

15/3/2016

SEMINAR

CAN WE LEARN SOMETHING FROM EXPERIMENTAL STUDIES IN THE SOCIAL FIELD?

Introduction

Welcome to everyone.

My name is Gian Paolo Barbetta and I'm the director of the Research Center on cooperative and nonprofit organizations of the UC.

IMPACT is a project recently started by our center, whose aim is to better understand, to promote and to increase the use of what is often called "evidence based policy making". The idea behind this approach is that particularly in a period of restrictions, when economic resources become more and more scarce, all efforts should be made to identify "what works" (and what does not work) when collective entities (public bodies, but also private organizations acting for a public purpose) address some social problems. If we (as a society) were able to distinguish policies (or interventions, or approaches) that work from those that do not deliver as promised, we could possibly improve the living conditions of many individuals and save a lot of money.

But – of course - a clear understanding of "what works" is not always easy to reach, for many different reasons. Over the last years, RCT have emerged in several fields (from agriculture to medicine) as the best available option to get a clear understanding of the impact of a policy intervention. While this approach has been widely accepted in many fields – nowadays nobody would take a medicine whose effect has not been tested in a RCT (well, almost nobody) – this is not always the case in the field of social interventions (from education to welfare). In this last case, the use of RCT is often criticized and the "diversity" of social sciences from medical or hard sciences is often mentioned as a good reason for rejecting the approach,

This is the reason why we decided to devote this seminar to the possible role of "experimental studies" in the social field. We are so lucky as to have with us two real experts in the field, D. and C. Torgerson and I really want to thank them for having accepted to be with us today.

David Torgerson (who, incidentally, is the husband of Carole Torgerson) is the Director of the York Trials Unit at the University of York. Originally a health economist he is now a trial methodologist and has published widely on the design and conduct of randomized controlled trials ... so widely that he has over 200 peer reviewed publications.

Carole Torgerson is Professor of Education at Durham University. Carole's career is – if David allows it to me – even more interesting. In fact, Carole has degrees in English Literature and Modern English Poetry from the Universities of York and Stirling and she taught English in a wide variety of schools in England and

Scotland. She then undertook her doctorate in “design issues in experimental research” at the University of Sheffield before starting her academic career. Her list of publication is equally impressive.

In case any of you is interested in today’s topic, I would recommend this one – slide.

We asked David to focus his speech on the “differences and similarities” that characterize the use of RCT in medicine and in the fields of education and the social welfare. Following, Carole will concentrate her speech on what is going in these latter fields, what are the trends and the lessons learned in the UK and the Anglo-saxon countries where these activities are more frequent, particularly when considering the differences mentioned by David and the criticism coming from the field. They have about 30’ each.

As I mentioned before, RCT are far from accepted in the social field, and quite opposite, they are the target of wide criticism.

RCT are sometimes criticized from a philosophical point of view, when asserting that theoretical statements cannot be tested. Or from a practical point of view, when considering the ethical concerns that makes it difficult their introduction in the social field, or the issue of non-compliance and contamination that makes individuals different in final tests, even though they were equal at the beginning of the experiment. Moreover, further criticism comes when considering that the “decision model” adopted by policy makers rarely relies on classical theories of rationality, so that the production of hard evidence does not necessarily improve the quality of policy making. Furthermore, other evaluation tools – less expensive and less requiring – could also lead to the production of plausible information, for the policy maker.

These are just some of the questions that have been raised when dealing with this particular method of producing evidence on what works.

We asked some friends and colleagues to help us discussing some of these issues and I want to warmly thank them for being here today. They come from different disciplines and back grounds: sociologists, like Emanuela Mora and Gianluca Argentin, or economists, such as Claudio Lucifora and Mario Maggioni. I have to excuse Alessandro Rosina, a statistician and demographer, who has not been able to be here today, but I know that his colleague Laura Deldossi will also help us in the discussion.

Therefore, after the two presentations I will ask these colleagues to enter the discussion and we’ll also be able to open the floor to all our hosts today.

Thank you very much to everybody.

CAN WE LEARN SOMETHING FROM EXPERIMENTAL STUDIES IN THE SOCIAL FIELD?

Randomized controlled trials. From medicine to education and the social field: differences and similarities

Professor David TORGERSON, Director York Trials Unit, University of York, York, UK
david.torgerson@york.ac.uk

Randomized controlled trials in education and the social field: recent trends and lessons

Professor Carole TORGERSON, Chair in Education, Durham University, Durham, UK
carole.torgerson@durham.ac.uk

Designing Randomised Trials

in Health, Education and the Social Sciences

An Introduction



David J. Torgerson and Carole J. Torgerson



Background

- To establish the effectiveness of an intervention we need a 'counterfactual'
 - » What would have happened in the absence of the intervention?
- In some rare cases we can establish this by testing a treatment or intervention on the same individual – 'a cross-over' study
- Most times we need to use a parallel group
- RANDOMISED CONTROLLED TRIAL

Randomisation

- We form two or more groups through random allocation
- Random allocation ensures that all groups are similar in both measured (e.g., age) and unmeasured (e.g., social status) and unknowable (e.g., genetic differences) characteristics

