The need for using theory to consider the transferability of interventions

Chris Bonell

LONDON
SCHOOL of
HYGIENE
&TROPICAL
MEDICINE

Randomised trials

Randomising individuals/clusters of individuals to intervention/comparison arms reduces bias

Randomising also aims to balance arms on measured and unmeasured factors to reduce confounding

In a non-random schools study for example the schools already taking more effective action to reduce bullying might opt for the intervention
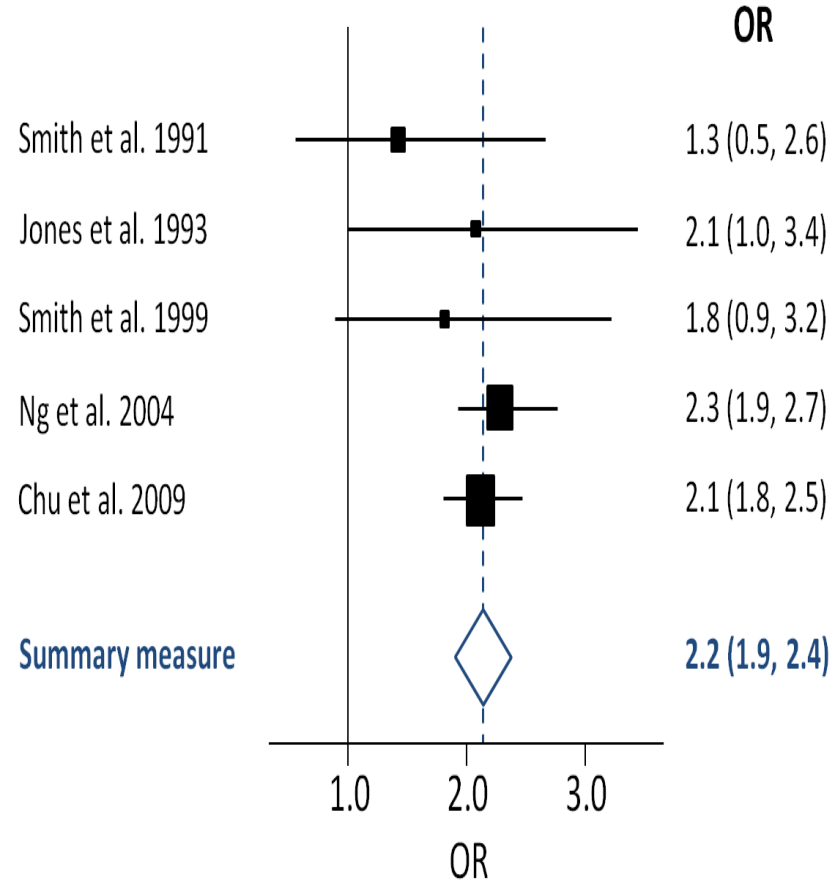
Then aim to minimise error and bias in measurement and analysis e.g. retention, blinding, validated measures, intention-to-treat analysis
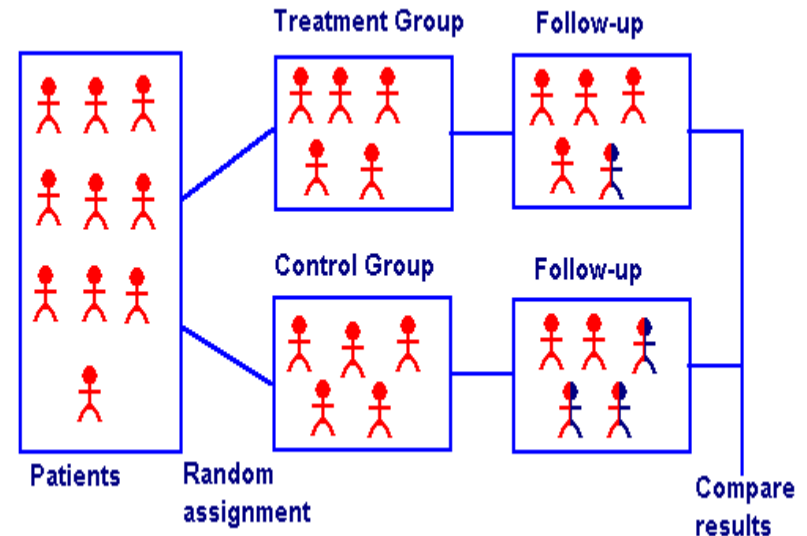
Systematic reviews

