibid.*

276   Ground Up (2013) *DEVTA Trial – will it help or harm our treatment programs?* [Online] http://www.
       groundup.org.za/content/devta-trial-will-it-help-or-harm-our-treatment-programs [Accessed: 10.13].

277   Clarke, M. (2013) Queen's University. Interview with Fiennes, C., 07.13.

278   Radiolab. *Cut and Run*. [Online] http://www.radiolab.org/story/runners/ [Accessed: 11.02.14].

279   Novy Marx, M. (2013) Interview with Fiennes, C., 18.07.13.

280   *Ibid.*

281   Laughan, P. (2013) Firelight Foundation. Interview with Fiennes, C., 03.07.13.

282   Whitty, C. (2013) UK Department for International Development and London School of Hygiene and
       Tropical Medicine, Interview with Fiennes, C., 30.07.13.

283   Powers, S. (2013) MIT. Interview with Fiennes, C., 31.05.13.

284   Chavan, M., Banerji, R. (2013) Pratham. Interview with Fiennes, C., 07.13.

285   Heady, L. (2013) Children's Investment Fund Foundation. Interview with Fiennes, C., 26.06.13.

286   Goldacre, B. (2013) *Building Evidence into Education.*

287   Morris., Devlin., Parkin (2007) Principles of Health Economics (including Notions of Scarcity, Supply
       and Demand, Marginal Analysis, Distinctions between Need and Demand, Opportunity Cost, Margins,
       Efficiency and Equity) [Online] http://www.healthknowledge.org.uk/public-health-textbook/medical-
       sociology-policy-economics/4d-health-economics/principles-he

288   Matchar, D.B. et al. (1997) The Stroke Prevention Policy Model: linking evidence and clinical decisions.
       (Review) *Annals of Internal Medicine*, 127(8 (Pt2), 704-711.

289   *Ibid.*

290   Klarman, H., Francis, J., Rosenthal, G. 'Cost Effectiveness Analysis Applied to the Treatment of Chronic
       Renal Disease'. *Medical Care.* 6(1), pp. 48-54.

291   National Institute for Health and Clinical Excellence (2008) 'Guide to the methods of technology appraisal'
       [Online] www.nice.org.uk/media/B52/A7/TAMethodsGuideUpdatedJune2008.pdf [Accessed: 10.03.10].

292   National Institute for Health and Care Excellence. *Glossary* [Online] http://www.nice.org.uk/website/
       glossary/glossary.jsp?alpha=Q

293   Wagstaff, A., Culyer, A.J. (2011) 'Four Decades of Health Economics through a Bibliometric Lens', *The
       World Bank* [Online] http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-5829

294   Abdul Latif Jameel Poverty Action Lab. *Student Participation* [Online] *http://www.povertyactionlab.org/
       policy-lessons/education/student-participation*

295   Abdul Latif Jameel Poverty Action Lab. Teacher Attendance and Incentives [Online] http://www.povertyactionlab.org/policy-lessons/education/teacher-attendance-incentives

296   Powers, S. (2013) MIT. Interview with Fiennes, C., 31.05.13.

297   Wagstaff, A., Culyer, A.J. (2011) 'Four Decades of Health Economics through a Bibliometric Lens', *The World Bank* [Online] http://elibrary.worldbank.org/doi/book/10.1596/1813-9450-5829

298   Kremer, M., Brannen, C., Glennerster, R. (2013) 'The Challenge of Education and Learning in the Developing World', *Science*, 340(6130), pp. 297-300 [Online] http://www.sciencemag.org/content/340/6130/297.full?ijkey=tq1ax.4Tmcjacandkeytype=refandsiteid=sci

299   *Ibid.*

300   British Medical Journal. (2012) 'High reprint orders in medical journals and pharmaceutical industry funding: case-control study' *British Medical Journal.*

301   Global Partnership for Education. Sharing Innovative Solutions for Education [Online] http://www.globalpartnership.org/our-work/Sharing-Innovative-Solutions-for-Education/ [Accessed: 03.12.13].

302   *Ibid.*

303   Children's Investment Fund Foundation / Hewlett Foundation (2013) Interview of staff with Fiennes, C., 23.07.13 & 24.07.13.

