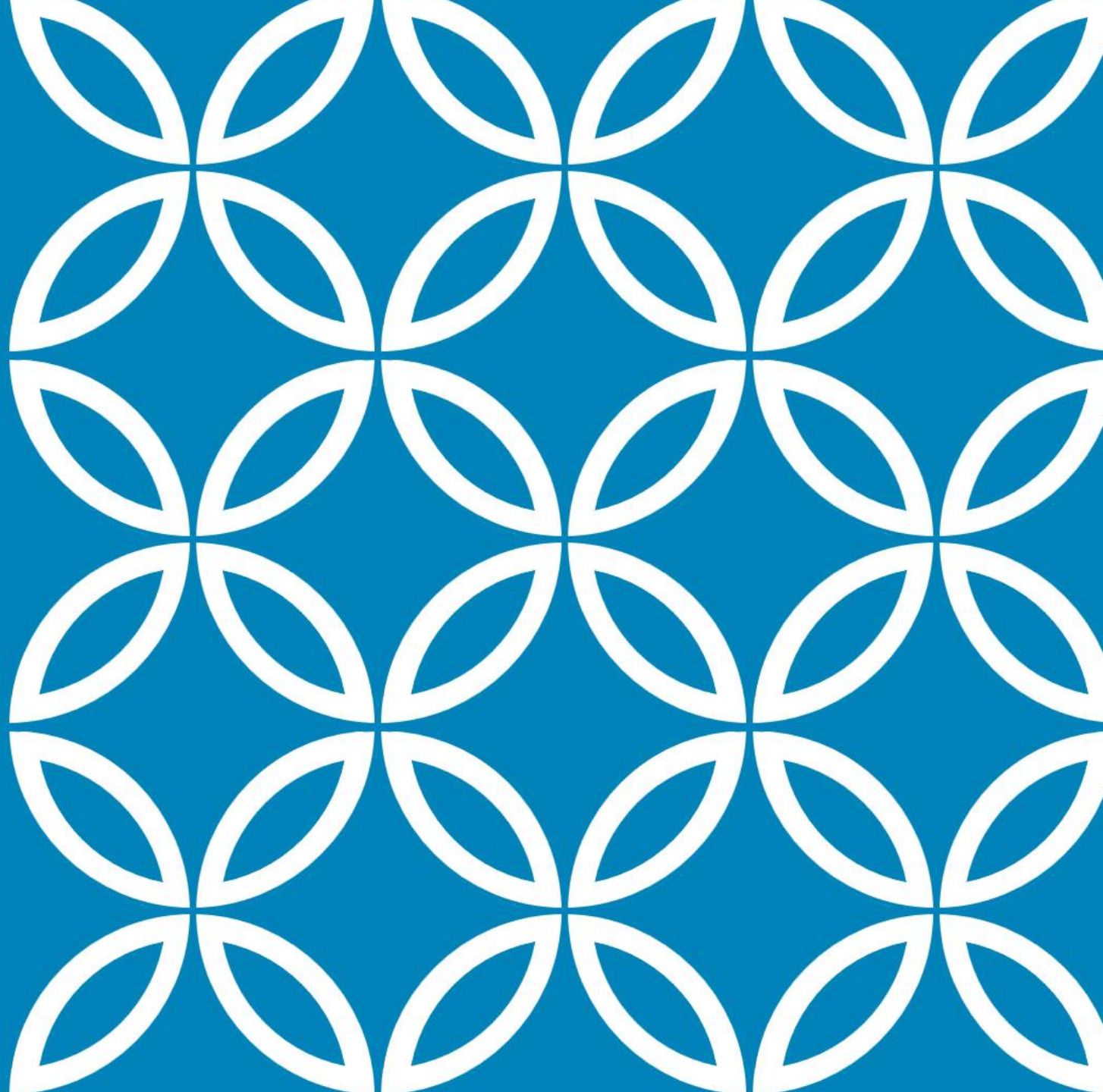
WHAT CAN HEALTH AND SOCIAL SCIENTISTS LEARN FROM ONE ANOTHER'S REPLICATION EFFORTS?

Hugh Sharma Waddington

5 December 2023



HOW CAN WE QUANTIFY BIAS IN REPLICATION?