Trials generate effect estimates in situ usually for overall population

Systematic reviews pool effect estimates of studies homogeneous for PICO (population, intervention comparisons and outcomes) to produce general effect estimate

|	OR
Smith et al. 1991	1.3 (0.5, 2.6)
Jones et al. 1993	2.1 (1.0, 3.4)
Smith et al. 1999	1.8 (0.9, 3.2)
Ng et al. 2004	2.3 (1.9, 2.7)
Chu et al. 2009	2.1 (1.8, 2.5)
Summary measure	2.2 (1.9, 2.4)

OR

1.0   2.0   3.0

Trials and reviews don't tell us whether the intervention will work elsewhere

Trial reports often make bland statements about uncertain generalisability

Systematic reviews pool trial results more often than explaining differences in results between trials

Trials in meta-analysis may be homogeneous for PICO but heterogeneous for other factors that moderate intervention effects

Moderators
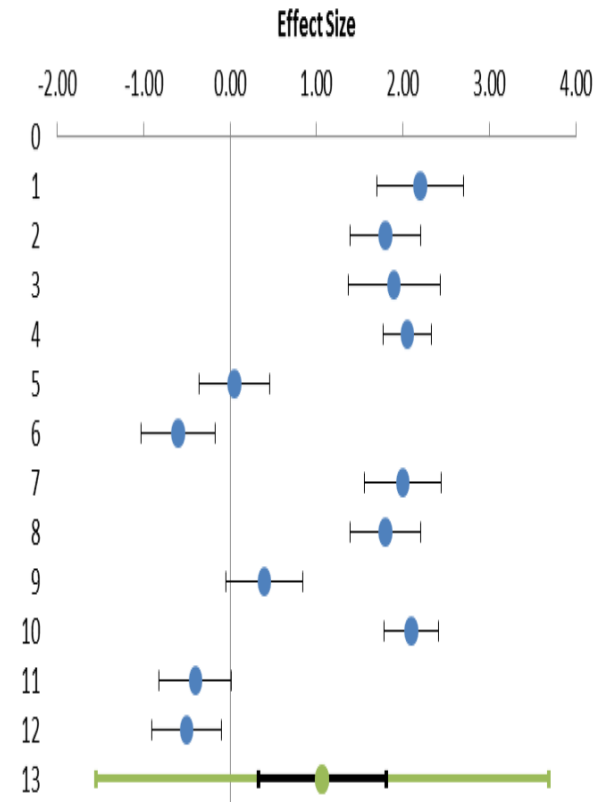
Variables that directly measure or act as a proxy measure for some factor modifying the effect of allocation to receive an intervention on an outcome.

Can explain

-differences in effect between sub-groups within a trial or

-differences in the effect found in one trial and that found in another trial (or that which would be found were the intervention delivered elsewhere)
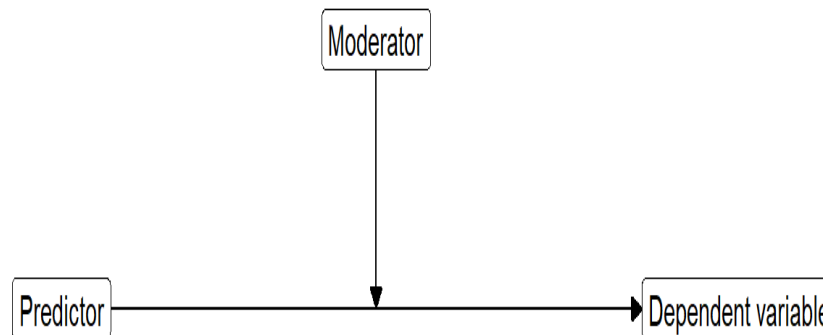
Example: sexual health promotion with MSM

Peer education effective in reducing risk of STIs among MSM in small US mid-western cities in early 1990s

Not effective in reducing risk among Scottish MSM in late 1990s – why?