Differences and similarities between health and social sciences

- Differences are few
 - » Health care trials are highly regulated, which increases their cost compared with social science trials – some new treatments can be life threatening; health care practitioners are expected to understand RCT evidence
- Similarities are common
 - » Both evaluate complex interventions; both can change the life chances of people; randomisation is the same; analysis is the same

A brief history of controlled trials

Use of control groups

- It has been known that a control group is required for many years
- *“On the 20th of May 1747, I took twelve patients in the scurvy, on board the Salisbury at sea....Two of them were ordered each a quart of cyder a-day. Two others took twenty-five gutts of elxir vitriol three times a day.. Two others had each two oranges and one lemon given them every day.”*

Scurvy

- *“The consequence was, that the most sudden and visible good effects were perceived from the use of the oranges and lemons; one of those who had taken them, being at the end of six days fit for duty.”*

Peninsular War 1815

- It had been so arranged, that this number [366] was admitted, alternately, in such a manner that each of us had one third of the whole. The sick were indiscriminately received, and were attended as nearly as possible with the same care and accommodated with the same comforts. One third of the whole were soldiers of the 61st Regiment, the remainder of my own (the 42nd) Regiment. Neither Mr Anderson nor I ever once employed the lancet. He lost two, I four cases; whilst out of the other third [treated with bloodletting by the third surgeon] thirty five patients died ([Lesassier Hamilton 1816](#)).*

Chalmers et al. (<http://www.jameslindlibrary.org/articles/the-advent-of-fair-treatment-allocation-schedules-in-clinical-trials-during-the-19th-and-early-20th-centuries/>)