Two main ways of measuring bias:

1. Measure distance from benchmark study in **internal replication** (within-study comparison)
2. Measure distance from benchmark in **external replication** (incorporating information about risk of bias)

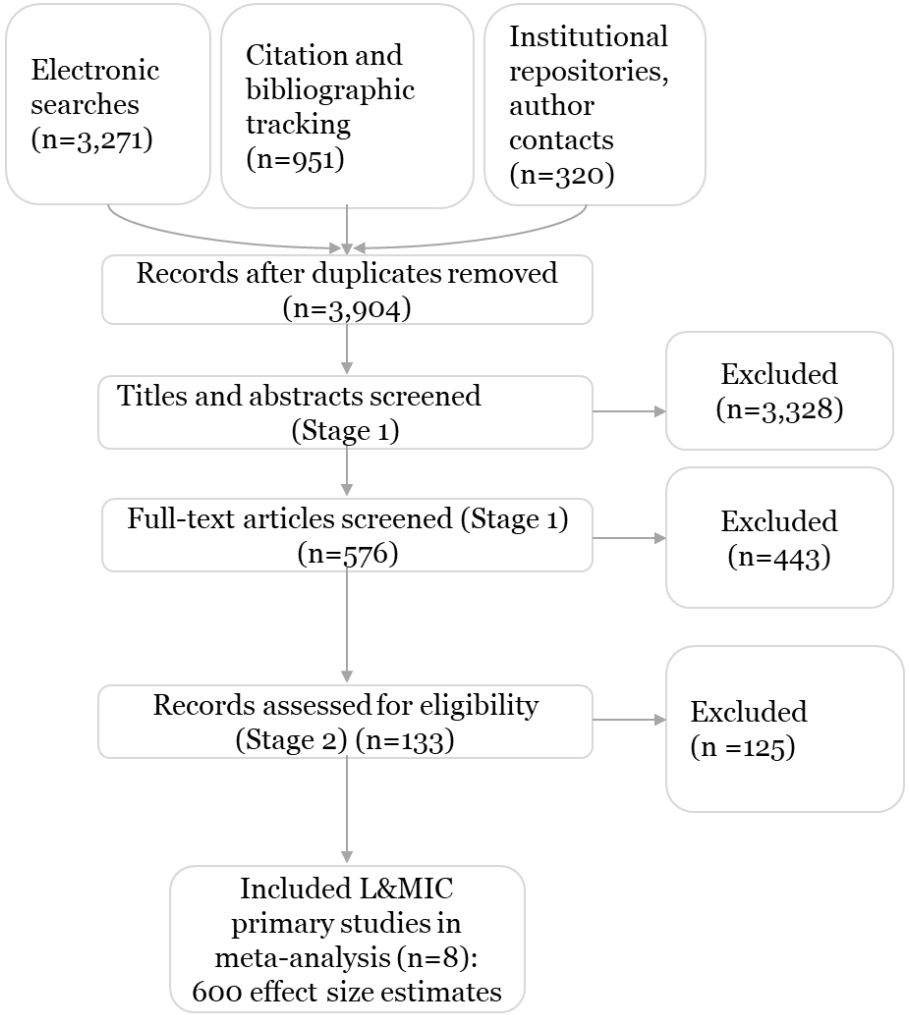
where the benchmark study is a well-conducted RCT, and the replications are quasi-experimental designs (QEDs), also called non-randomised studies of interventions (NRSI)

WHY RESULTS FROM RCTS AND QED/NRSI DIFFER

1. Bias in parameter estimate (internal validity)
   - Main sources of bias in RCTs: subversion of randomisation (confounding), attrition (selection bias), motivation bias
   - Main sources of bias in QEDs: confounding, selection bias, selection of the reported result
2. Sampling bias (external validity)
   - Discrepancy in treatment effect estimands due to differences in the target population (e.g. average treatment effect vs. local average treatment effect)

SYSTEMATIC REVIEW OF INTERNAL REPLICATIONS IN DEVELOPMENT ECONOMICS

Study	Intervention	QED replication
Buddelmeyer and Skoufias (2004)	Cash transfer (PROGRESA)	RDD
Diaz and Handa (2006)	Cash transfer (PROGRESA)	OLS, matching
Handa and Maluccio (2010)	Cash transfer (RPS)	Matching
McKenzie et al. (2010)	Immigration entitlement	DID, IV, OLS, matching
Galiani and McEwan (2013)	Cash transfer (PRAF)	RDD
Barrera-Osorio et al. (2014)	Scholarship	RDD
Chaplin et al. (2017)	Subsidy	Matching
Galiani et al. (2017)	Cash Transfer (PRAF)	GDD



Source: Sharma Waddington, Villar and Valentine 2022 Evaluation Review

RESULTS FROM FIXED EFFECT META-ANALYSIS

Standardised bias $|D_i| = \frac{|\bar{Y}^C_{NRS} - \bar{Y}^C_{RCT}|}{S_{RCT}}$; Mean squared error $MSE_i = D_i^2 + s_i^2$; % bias removed $|D_R| = \left(1 - \frac{\bar{Y}^C_{NRS} - \bar{Y}^C_{RCT}}{|\bar{Y}^C_{PF} - \bar{Y}^C_{RCT}|}\right) x\ 100$