=Inadequate fidelity?

=Difference in moderators?

- Low knowledge less prevalent and/or weaker moderator in late 1990s?
- Drug use more prevalent and stronger moderator in late 1990s?
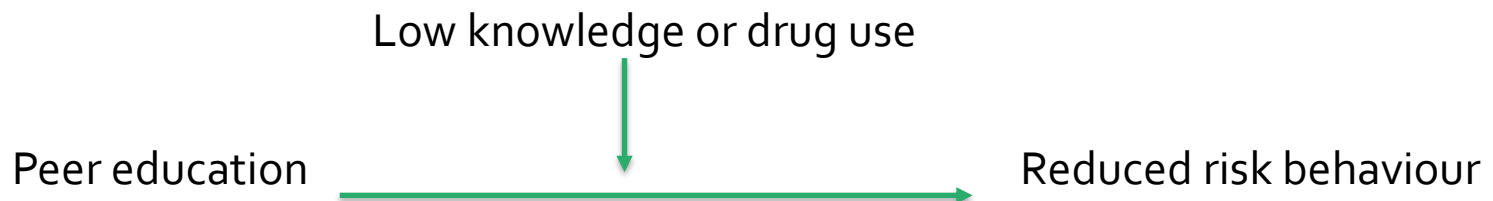
Low knowledge or drug use

Peer education ———————————————→ Reduced risk behaviour

Implications for assessing transferability

You could reanalyse the original trial data, reweighting strata defined by within-trial moderators to take account of different prevalence of moderators in the new context

e.g. model overall effect to take account of different prevalence of low knowledge

-requires evidence about prevalence of moderators in trial and in new context

Moderator	Original trial		New setting	
	Effect	Prevalence	Effect	Prevalence
Low knowledge	OR=0.5	75	OR=0.5	25
High knowledge	OR=1.00	25	OR=1.00	75
Overall	Or=0.625	100	OR=0.875	100

-addresses issue in difference between contexts in prevalence but not strength of moderators or existence of new moderators

Implications for assessing transferability

You could reanalyse the original trial data, reweighting strata defined by within-trial moderators to take account of different strength of moderators in the new context

Moderator	Original trial		New setting	
	Effect	Prevalence	Effect	Prevalence
Low knowledge	OR=0.5	75%	OR=0.75	75%
High knowledge	OR=1.0	25%	OR=1.0	25%
Overall	OR=0.625	100%	OR=0.8	100%

-but you won't know the strength of moderator across different contexts

-we might guesstimate this based on observational research from new context e.g. association between drug use and sexual risk - but very crude

Need to generalise on basis of theory

Statistical modelling challenging:

- we won't know which moderators to examine in trials
- we will never have perfect information about prevalence and strengths of these moderators in new settings to develop precise estimates

And we don't understand what real mechanisms of action underlie – e.g. is low knowledge a moderator because mechanism concerns education or is it a proxy for something else?