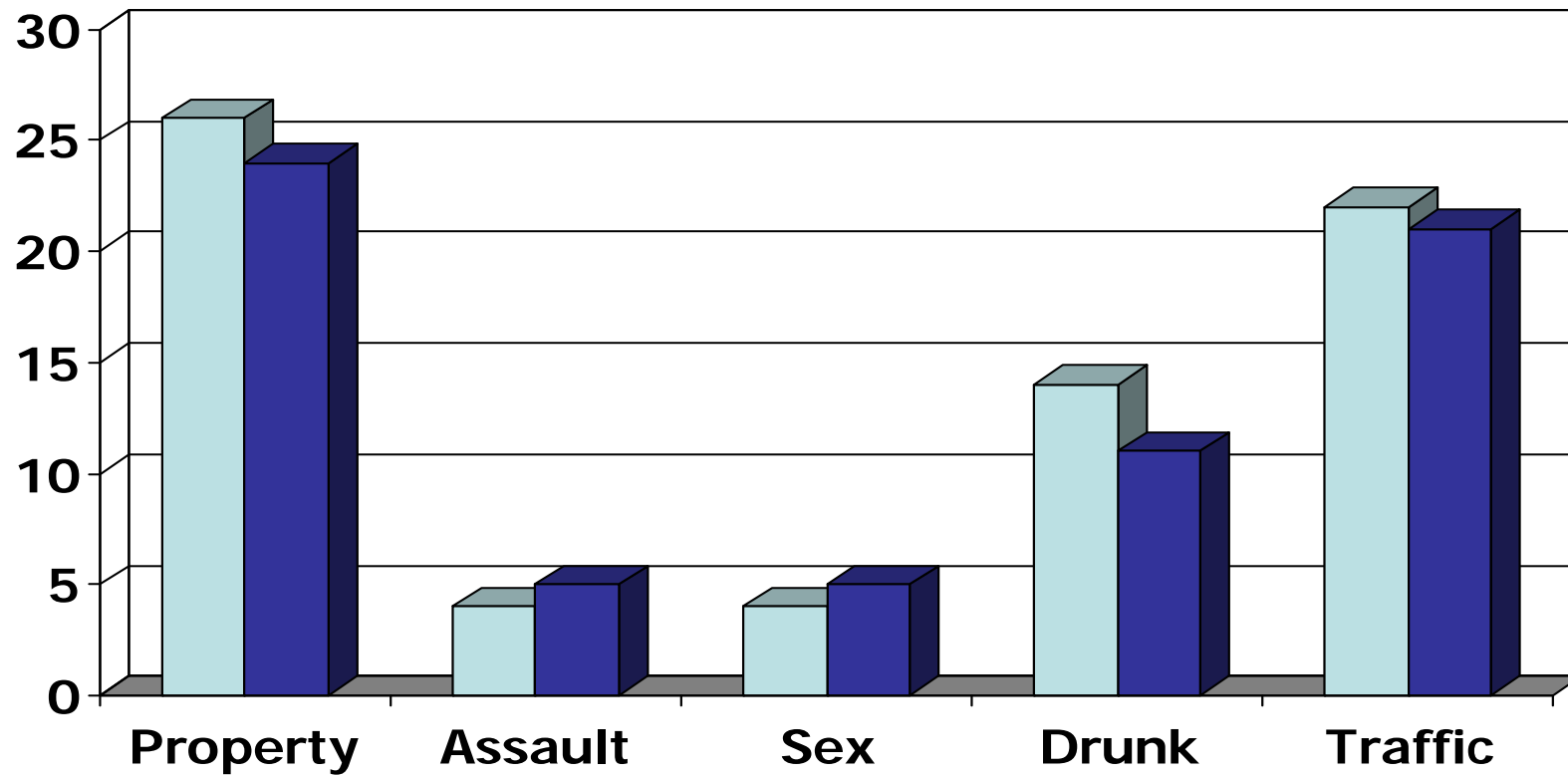
First known RCT

“Five seniors, each of whom had a good scholarship record, pleasing personality, excellent health and fine social environment, were chosen to act as personnel counselors for the members of the freshman class, who at the end of the first eight weeks of school happened to be delinquent in scholarship in the School of Mechanical Engineering in 1929-30. The 220 delinquent freshmen were divided into two groups by *random sampling*.” [italics added]

Cambridge-Somerville Youth Trial

- In 1937 a classic experiment – the Cambridge-Somerville trial was launched
- The aim was to show that social worker intervention among 'delinquent' boys would reduce 'criminality'
- 650 boys were identified by their teachers as having delinquent behaviour that put them at later risk of criminal activity
- 325 pairs were formed and one from each pair was allocated a social worker supported by psychiatrists

Results: early follow-up



Results: later follow-up

- In 1975 'boys' were followed up again when middle aged men
- 42% of intervention group had had a criminal conviction
- BUT only 32% of control group had had a conviction
- Another outcome was social workers refused to engage with RCTs

Some landmark experimental evaluations in education in the US and the developing world

- Cambridge Somerville youth experiment (US)
 - » Early intervention to reduce 'juvenile delinquency'
- Tennessee class-size experiment (US)
 - » Class-size reduction intervention in early years education
- *PROGRESA* experiment (Mexico)
 - » Rural anti-poverty intervention
- *Balsakhi* experiment (India)
 - » Teaching assistants intervention in literacy and numeracy

Streptomycin

- The second randomised trial in health care of a BENEFICIAL treatment was the 1948 MRC's streptomycin trial
- Streptomycin expensive antibiotic, unknown efficacy for pulmonary TB
- UK could not afford to buy much due to virtual post-war bankruptcy
- Randomisation was seen as fair method of allocating scarce drug

Streptomycin Results

	Streptomycin	Usual care
Improvement	69%	33%
No change	4%	6%
Worse	20%	34%
Death	7%	27%

1950s to 1970s

- The use of trials expanded rapidly within and beyond medicine
- In the social sciences experiments included
 - » Negative income tax
 - » Adoption
 - » Busing
 - » Public versus private schools
 - » Prevention of spousal abuse

Why more trials in health care?

- Many early social science trials were small and missed modest but important effects. Also at least 50% of RCTs show that something does not work, which policy makers do not like
- In health care the same problems but often trials expose harmful effects with mortality, which doctors and policy makers cannot ignore

Disasters among babies

- Routine practice in 40s and 50s to give premature infants pure oxygen. At the same time it was noted that there was an 'epidemic' of blindness among babies. Linked to oxygen use
- Routine practice in 50s to give prophylactic antibiotics to premature infants, caused brain damage and death
- BOTH of these problems only discovered AFTER an RCT was undertaken

Trial sabotage

- Interestingly an early trial of pure oxygen for neonates was sabotaged by nurses who secretly gave oxygen to some of the controls because they KNEW that it was effective
- Because of this ARROGANCE they contributed to the blinding of healthy babies

Pregnancy loss

- Routine practice to give a synthetic hormone to women at risk of miscarriage (DES)
- RCTs showed it to be completely useless
- WORSE years later it was shown that in female babies born to treated women, rare reproductive cancers were common also treated women had an increased risk of breast cancer

Heart Attacks

- After heart attack survivors can be placed into two groups:
 - » Those with arrhythmia & those without
 - » Arrhythmia patients have a much HIGHER risk of subsequent death
 - » There are drugs that can 'correct' cardiac arrhythmias
 - » These were widely used on millions of patients

CAST Trials

- For decades researchers tried to evaluate arrhythmia drugs
- Not ETHICAL argued many cardiologists
- Finally, CAST trial was started, terminated early in 1989 because of INCREASE in deaths on active drug
- 100000s died because of this

Could it still HAPPEN?