Estimator	Standardised bias	Mean squared error	Percent of bias removed	Num. of estimates
Adjusted regression (cross-section data)	0.23	0.18	34%	10
Baseline adjustment (panel data) (DID, PSM)	0.05	0.01	56%	17
Discontinuity design	0.01	0.00	95%	173
Interrupted time series	-	-	-	-
Instrumental variables (strong instrument)	0.01	0.00	95%	1
Instrumental variables (weak instrument)	0.18	0.14	-92%	2
Matching (nearest neighbour)	0.04	0.07	52%	59
Matching (kernel)	0.14	0.15	34%	70

Source: Sharma Waddington, Villar and Valentine 2022 Evaluation Review

CAVEAT: ALL COMPARISONS AT SOME RISK OF BIAS

Risk of bias from within study comparison	Buddelmeyer and Skoufias (2004)*	Diaz and Handa (2006)*	Handa and Maluccio (2010)**	McKenzie et al. (2010)***	Barrera-Osorio et al. (2014)****	Galiani and McEwan (2013); Galiani et al. (2017)*****	Chaplin et al. (2017)
Confounding bias due to randomisation process	Some concerns	Some concerns	Some concerns	Some concerns	Low risk	Low risk	Some concerns
Selection bias in recruitment	Some concerns	Some concerns	Some concerns	Some concerns	Low risk	Low risk	Some concerns
Attrition bias due to missing outcome data	High risk	Some concerns	Some concerns	Some concerns	Some concerns	Low risk	Low risk
Departures from intended intervention	Some concerns	Some concerns	Low risk	Low risk	Low risk	Low risk	Some concerns
Bias in measurement of the outcome	Some concerns	Some concerns	Some concerns	Some concerns	Low risk	Low risk	Low risk
Selective analysis and reporting	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk	Low risk
Bias in NRS estimate	Low risk	Some concerns	Some concerns	Low risk	Low risk	Some concerns	Low risk
Overall bias in within-study comparison	**High risk**	**Some concerns**	**Some concerns**	**Some concerns**	**Some concerns**	**Some concerns**	**Some concerns**

Source: Sharma Waddington, Villar and Valentine 2022 Evaluation Review

CAN WE PREDICT THE DIRECTION OF BIAS IN EXTERNAL REPLICATIONS?

- We might expect absolute effect size magnitudes from NRSI to **exceed** effects from RCTs:
  - Non-adherence in trials (i.e., Intent-to-treat > Treatment-on-the-treated estimand)
  - Publication bias (e.g., p-hacking)
- But even theoretically it depends on context:
  - Direction of omitted variable bias (confounding) depends on the relationship between omitted variable and probability of treatment
  - Bias due to participant motivation in unblinded trials may increase treatment effect magnitude (e.g. Hawthorne effects) or reduce it (e.g. compensatory rivalry in controls)
  - Bias due to reported outcomes may increase treatment effect magnitude (e.g. social desirability bias) or reduce it (e.g. participant fatigue in repeated panel surveys)
  - Site-selection effects in RCTs acts to increase effect size magnitude

REPLICATION STUDIES CAN HELP VALIDATE RISK-OF-BIAS (ROB) ASSESSMENT TOOLS

Source of bias (bias domain)	Examples on which signalling questions are based
Confounding	Programme placement bias and self-selection into treatment
Selection bias	Differential exclusion of eligible treatment units and follow-ups
Departures from intended interventions	Performance bias (e.g. spillovers), motivation bias (e.g. Hawthorne effects), implementation fidelity
Measurement error	Errors in measuring exposures or in defining and reporting outcomes
Selection of the reported result	Outcomes, sub-groups or methods of analysis
Overall 'risk of bias'	'Low risk', 'some concerns', 'high risk'

Source: Hombrados and Waddington (2012); Sterne et al. (2016); Waddington (2021)