304   Glewwe, P. (2013) University of Minnesota. Interview with Fiennes, C., 07.07.13.

305   Journal of Medical Genetics. *Journals* [Online] http://jmg.bmj.com/ [Accessed 12.02.14].

306   British Medical Journal. *Journals* [Online] http://journals.bmj.com/ [Accessed: 25.10.13].

307   Simpson, P. (2013) Senior Editor, Public Library of Science Medicine. Interview with Fiennes, C., 26.07.13.

308   The Lancet, *Rate Card* [Online] download.thelancet.com/flatcontentassets/RateCard2012.pdf [Accessed 16.04.14].

309   British Medical Journal (2012) 'High reprint orders in medical journals and pharmaceutical industry funding: case-control study' *British Medical Journal.*

310   *PLOS*. 'Global Participation Initiative [Online] http://www.plos.org/newsroom/viewpoints/global-participation-initiative/ [Accessed 16.02.14].

311   *PLOS* Blogs. '*PLOS One* Launched Reproducibility Initiative' [Online] http://blogs.plos.org/everyone/2012/08/14/plos-one-launches-reproducibility-initiative/ [Accessed 16.02.14].

312   Bad Science. 'Matthias Rath drops his million pound legal case against me and the Guardian' [Online] http://www.badscience.net/2008/09/matthias-rath-pulls-out-forced-to-pay-the-guardians-costs-i-think-this-means-i-win/ [Accessed: 18.12.13].

313   The Guardian. 'Why we sued Simon Singh: the British Chiropractic Association speaks' [Online] http://www.theguardian.com/science/blog/2012/feb/22/simon-singh-british-chiropractic-association [Accessed: 18.12.13].

314   The Guardian. 'Setback for US company suing cardiologist Peter Wilmshurst for libel' [Online] *http://www.theguardian.com/science/2010/dec/01/company-suing-peter-wilmshurst-libel?guni=Article:in%20body%20link [Accessed: 12.12].*

315   Budapest Open Access Initiative. 'Ten years on from the Budapest Open Access Initiative: setting the default to open'. *WebCite* [Online] http://www.webcitation.org/6AdLcI1TC [Accessed: 29.11.13].

316   Simpson, P. (2013) Senior Editor, Public Library of Science Medicine. Interview with Fiennes, C., 26.07.13.

317   Jeffery, K.G. *Open Access: An Introduction* [Online] http://www.ercim.eu/publication/Ercim_News/enw64/jeffery.html [Accessed: 29.11.13].

318   Simpson, P. (2013) Senior Editor, Public Library of Science Medicine. Interview with Fiennes, C., 26.07.13.

319   *Ibid.*

320   Public Library of Science. *How open is it?* [online] http://www.plos.org/open-access/howopenisit/ [Accessed: 29.11.13].

321   Children's Investment Fund Foundation / Hewlett Foundation (2013). Interview of staff with Fiennes, C., 23.07.13 & 24.07.13.

322   Kremer, M., Miguel, E. (2004) *Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities*, Econometrica 72 (1): 159-217.

323   Glewwe, P. et al. (2011) 'School resources and educational outcomes in developing countries: A review of the literature from 1990 to 2010' *Center for International Food and Agricultural Policy*, p. 17.

324   Children's Investment Fund Foundation / Hewlett Foundation (2013). Interview with Fiennes, C., 23.07.13 & 24.07.13.

325   U.S. National Library of Medicine [Online] http://www.nlm.nih.gov [Accessed: 22.10.13].

326   U.S. National Library of Medicine. *Fact Sheet: MEDLINE* [Online] http://www.nlm.nih.gov/pubs/factsheets/medline.html

327   U.S. National Library of Medicine. *PMC FAQs* [Online].

328   Center for Education Innovations. *Programs Database* [Online] http://www.educationinnovations.org/programs [Accessed: 21.08.13].

329   National Center for Education Evaluation and Regional Assistance. *Education Resources Information Center* [Online] http://ies.ed.gov/ncee/projects/eric.asp [Accessed: 21.08.13].

330   United Nations Educational, Scientific and Cultural Organization. *Library Resources* [Online] http://www.iiep.unesco.org/information-services/documentation-centre/library-resources.html [Accessed: 21.08.13].

331   Abdul Latif Jameel Poverty Action Lab. *Evaluations* [Online] http://www.povertyactionlab.org/search/apachesolr_search?filters=type:evaluation [Accessed: 21.08.13].