- High flow oxygen supplementation in acute COPD increases mortality (2010)
- Probiotics for acute pancreatitis (2008)
- Use of intravenous steroids after head injury increases mortality (2004)
- HRT INCREASES strokes, myocardial infarction and dementia (2002)
- Counselling women after difficult labour INCREASES depression (2001)

Publication Bias

- In 1981 a trial of an anti- arrhythmic drug was completed – showed an increase in mortality
- Authors tried unsuccessfully to get the study published (journals refused) only published in 1992, 11 years late – had it been published could have saved many lives

Non-publication of negative trials

- A key problem affecting the experimental literature in health and social sciences is failure to publish negative or null trials
- Pharmaceutical companies may leave 50% of trials unpublished
- Researchers delay or fail to publish many key trials



Population deworming every 6 months with albendazole in 1 million pre-school children in north India: DEVTA, a cluster-randomised trial

Shally Awasthi, Richard Peto, Simon Read, Susan M Richards, Vinod Pande, Donald Bundy, and the DEVTA (Deworming and Enhanced Vitamin A) team

Summary

Lancet 2013; 381: 1478–86

Published Online

March 14, 2013

[http://dx.doi.org/10.1016/](http://dx.doi.org/10.1016/S0140-6736(12)62126-6)

[S0140-6736\(12\)62126-6](http://dx.doi.org/10.1016/S0140-6736(12)62126-6)

See [Comment](#) page 1439

King George's Medical
University, Lucknow, Uttar
Pradesh, India
(Prof S Awasthi MD, V Pande MD,
DEVTA team); Clinical Trial
Service Unit and
Epidemiological Studies Unit
(CTSUS), University of Oxford,
Oxford, UK (Prof R Peto FRCS,
S Read PhD, S M Richards PhD);
and Human Development
Network, The World Bank,
Washington, DC, USA
(D Bundy PhD)

Correspondence to:
Prof Shally Awasthi, Department
of Paediatrics, King George's
Medical University, Chowk,
Lucknow, Uttar Pradesh 226003,
India
shally07@gmail.com

Background In north India many pre-school children are underweight, many have intestinal worms, and 2–3% die at ages 1·0–6·0 years. We used the state-wide Integrated Child Development Service (ICDS) infrastructure to help to assess any effects of regular deworming on mortality.

Methods Participants in this cluster-randomised study were children in catchment areas of 8338 ICDS-staffed village child-care centres (under-5 population 1 million) in 72 administrative blocks. Groups of four neighbouring blocks were cluster-randomly allocated in Oxford between 6-monthly vitamin A (retinol capsule of 200 000 IU retinyl acetate in oil, to be cut and dripped into the child's mouth every 6 months), albendazole (400 mg tablet every 6 months), both, or neither (open control). Analyses of albendazole effects are by block (36 vs 36 clusters). The study spanned 5 calendar years, with 11 6-monthly mass-treatment days for all children then aged 6–72 months. Annually, one centre per block was randomly selected and visited by a study team 1–5 months after any trial deworming to sample faeces (for presence of worm eggs, reliably assessed only after mid-study), weigh children, and interview caregivers. Separately, all 8338 centres were visited every 6 months to monitor pre-school deaths (100 000 visits, 25 000 deaths at age 1·0–6·0 years [the primary outcome]). This trial is registered at ClinicalTrials.gov, NCT00222547.

Findings Estimated compliance with 6-monthly albendazole was 86%. Among 2589 versus 2576 children surveyed during the second half of the study, nematode egg prevalence was 16% versus 36%, and most infection was light. After at least 2 years of treatment, weight at ages 3·0–6·0 years (standardised to age 4·0 years, 50% male) was 12·72 kg albendazole versus 12·68 kg control (difference 0·04 kg, 95% CI –0·14 to 0·21, $p=0·66$). Comparing the 36 albendazole-allocated versus 36 control blocks in analyses of the primary outcome, deaths per child-care centre at ages 1·0–6·0 years during the 5-year study were 3·00 (SE 0·07) albendazole versus 3·16 (SE 0·09) control, difference 0·16 (SE 0·11, mortality ratio 0·95, 95% CI 0·89 to 1·02, $p=0·16$), suggesting absolute risks of dying between ages 1·0 and 6·0 years of roughly 2·5% albendazole versus 2·6% control. No specific cause of death was significantly affected.

Interpretation Existing ICDS village staff can be organised to deliver simple pre-school interventions sustainably for many years at low cost, but regular deworming had little effect on mortality in this lightly infected pre-school population.

Funding UK Medical Research Council, USAID, World Bank (albendazole donated by GlaxoSmithKline).

Introduction

Heavy worm infection can constrain physical development.^{1,2} In areas where heavy infection is common, periodic deworming of school-age children can improve weight gain.^{1,5} Of three studies of urban pre-school children in Lucknow, north India, two suggested little

and sustainably deliver anthelmintics at operational scale to rural pre-school children in Uttar Pradesh. The primary aim of the trial was to assess effects of a widely practicable periodic deworming regimen on mortality at ages 1·0–6·0 years.¹²

Plans for this deworming trial were eventually revised

DEVTA: results from the biggest clinical trial ever



In developing countries health resources are short, so it is particularly important that science should inform policy. In *The Lancet*, a landmark trial^{1,2} tests two so-called magic bullets of the international health community's armoury: pre-school vitamin A supplementation, assumed to reduce child mortality by a quarter;³ and intestinal deworming, assumed to improve child nutrition, growth, and cognitive development.⁴