Therefore important complement to statistical modelling is theory

Realist approaches to social science and evaluation provide useful insights

Realist evaluation

Informed by critical realist philosophy

Interventions viewed as providing resources for actors who will change their practices to trigger mechanisms which generate outcomes

Implementation and mechanisms will vary by context (person or place) and therefore so will outcomes

Causation is unobservable but real

Measures are observable but indirect window on reality
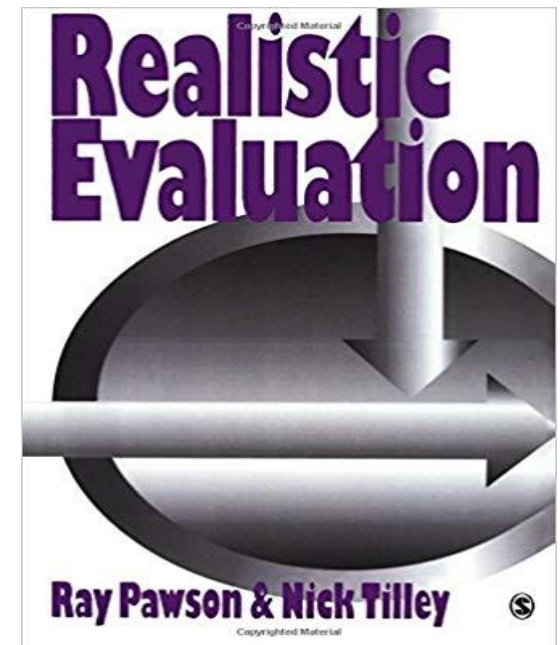
Realist evaluation

Realist evaluators formulate hypotheses about how mechanisms interact with context to generate outcomes (CMO configurations)

Test these hypotheses using natural experiments

For example, CCTV in carparks working via increasing natural surveillance, removing core offenders or signalling priority

Realists against RCTs but their insights could inform interpretation of trial evidence in order to predict effectiveness of interventions in new settings

CMO Configurations
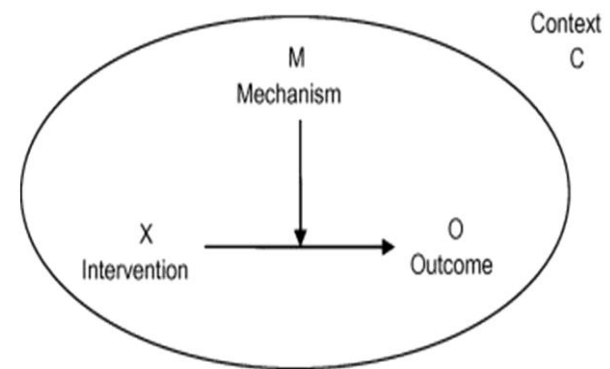
Goals of realist intervention research

Conventional evaluation goal of estimating intervention effects and understanding what factors moderate this: what works for whom under what conditions

Less conventional (at least within evaluation research) goal of using RCTs and systematic reviews to build and refine scientific theories about:

- how interventions trigger mechanisms that then generate outcomes
- how these mechanisms will vary
   with context

Might provide more informed
predictions about transferability
of effects

Potential for realist RCTs?

Use qualitative research nested within RCT to build/refine CMO hypotheses

Test these in additional trial analyses e.g. moderator and mediator analyses

Randomised trials can encompass sufficient variety of contexts to test some but not all CMOs

(variation in context can exist as long as it is similar in each arm)

Systematic reviews can encompass more variety of contexts to test more CMOs

Example: INCLUSIVE trial

Conventional basic design with post-baseline random allocation of 40 schools across SE England (2014-7)

Single blinded follow up of students in year 7 t baseline to 24 and 36 months

Encompass variation in students (SES, ethnicity, sex) and schools (e.g. good/poor inspection rating, inner city/suburban)

Examine intervention which 'disrupts' school environment via multiple mechanisms

Starting theory of change

'Intervention' provides data, manual, training, facilitator, curriculum materials

Teachers use these to implement student-staff policy-making, restorative practices, social & emotional skills lessons (*especially in schools with higher management capacity*)

Teachers and students develop better relationships and new skills, and build student commitment to school (*especially of low SES students*)

Reduced student engagement with anti-school peer groups and health risk behaviours including bullying (*especially in schools with high capacity / more students of low SES*)