332   Global Partnership for Education. Monitoring Education Results [Online] http://www.globalpartnership. org/our-work/our-work-monitoring-education-results/

333   United Nations Educational, Scientific and Cultural Organization. *Data to make a difference* [Online] http://www.uis.unesco.org/Pages/default.aspx World Bank. *World Bank open data* [Online] http://data. worldbank.org/ [Accessed: 03.11.13].

334   Centers for Disease Control and Prevention. *Conferences & Events* [Online] http://www.cdc.gov/about/ newsEvents/events.htm [Accessed: 22.10.13].

335   Glewwe, P. (2013) University of Minnesota. Interview with Fiennes, C., 07.07.13.

336   Innovations for Poverty Action (2012) Evidence-Based Education: Policy-Making and Reform in Africa conference, Alisa Hotel, Accra, Ghana, 14-15.05.12 [Online] http://poverty-action.org/Accra_Education_ May2012 [Accessed: 11.07.13].

337   Brookings Institution. *Upcoming event: New Ideas to Scale Up and Finance Global Education* [Online] http://www.brookings.edu/events/2014/02/24-scale-up-finance-education [Accessed 16.02.14].

338   Glewwe, P. (2013) University of Minnesota. Interview with Fiennes, C., 07.07.13.

339   Global Partnership for Education. *Sharing Innovative Solutions for Education* [Online] http://www. globalpartnership.org/our-work/Sharing-Innovative-Solutions-for-Education/ [Accessed: 25.10.13].

340   The Royal Australian College of General Practitioners [Online] http://www.racgp.org.au/ [Accessed: 22.10.13].

341   Royal College of General Practitioners. *A Guide To Membership* [Online] http://www.rcgp.org.uk/~/media/ Files/Membership/RCGP-Membership-Guide-2013.ashx [Accessed: 22.10.13].

342   Laughan, P. (2013) Firelight Foundation. Interview with Fiennes, C., 03.07.13.

343   Davis, D.A. (1995) 'Changing Physician Performance: A Systematic Review of the Effect of Continuing Medical Education Strategies', *Journal of the American Medical Association* 274(9), pp. 700-5.

344   MedicineNet.com. *Definition of Grand rounds* [Online] http://www.medterms.com/script/main/art. asp?articlekey=40370 [Accessed: 22.10.13].

345   Goldacre, B. *Bad Pharma.* Fourth Estate.

346   Rhee, J. (2008) 'The Influence of the Pharmaceutical Industry on Healthcare Practitioners' Prescribing Habits'. *The Internet Journal of Academic Physician Assistants*. 7(1).

347   The World Bank. *Education* [Online] http://data.worldbank.org/topic/education. [Accessed: 07.07.2013].

348   Children's Investment Fund Foundation / Hewlett Foundation (2013) Interview with Fiennes, C., 23.07.13 & 24.07.13.

349   *Ibid.*

350   *Ibid.*

351   *Ibid.*

352   *Ibid.*

353   Nobel Prize. *15 Questions and Answers about the Nobel Prize in Physiology or Medicine* [Online] http:// www.nobelprize.org/nomination/medicine/questions-goran-hansson-2013.html [Accessed: 08.01.14].

354   The Royal Society of Medicine. *RSM Awards* [Online] http://www.rsm.ac.uk/academ/awards/ [Accessed: 10.12.13].

355   World Innovation Summit for Education [Online] http://www.wise-qatar.org/content/wise-initiative#tabslink-1 [Accessed: 10.12.13].

356   World Innovation Summit for Education. *WISE 2012: The WISE Prize for Education.* [Online] http://www. wise-qatar.org/content/wise-prize-education [Accessed: 10.12.13].

357   USAID. *Grand Challenges for Development* [Online] http://www.usaid.gov/grandchallenges [Accessed: 10.13].

358   STIR Education. *STIR in India* [Online] http://www.stireducation.org/ambition/india/ [Accessed: 10.12.13].

359   HP-101. *Content Discussion: Social Learning Theory* [Online] http://www.thcu.ca/hp101/eng/mod4/ module4c7.cfm [Accessed: 10.12.13].