Richard Peto and colleagues taught us in the 1980s to "ask an important question, and answer it reliably" with large, simple randomised trials testing the effects on mortality of practicable treatments for common conditions.⁵ DEVTA epitomises this vision. It is the biggest drug trial ever undertaken, with 2 million pre-school children in India consuming four tonnes of vitamin A and eight tonnes of albendazole over 5 years. It is a credit to all participating: the scientists in Lucknow and elsewhere, the field staff, the anganwadi workers, and the mothers and children. DEVTA is of staggering size, but the delay in publication is equally staggering, in view of the important policy implications for India and the rest of the developing world. Data collection was completed in 2006 and preliminary results for vitamin A (but not deworming) were presented at a meeting in Turkey in 2007, with the presentation available online.⁶ Then silence until

now, punctuated only by calls from us and others for the team to publish.^{7,8} Today, the authors reveal that neither intervention showed a significant effect on child mortality in their trial.

Before DEVTA, the randomised evidence base showed that vitamin A reduced mortality by a quarter, based on 17 trials in 194 795 children.⁹ DEVTA, with ten times the number of children (although contributing only about twice as much information, because of cluster effects) estimates mortality reduction at 4%, with 95% CIs ranging from an 11% reduction to a 3% increase ($p=0.22$). On the basis of their meta-analysis, the trial authors propose that DEVTA attenuates the global estimates of mortality reduction by half. Their forest plot suggests heterogeneity, confirmed in a sensitivity analysis in the Cochrane review:⁹ so an alternative interpretation is that vitamin A does not affect child mortality in all settings all of the time. If we accept the authors' analysis that this trial simply reduces the global estimate of benefit by half, the question still remains as to why the maximum estimated mortality reduction of 11% is so substantially different from the estimate of the other 17 trials.⁹ Since there is clear evidence of vitamin A deficiency in the study site, this finding is unexpected. Had access to preventive and curative health-care interventions improved, as in many

Published Online
March 14, 2013
[http://dx.doi.org/10.1016/S0140-6736\(13\)60611-1](http://dx.doi.org/10.1016/S0140-6736(13)60611-1)
See [Articles](#) page 146

should invest—and invest heavily—in such studies, and not shy away from funding big trials because DEVTA did not show a difference.

And what can researchers learn from DEVTA? Undoubtedly, the fact that there was no apparent effect detected delayed publication. The best way to underline this lesson for us all is to modify the criteria of a good trial to: “Ask an important question, answer it reliably; and publish the results promptly, irrespective of the findings”.

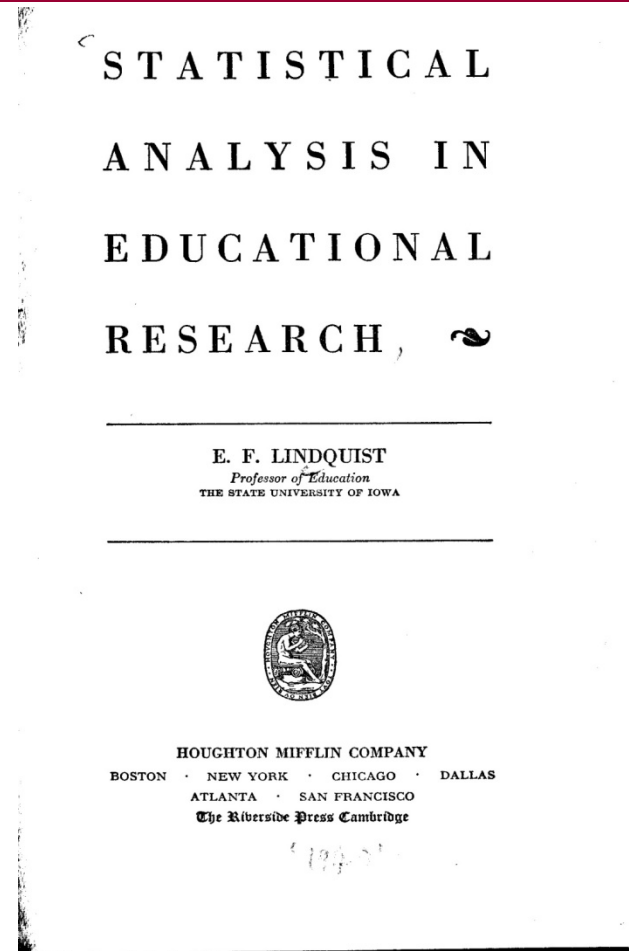
**Paul Garner, David Taylor-Robinson, Harshpal Singh Sachdev*
Liverpool School of Tropical Medicine, Liverpool L3 5QA, UK (PG);

Solutions?