Trial analyses

Overall intervention effective in

Reductions in:

-bullying (but not school based aggression

-smoking

-alcohol

-drug use

But not:

-school based aggression

-sexual risk behaviour

 Improvements in:

-mental health

-psychological functioning

-quality of life

Outcome	Mean difference (95% CI)
GBS overall score	-0.03 (-0.06, 0.00)
Teasing	-0.05 (-0.10, 0.00)
Rumours	-0.07 (-0.11, 0.02)
Deliberate exclusion	-0.04 (-0.08, -0.01)
Threatened or hurt	-0.01 (-0.03, 0.05)

-.11   0   .11

Outcome	Odds Ratio (95% CI)
Ever smoked	0.58 (0.43, 0.80)
Ever drunk alcohol	0.72 (0.56, 0.92)
Ever really drunk	0.47 (0.31, 0.71)
Contact with police	0.74 (0.56, 0.97)

.25   .5   1   2

Other analyses
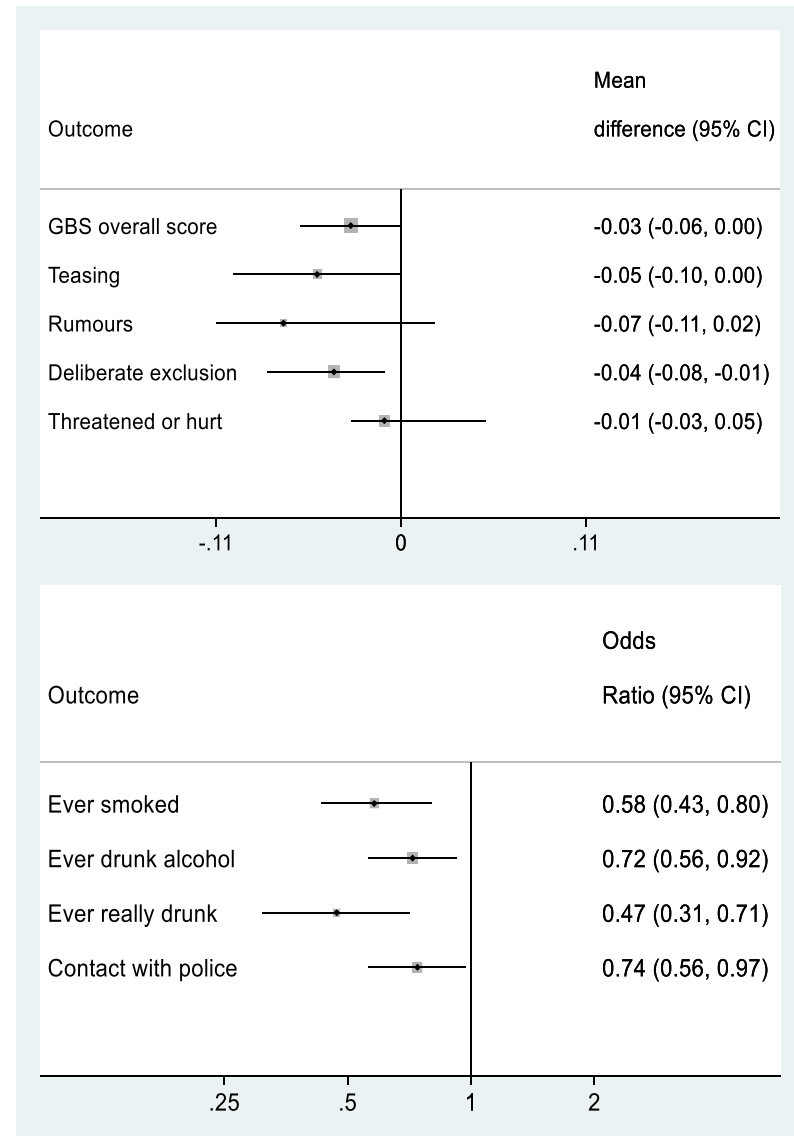
Process evaluation

- better fidelity in schools with more capacity but also with more baseline orientation towards student inclusion or holistic development
- social & emotional skills curriculum very poorly delivered
- intervention encouraged students and staff to see each other's points of view