360   World Health Organization (2007) *A Guide for Fostering Change to Scale Up Effective Health Services*, p. 2.

361   Glewwe, P. (2013) University of Minnesota. Interview with Fiennes, C., 07.07.13.

362   Children's Investment Fund Foundation / Hewlett Foundation (2013) Interview with Fiennes, C., 23.07.13 & 24.07.13.

363   *Ibid.*

364   Chavan, M., Banerji, R. (2013) Pratham. Interview with Fiennes, C., 07.13.

365   British Broadcasting Corporation (2006) *Does the MMR Jab Cause Autism?* [Online] http://www.bbc. co.uk/sn/tvradio/programmes/horizon/mmr_prog_summary.shtml [Accessed: 10.13].

366   Goldacre, B. *Bad Pharma*. Fourth Estate. Afterword: what happened next?

367   *The Economist* (2011) 'Flipping the classroom: Hopes that the internet can improve teaching may at last be bearing fruit'. *The Economist.*

368   Powers, S. (2013) MIT. Interview with Fiennes, C., 31.05.13.

369   Stokes, K., Barker, R., Pigott, R. (2014) Which doctors take up promising ideas? New insights from open data, London, Nesta, p. 16 http://www.nesta.org.uk/sites/default/files/which_doctors_take_up_promising. pdf [Accessed: 28.02.14].

370   *Ibid.*, p. 29.

371   TES. *About us* [Online] http://www.tes.co.uk/article.aspx?storyCode=6000208&navCode=285 [Accessed: 04.03.14].

372   Simpson, P. (2013) Editor, Public Library of Science Medicine. Interview with Fiennes, C., 26.07.13.

373 Friedrich, M.J. (2013) 'The Cochrane Collaboration Turns 20', *Journal of the American Medical Association*, 309(18), p 1881.

374 *Ibid.*

375 Simpson, P. (2013) Editor, Public Library of Science Medicine. Interview with Fiennes, C., 26.07.13.

376 eLife Sciences. *About* [Online] http://www.elifesciences.org/about/ [Accessed: 10.13].

377 World Health Organization. *Data and Statistics* [Online] http://www.who.int/research/en/ [Accessed 10.13].

378 Simpson, P. (2013) Editor, Public Library of Science Medicine. Interview with Fiennes, C., 26.07.13.

379 *Ibid.*

380 American Economic Association [Online] http://www.aeaweb.org/ [Accessed: 07.02.14].

381 Simpson, P. (2013) Editor, Public Library of Science Medicine. Interview with Fiennes, C., 26.07.13.

382 Winthrop, R. (2014) Brookings Institution. Conversation with Fiennes, C., 03.14.

383 Miller, C. (2014) FB Heron Foundation. Conversation with Fiennes, C., 03.14.

384 Shepherd, J.P. (2007) 'The production and management of evidence for public service reform'. *Evidence and Policy*. 3(2): pp. 231–251.

385 McMaster University. *Health Systems Evidence* [Online] http://www.healthsystemsevidence.org/top-5-one-pagers-all.aspx [Accessed: 11.09.13].

386 Thaler, R.H., 'Watching Behavior Before Writing the Rules', *The New York Times* [Online] http://www.nytimes.com/2012/07/08/business/behavioral-science-can-help-guide-policy-economic-view.html?_r=2and [Accessed: 14.11.13].

387 Banerjee, A.V., Duflo, E. (2011) *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*. PublicAffairs,U.S.

388 Goldacre B. *Building Evidence into Education*, 03.13.

389 UNICEF. *Education and Literacy*. [Online] http://www.unicef.org/specialsession/about/sgreport-pdf/sgrep_adapt_part2b_eng.pdf [Accessed: 11.02.14].

390 Conversations which are anonymised as requested.

391 Institute of Education Sciences. What works Clearinghouse [Online] http://ies.ed.gov/ncee/wwc/ [Accessed: 10.13].

392 Education Endowment Foundation [Online] http://educationendowmentfoundation.org.uk/ [Accessed: 10.13].

393 Dr Foster Health. *The Hospital Guide 2013* [Online] http://www.drfosterhealth.co.uk/consultant-guide/ [Accessed: 10.11.13].

394 British Broadcasting Corporation (2007) *News: Heart op report criticises Trust* [Online] http://news.bbc.co.uk/1/hi/england/oxfordshire/6499457.stm [Accessed: 01.05.13].

395 Care Quality Commission. *About Us* [Online] http://www.cqc.org.uk/about-us [Accessed: 10.11.13].

396 Coomarasamy, A., Lissauer, D. (2013) Professors from Birmingham Women's Hospital. Interview with Fiennes, C., 19.08.13.