- All health care trials have to be registered in public database – unpublished ones can be assumed to be null or negative
- Federal Drug Authority (USA) can fine \$10,000 a day for delayed publication
- In UK public funder (NHS HTA) withhold 10% of trial grant until publishable report has been delivered (>98% are published)

Lessons from education

- In 1940 Lindquist demonstrated that when groups (classes, or schools) of individuals are randomised cluster level analysis should be undertaken
- In health care – Cornfield argued in 1978 that:
 - » “Randomisation by cluster accompanied by analysis appropriate to randomisation by individual is an exercise in self-deception”
- It wasn't until late 1990s that biostatisticians routinely used appropriate statistical methods



Cornfield, 1978 Am J Epidemiol 108:100-102

Lessons from social sciences: Impact of attrition

- Most health care trials undertake baseline comparisons of randomised groups (either descriptively or worse inferentially)
- Social science trials look at baseline characteristics of 'analysed groups' which assess balance after attrition
- The latter is more useful than the former

Lessons from social sciences: Impact of non-compliance

- Health care trials tend to do either
 - » Per protocol analysis or on treatment analysis – both of which are usually biased as well as intention to treat analysis, which is not.
- Social science trials tend to undertake:
 - » Instrumental analysis or complier average causal effect (CACE) analysis – which is usually unbiased.
- Health care trialists are slowly adopting CACE

Allocation concealment

- Most social science trials do not describe how randomisation was performed and was it done by a third party?
- Health care trials have to do this and generally follow the CONSORT statement in their reporting
- Recent social science trials are reporting more fully the randomisation process and adopting CONSORT

Comparison of adequate, unclear and inadequate concealment

Allocation Concealment	Effect Size OR	
Adequate	1.0	
Unclear	0.67	P < 0.01
Inadequate	0.59	

[Schulz et al. *JAMA* 1995;273:408]

Conclusions

- The use of RCTs in health care has transformed the treatment of patients and can transform the well being of society in general
- There are some 'silos' between methodologists and lessons learnt about how to do robust trials are not communicated quickly across different areas

Education trials in 20th Century

- 1900 to 1960s:
 - » many 'explanatory' experiments in educational psychology, probably sometimes using randomisation
 - » a few 'pragmatic' experiments (e.g., Pearson, 2011)
- 1931: first known RCT in modern period - Walters' counselling experiment
- 1932: Walters' replication trial

Education trials in 20th Century

- 1930s to 1970s: Many RCTs in education (US) (some large scale), but dearth of high quality RCTs in education research (UK)
- 1970s to 2000s: few large scale RCTs in education (US), notable exceptions:
- 1985: Tennessee class-size experiment

Education trials in 21st Century

Author, date	Topic area	Total RCTs	US	UK	Other
Torgerson and Elbourne, 2002	ICT and spelling	7	6		1 (Austr.)
Torgerson, King and Sowden, 2003	Volunteering and early reading	10	8	2	
Torgerson, Porthouse and Brooks, 2003	Adult literacy and numeracy	9	8	1	
Torgerson, See and Andrews, 2006	Argumentation in HE	10	10		
Torgerson, Brooks and Hall, 2006	Phonics instruction	12	11	1	

Education trials in 21st Century

- 2002: US legislation to create Institute of Education Sciences (IES)
- 2009: UK government funding of first RCT evaluation of curriculum intervention (Every Child Counts Trial)
- 2011: UK government setting up of Education Endowment Foundation (EEF)
- 2002 to present: 200+ large scale RCTs in education (US); 100+ RCTs in education (UK)



US legislation: Institute of Education Sciences

One Hundred Seventh Congress
of the
United States of America

AT THE SECOND SESSION

*Began and held at the City of Washington on Wednesday,
the twenty-third day of January, two thousand and two*

An Act

To provide for improvement of Federal education research, statistics, evaluation,
information, and dissemination, and for other purposes.

“Scientifically valid educational evaluation employs experimental designs using random assignment, when feasible, and other research methodologies that allow for the strongest possible causal inferences when random assignment is not feasible.” [page 5]

[italics added]



Education Endowment Foundation (EEF)



“...all EEF-funded projects are independently and rigorously evaluated...The impact of projects on attainment will be evaluated where possible, using randomised controlled trials.”
[italics added]

Recent developments in social science

- Last 10-15 years - large impetus to increase trials into education and social sciences in the UK and US
- UK Government invested £13+ million to support education RCTs
- Work and Pensions doing RCTs; Tax office RCTs; Criminal Justice RCTs – driven by Behavioural Sciences Unit in Cabinet Office

Test, Learn, Adapt:

Developing Public Policy with Randomised Controlled Trials

Laura Haynes

Owain Service

Ben Goldacre

David Torgerson



CabinetOffice

Behavioural Insights Team

http://www.consort-statement.org/


File Edit View Favorites Tools Help

Consort - Welcome to the ...

CONSORT TRANSPARENT REPORTING OF TRIALS

Search Sign In

Home CONSORT 2010 Extensions Downloads Examples Resources About CONSORT






Welcome to the CONSORT Website

CONSORT stands for Consolidated Standards of Reporting Trials and encompasses various initiatives developed by the CONSORT Group to alleviate the problems arising from inadequate reporting of randomized controlled trials.