Mediation analyses

- effects on bullying might be partly explained by increased commitment to school and reduced engagement with anti-school peers

Moderation analyses

- no evidence that more effective for students of low SES but evidence that more effective for boys and those reporting baseline bullying victimisation

Exploratory analyses of other outcomes

- reduced truancy, aggression in/outside school, involvement in school discipline systems

Refined theory of change

'Intervention' provides data, manual, training, facilitator but not curriculum materials

Teachers use these to implement student-staff policy-making, restorative practices, (*especially in schools with high capacity & inclusive ethos*) but not social & emotional skills lessons

Teachers and students develop more empathy, and build student commitment to school (*especially of boys and those involved in risk*)

Reduced student engagement with anti-school peer groups and multiple health and educational risk behaviours (*especially in schools with high capacity / inclusive ethos / more boys / more baseline risk*)

Transferability

Use refined theory to predict effects (description not quantitative estimate) in other settings, for example:

- Intervention stronger candidate in

- inclusive schools
- schools with more boys and baseline risk

- Need to focus on activities that allow staff and students to build empathy

- Social & emotional skills education not a key ingredient in this intervention

- Need to work harder on gender equity through refined or additional intervention

- Might be an intervention for school improvement not just public health – more marketable to schools

Better:

understanding of how outcomes vary between people and places

understanding of actual mechanisms underlying these from iterative qualitative and quantitative analyses

predictions of potential effectiveness in new settings

as theory refined iteratively across trials in diverse settings, we might move from descriptive predictions of effectiveness to more precise quantitative estimates of effects
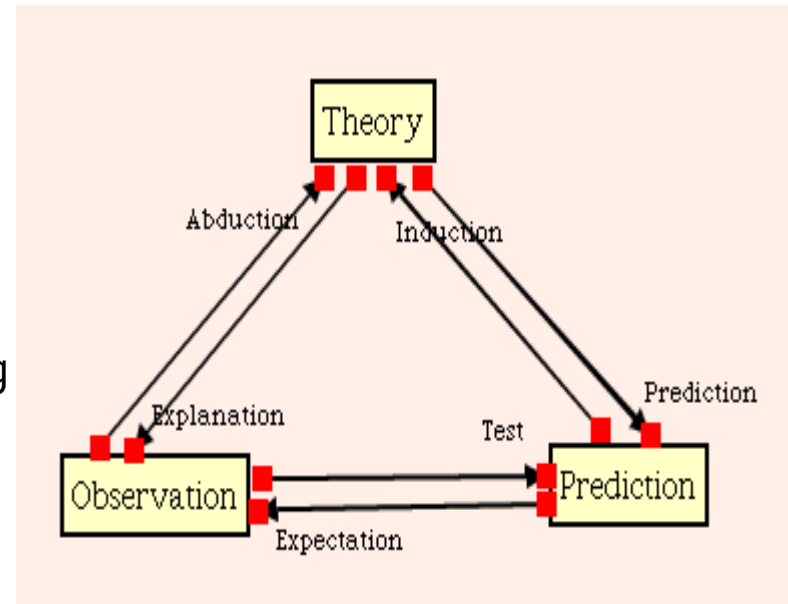
Potential limitations

Larger RCTs powered to assess moderators cost more money

Within trial, won't always have the quantitative measures to test hypotheses emerging from qualitative research e.g. inclusive culture

Hard to get multiple studies to focus on similar mechanisms across divergent contexts using comparable measures

More analysis mean more false positives but at least hypothesis-led– could add limited no. of new hypotheses to amended protocol to minimise bad practice

None of these problems are specific to randomised trials

Thank you – any questions?

chris.bonell@lshtm.ac.uk