397 Damberg, C.L. et al. (2011) 'An Evaluation of using Performance Measures in Healthcare' [Online] http://www.rand.org/content/dam/rand/pubs/technical_reports/2011/RAND_TR1148.pdf *Rand Health*. [Accessed: 02.05.13].

398 Children's Investment Fund Foundation / Hewlett Foundation (2013) Interview of staff with Fiennes, C., 23.07.13 & 24.07.13.

399 Banerjee, A.J., Duflo, E. (2011) *Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty*, PublicAffairs, U.S.

400 Kahneman, D. (2012) *Thinking, Fast and Slow*. London: Penguin. p. 63.

401 *Ibid*., p. 45.

402 Townsend, J. (2013) STIR. Interview with Fiennes, C., 15.06.13.

403 Sampath, B. (2013) EurekaChild. Interview with Fiennes, C., 19.06.13.

404 Novy Marx, M. (2013). Interview with Fiennes, C., 18.07.13.

405 Kahneman, D. (2012) *Thinking, Fast and Slow*. London: Penguin. p. 30.

406 Sampath, B. (2013) EurekaChild. Interview with Fiennes, C., 19.06.13.

407 Kahneman, D. (2012) *Thinking, Fast and Slow*. London: Penguin. p. 45.

408 *Ibid*., p. 367, p. 249, p. 263, p. 228, p. 139, p. 45, p. 130.

409 Gerber, A.S., Patashnik, E.M., Doherty, D (2013) 'Doctor Knows Best: Physician Endorsements, Public Opinion, and the Politics of Comparative Effectiveness Research'. *Journal of Health, Politics, Policy and Law.*

410 Haslam, D. (2014) Chair of NICE, speaking at NESTA event about experiences of NICE, apparently citing Bruce Keogh, Medical Director of the NHS in England.

411 Gawande, A. (2013) 'Slow Ideas: Some innovations spread fast. How do you speed the ones that don't?', *New Yorker Magazine* [Online] http://www.newyorker.com/reporting/2013/07/29/130729fa_fact_gawande?currentPage=all

412 *Ibid.*

413 Novy Marx, M. (2013) Interview with Fiennes, C., 18.07.13.

414 Gawande, A. (2013) 'Slow Ideas: Some innovations spread fast. How do you speed the ones that don't?', *New Yorker Magazine* [Online] http://www.newyorker.com/reporting/2013/07/29/130729fa_fact_gawande?currentPage=all

415 Patashnik, E (2014) University of Virginia. Interview with Fiennes, C., 05.13

416 Townsend, J. (2013) STIR. Interview with Fiennes, C., 15.06.13.

417 Banerjee, A. et al. (2006) *An Evaluation of World Bank Research, 1998-2005* [Online] http://siteresources.worldbank.org/DEC/Resources/84797-1109362238001/726454-1164121166494/RESEARCH-EVALUATION-2006-Main-Report.pdf [Accessed: 10.11.13].

418 *Ibid.*

419 Children's Investment Fund Foundation / Hewlett Foundation (2013) Interview of staff with Fiennes, C., 23.07.13 & 24.07.13.

420 Van Dijk, N., Hooft, L., Wieringa-de Waard, M. (2010) 'What Are the Barriers to Residents' Practicing Evidence-Based Medicine?' Epub ahead of print.

421 Haynes, R.B., Hayward, R.S.A., Lomas, J. (1995) 'Bridges between healthcare research evidence and clinical practice'. *Journal of the American Medical Informatics Association* 2(6), pp. 342-350.

422 Glewwe, P. (2013) University of Minnesota. Interview with Fiennes, C., 07.07.13.

423 Powers, S. (2013) MIT. Interview with Fiennes, C., 31.05.13.

424 DFID, ESRC and UKCDS workshop, 21 June 2013.

425 Lloyd, M. (2008) 'Striking Mexico teachers see jobs as things to inherit, sell', *Houston Chronicle* [Online] http://www.chron.com/life/mom-houston/article/Striking-Mexico-teachers-see-jobs-as-things-to-1642091.php
Press TV (2013) 'Mexican teachers strike over assessments' *Press TV* [Online] http://www.presstv.com/detail/2013/08/20/319613/mexican-teachers-strike-over-assessments/

426 Mulkeen, A. (2010) 'Teachers in Anglophone Africa: Issues in Teacher Supply, Training, and Management', *World Bank, Washington, DC*. p 4.