The CONSORT Statement

The main product of CONSORT is the [CONSORT Statement](#), which is an evidence-based, minimum set of recommendations for reporting randomized trials. It offers a standard way for authors to prepare reports of

CONSORT 2010 Key Documents

- ☒ [CONSORT 2010 Checklist](#)
-  [CONSORT 2010 Flow Diagram](#)
-  [CONSORT 2010 Statement](#)
-  [CONSORT 2010 Explanation and Elaboration Document](#)

Tweets

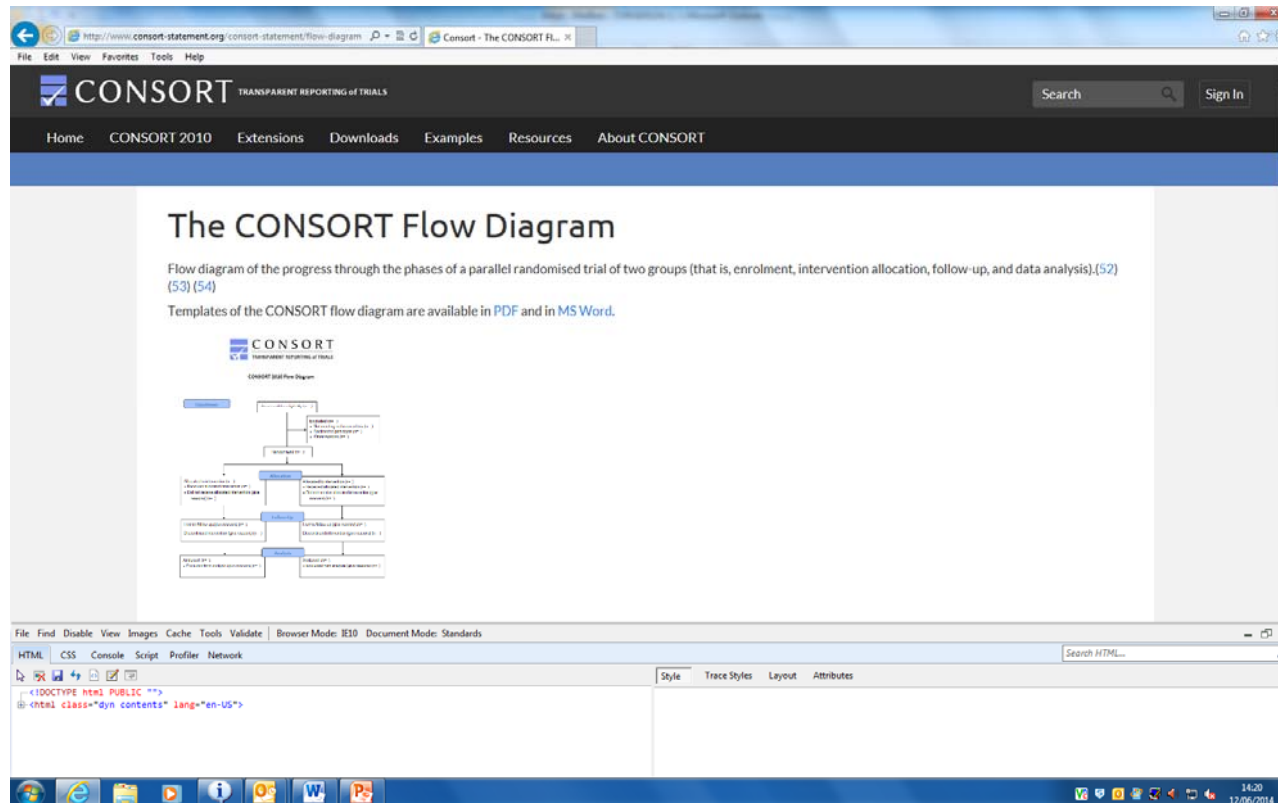
 **BMC Medical Evidence** @MedicalEvidence 30 May
From Miró to Pollock - What impact has the @CONSORTing Statement had & what was discussed at the latest meeting? bit.ly/1prdcw3
13 Retweeted by CONSORT Statement

11:56 11/05/2014

The CONSORT guidelines, adapted for trials in educational research

- Was attrition from the trial documented at all stages using a CONSORT flow diagram?
- Was the randomisation process concealed from the investigators? (i.e. were the researchers who were recruiting children to the trial blind to the child's allocation until after that child had been included in the trial?)
- Was intention to treat analysis used? (i.e. were all children who were randomised included in the follow-up and analysis?)
- Were follow-up measures administered blind? (i.e. were the researchers who administered the outcome measures blind to treatment allocation?)

CONSORT flow diagram



Education: comparison with health education

Characteristic	Health Ed	Education Ed
Cluster Randomised	36%	18%
Sample size justified	28%	0%
Concealed randomisation	8%	0%
Blinded Follow-up	30%	14%
Use of CIs	41%	1%
Low Statistical Power	41%	85%

Torgerson CJ, Torgerson DJ, Birks YF, Porthouse J. (2005) A comparison of randomised controlled trials in health and education. British Educational Research Journal,31:761-785. (based on n = 168 trials)

Lessons learnt

1. Funders and developers interfere
 - » Potential for conflict of interest has to be managed
2. Most educational interventions do not work
 - » Lack of support for negative trials
3. Recruitment and retention can be challenging
 - » Acceptability of intervention (Buttle and Teensleep)
4. Explaining random allocation effectively is critical (SHINE)
5. Most barriers to undertaking RCTs have design solutions
 - » Essential to have methodologist on trial team