427 *Ibid.*

428 *Ibid.*

429 Children's Investment Fund Foundation / Hewlett Foundation (2013) Interview of staff with Fiennes, C., 23.07.13 & 24.07.13.

430 Coomarasamy, A., Lissauer, D. (2013) Professors from Birmingham Women's Hospital. Interview with Fiennes, C., 19.08.13.

431 US Department of Health & Human Services. *Guideline Index* [Online] http://www.guideline.gov/browse/index.aspx?alpha=A

432 World Health Organization. *Documents listed alphabetically* [Online] http://www.who.int/publications/guidelines/atoz/en/index.html

433 Coomarasamy, A., Lissauer, D. (2013) Professors from Birmingham Women's Hospital. Interview with Fiennes, C., 19.08.13.

434 National Institute for Health and Care Excellence [Online] www.evidence.nhs.uk

435 Fontanarosa, P.B., DeAngelis, C.D. (2002) 'Basic Science and Translational Research in JAMA'. *Journal of the American Medical Association*. 287(13), pp. 1728.

436 *Ibid.*

437 Cobelens, F. et al. (2012) 'Which New Diagnostics for Tuberculosis, and When?' *Journal of Infectious Diseases.* p. 205.

438 National Institute for Health and Care Excellence. *NICE Guidance* [Online] http://guidance.nice.org.uk/ [Accessed: 25.10.13].

439 National Institute for Clinical Excellence (2004) 'Clinical practice guideline for the assessment and prevention of falls in older people' *Royal College of Nursing* [Online] http://www.nice.org.uk/nicemedia/pdf/CG021fullguideline.pdf [Accessed: 11.01.14].

440 *Ibid.*

441 Clarke, M. (2013) Queen's University. Interview with Fiennes, C., 07.13.

442 Adapte. *The Adapte Collaboration* [Online] http://www.adapte.org/www/rubrique/the-adapte-collaboration.php [Accessed: 11.12.13].

443 National Institute for Clinical Excellence (2004) 'Clinical practice guideline for the assessment and prevention of falls in older people' *Royal College of Nursing* [Online] http://www.nice.org.uk/nicemedia/pdf/CG021fullguideline.pdf [Accessed: 11.01.14].

444 NICE Pathways. *Self-limiting respiratory tract infections – antibiotic prescribing overview* [Online] http://pathways.nice.org.uk/pathways/self-limiting-respiratory-tract-infections---antibiotic-prescribing [Accessed: 11.12.13].

445 Gerber, A.S. et al. (2010) 'A national survey reveals public scepticism about research-based treatment guidelines', *Health Affairs*, 29(10), pp. 1882-1884.

446 *Ibid.*

447 Goldacre, B. *Building Evidence into Education*. 03.13.

448 World Health Organization. *Patient safety checklists* [Online] http://www.who.int/patientsafety/implementation/checklists/en/ [Accessed: 16.12.13].

449 Gawande, A. (2011) *The Checklist Manifesto* [Online] http://gawande.com/the-checklist-manifesto [Accessed: 16.12.13].

450 *Ibid.*

451 *Ibid.*

452 World Health Organization (2009) *Surgical Safety Checklist* [Online] http://whqlibdoc.who.int/publications/2009/9789241598590_eng_Checklist.pdf [Accessed: 16.12.13].

453 World Health Organization. *The Checklist Effect* [Online] http://www.who.int/patientsafety/implementation/checklists/background/en/index.html [Accessed: 16.12.13].

454   *Ibid.*

455   Gawande, A. (2011) *The Checklist Manifesto*. Profile books.

456   *Ibid.*

457   World Health Organization (2009) *Surgical Safety Checklist* [Online] http://whqlibdoc.who.int/
       publications/2009/9789241598590_eng_Checklist.pdf [Accessed: 16.12.13].

458   Gawande, A. (2011) *The Checklist Manifesto.* Profile Books.

459   A Checklist for Checklists (2010) [Online] http://www.projectcheck.org/uploads/1/0/9/0/1090835/
       checklist_for_checklists_final_10.3.pdf [Accessed: 16.12.13].

460   National Institute for Health and Clinical Excellence (2011) *Hypertension* [Online] http://www.nice.org.uk/
       nicemedia/live/13561/56015/56015.pdf [Accessed: 25.10.13].