Some examples of expensive educational interventions that don't work (UK)

- Financial incentives (Brooks *et al*, 2008)
- ICT and spelling (Brooks *et al*, 2006)
- Nurse numeracy intervention (Ainsworth *et al*, 2011)

Every Child Counts Trial

- *Every Child Counts (ECC)* was previous UK government's flagship policy to help children (age 7) at risk in numeracy
- Expensive one-to-one tutoring intervention (*Numbers Count*) delivered each day over one school term (12 weeks)
- Randomised experiment commissioned to establish effectiveness and cost-effectiveness (Torgerson *et al*, 2013a; Torgerson *et al*, 2013b)
- Pre-post test evaluation (undertaken by developer) demonstrated large effect size (>1 SD) and cost-effectiveness (using weak design)

Results

- *ECC* better than business as usual (0.33 SD) but expensive
- No evidence that one-to-one was better than *ECC* in pairs or triplets
- One-to-one *not* cost effective – more cost-effective to deliver in pairs or triplets

Design features that minimise alternative explanations for results

- Large sample size: excludes chance finding
- Randomisation: intervention and control groups are equivalent at start so design controls for history, maturation, regression to the mean, selection bias
- Intervention and control conditions are both numeracy interventions and both last for 30 minutes each day for 12 weeks: the comparison is a 'fair' one
- Independent 'blinded' testing: eliminates possibility of tester bias

Design limitations: Generalisability

- *ECC* schools were identified: by policy-makers/funders of programme - education policy 'roll out' in England, i.e., schools in disadvantaged areas
- Ideally, a random sample of *all* primary schools in England should have been approached and asked to take part

Design limitations: Intervention

- One-to-one teaching with intervention children being withdrawn from classroom
- Problem of attribution: was effect due to *Numbers Count* intervention? Or to one-to-one teaching? Or to withdrawal from classroom?
- Design could have included additional one-to-one arm

Design limitations: 'Contamination'/'spill over' effects

- Children withdrawn from usual classroom teaching – may have benefited remaining children
- Teachers using intervention have applied it to some control children.
- Instead of randomising *individual children*, design could have randomised by *school* (cluster randomisation, where school is the cluster) to avoid these problems.

Design limitations: Long term effects

- Wait list design prevented long term follow-up; effects may have 'washed out' soon after intervention was finished.
- Could have used cluster randomisation;
- Could have recruited children above threshold and randomised these to intervention or long term follow-up;
- **All options (above) rejected by funder.**

Conclusions

- Design and conduct warranted conclusion *Numbers Count* (as delivered) more effective than usual classroom teaching BUT because of design limitations couldn't answer some really important questions
- These questions could have been answered if a different experimental design had been used: cluster randomisation (randomisation of schools), long-term follow-up (control group that didn't receive intervention); one to one control group (literacy or other numeracy)
- Interference from policy makers prevented a more informative trial being done

References

- Ainsworth, H., Torgerson, D., Torgerson, C. *et al* (2011) Computer-based instruction for improving student nurses' general numeracy: Is it effective? Two RCTs, *Educational Studies*
- Brooks, G., Burton, M., Coles, P., Miles, J., Torgerson, C., Torgerson, D. (2008) Randomised controlled trial of incentives to improve attendance at adult literacy classes, *Oxford Review of Education*, 34(4)
- Brooks, G., Miles, J.N.V., Torgerson, C.J. and Torgerson, D.J. (2006) Is an intervention using computer software effective in literacy learning? A randomised controlled trial, *Educational Studies*, 32(1)
- Torgerson, C.J., Wiggins, A., Torgerson, D.J., Ainsworth, H., Hewitt, C. *Every Child Counts*: Testing policy effectiveness using a randomized controlled trial, designed, conducted and reported to CONSORT standards, *Journal of Research in Mathematics Education*, July 2013
- Torgerson, C.J., Wiggins, A., Torgerson, D.J., Ainsworth, H., Hewitt, C. The effectiveness of an intensive individual tutoring programme (Numbers Count) delivered individually or to small groups of children: A randomised controlled trial, *Effective Education*, Apr., 2013
- Torgerson, C.J., Wiggins, A., Torgerson, D.J., Ainsworth, H., Hewitt, C. The effectiveness of an intensive individual tutoring programme (Numbers Count) delivered individually or to small groups of children: A randomised controlled trial, *Effective Education*, Apr., 2013
- Torgerson, C.J. and Torgerson, D.J. (2007) The need for pragmatic experimentation in educational research, *Economics of Innovation and New Technology*, 16(5 & 6)
- Torgerson CJ, Torgerson DJ. The Need for Randomised Controlled Trials in Educational Research. *British Journal of Educational Studies* 2001;49:316-28.
- Torgerson DJ, Torgerson CJ. *The design of randomised trials in health, education and the social sciences*. Palgrave Macmillan, Basingstoke, 2008