461   Whitty, C. (2013) UK Department for International Development and London School of Hygiene and
       Tropical Medicine, Interview with Fiennes, C., London, 30.07.13.

462   National Institute for Health and Clinical Excellence. (2011) *Hypertension* [Online] http://www.nice.org.uk/
       nicemedia/live/13561/56015/56015.pdf [Accessed: 25.10.13].

463   Whitty, C. (2013) UK Department for International Development and London School of Hygiene and
       Tropical Medicine, Interview with Fiennes, C., London, 30.07.13.

464   Alltrials.net, *Missing Trial Data – briefing notes* [Online] http://www.alltrials.net/wp-content/
       uploads/2013/01/Missing-trials-briefing-note.pdf [Accessed: 11.02.14].

465   Goldacre, B. (2013) 'Are clinical trial data shared sufficiently today? No.' *British Medical Journal.347*
       [Online] http://www.bmj.com/content/347/bmj.f1880 [Accessed: 11/02/14].

466   Goldacre, B. (2013) *Bad Pharma*. Fourth Estate. Afterword: what happened next?

467   *Ibid.*

468   Handel, A. e*t al.* (2012) 'High reprint orders in medical journals and pharmaceutical industry funding:
       case-control study' *British Medical Journal, p.* 344.

469   Friedrich, M. J. (2013) 'The Cochrane Collaboration Turns 20'. *Journal of the American Medical
       Association.* 309(18).

470   Goldacre, B. (2013) *Bad Pharma*. Fourth Estate.

471   AllTrials. (2013) *European pharmacists group puts AllTrials advert in its journal* [Online] http://www.alltrials.
       net/2013/advert-alltrials-journal/ [Accessed: 11.02.14].

472   GlaxoSmithKline (2013) Press Release: *GSK announces support for AllTrials campaign for clinical data
       transparency* [Online] http://www.gsk.com/media/press-releases/2013/GSK-announces-support-forAll-
       Trials-campaign-for-clinical-data-transparency.html [Accessed: 11.02.14].
       Goldacre B. (2013) *Bad Pharma*. Fourth Estate. Afterword: What happened next?

473   Freedman, D.H. (2010) Lies, Damned Lies, and Medical Science*, The Atlantic* [Online] http://www.
       theatlantic.com/magazine/archive/2010/11/lies-damned-lies-and-medical-science/308269/?single_
       page=true [Accessed: 11.02.14].

474   Ioannidis, J.P.A. (2005) 'Why Most Published Research Findings Are False', *PLOS Medicine 2(8): e124*
       [Online] http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.0020124 [Accessed:
       11.02.14].

475   *Ibid.*

476   Freedman, D.H. (2010) Lies, Damned Lies, and Medical Science*, The Atlantic* [Online] http://www.
       theatlantic.com/magazine/archive/2010/11/lies-damned-lies-and-medical-science/308269/?single_
       page=true [Accessed: 11.02.14].

477   Bohannon, J. (2013) 'Who's Afraid of Peer Review?' *Science*. 342(6154), pp. 60-65.

478   Goldacre, B. (2013) *Bad Pharma*. Fourth Estate.

479   Comet Initiative. *A walk through core outcome sets – useful references for core outcome set developers*
       [Online] http://www.comet-initiative.org/resources/coreresourcepack [Accessed: 11.02.14].

480   *Ibid*.

481   Omeract. *An international initiative to improve outcome measurement in rheumatology.* [Online] http://
       www.omeract.org/pdf/OMERACT%20Tugwell%20Trials.pdf [Accessed: 11.02.14].

482   Sinha, I. et al. (2008) 'A Systematic Review of Studies That Aim to Determine Which Outcomes
       to Measure in Clinical Trials in Children', 10(1371) [Online] http://www.plosmedicine.org/article/
       info%3Adoi%2F10.1371%2Fjournal.pmed.0050096 [Accessed: 11.02.14].

483   Oxman, A.D., Navis, J.N., Fretheim, A. (2007) 'Use of evidence in WHO recommendations', *The Lancet*, p.
       369.

484   *Ibid.*

485   Glewwe, P. et al. (2011) 'School resources and educational outcomes in developing countries: A review of
       the literature from 1990 to 2010' *Center for International Food and Agricultural Policy*, p. 45.