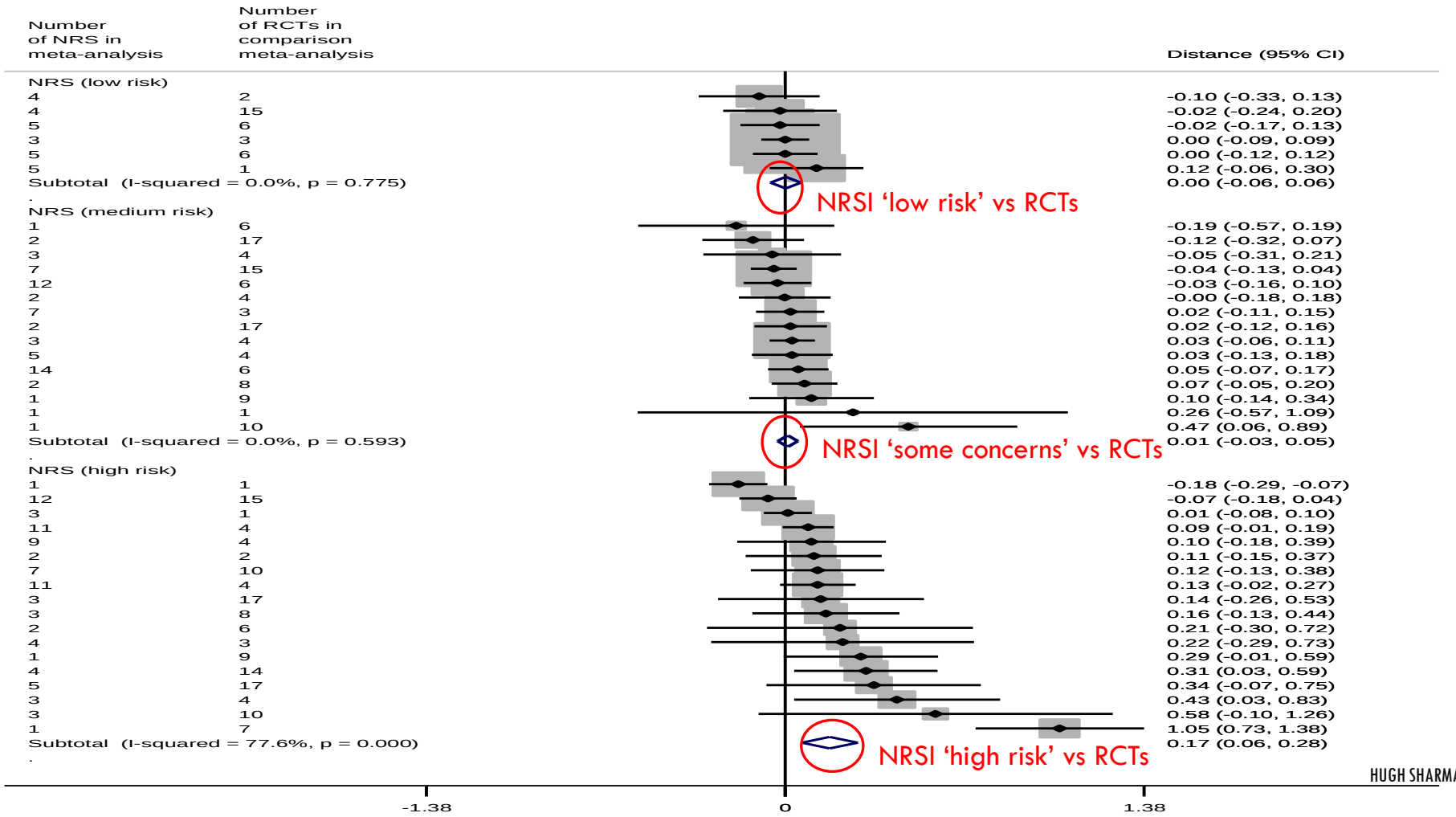
SYSTEMATIC REVIEWS THAT USED THIS ROB TOOL

Author	*Sector*	*Outcomes*	*# RCTs*	*# NRS*
Baird et al. (2013)	Education	School attendance	15	27
Brody et al. (2015)	Micro-finance	Women's empowerment	5	18
Carr-Hill et al. (2016)	Education	Drop-outs, test scores	9	17
Chinen et al. (2017)	Vocational training	Employment, earnings	26	9
Hemming et al. (2018)	Agriculture	Adoption, yield, income	2	13
Lawry et al. (2014)	Agriculture	Agricultural income	0	20
Molina et al. (2016)	Governance	Health outcomes	10	5
Oya et al. (2017)	Agriculture	Income, wages, schooling	0	43
Piza et al. (2016)	Vocational training and finance	Firm performance, employmen	6	23
Samii et al. (2014a)	Climate change	Environment, poverty	0	11
Samii et al. (2014b)	Climate change	Environment, poverty	0	8
Stone et al. (2019)	Education	Literacy	9	7
Ton et al. (2017)	Agriculture	Agricultural yield	0	22
Tripney et al. (2013)	Vocational training	Employment, income	3	23
Vaessen et al. (2014)	Micro-finance	Women's empowerment	4	21
Waddington et al. (2014)	Agriculture	Knowledge, adoption, yields, income	0	93
Waddington et al. (2019)	Governance	Community engagement, service access, service use	19	16

Note: All published in *Campbell Systematic Reviews*

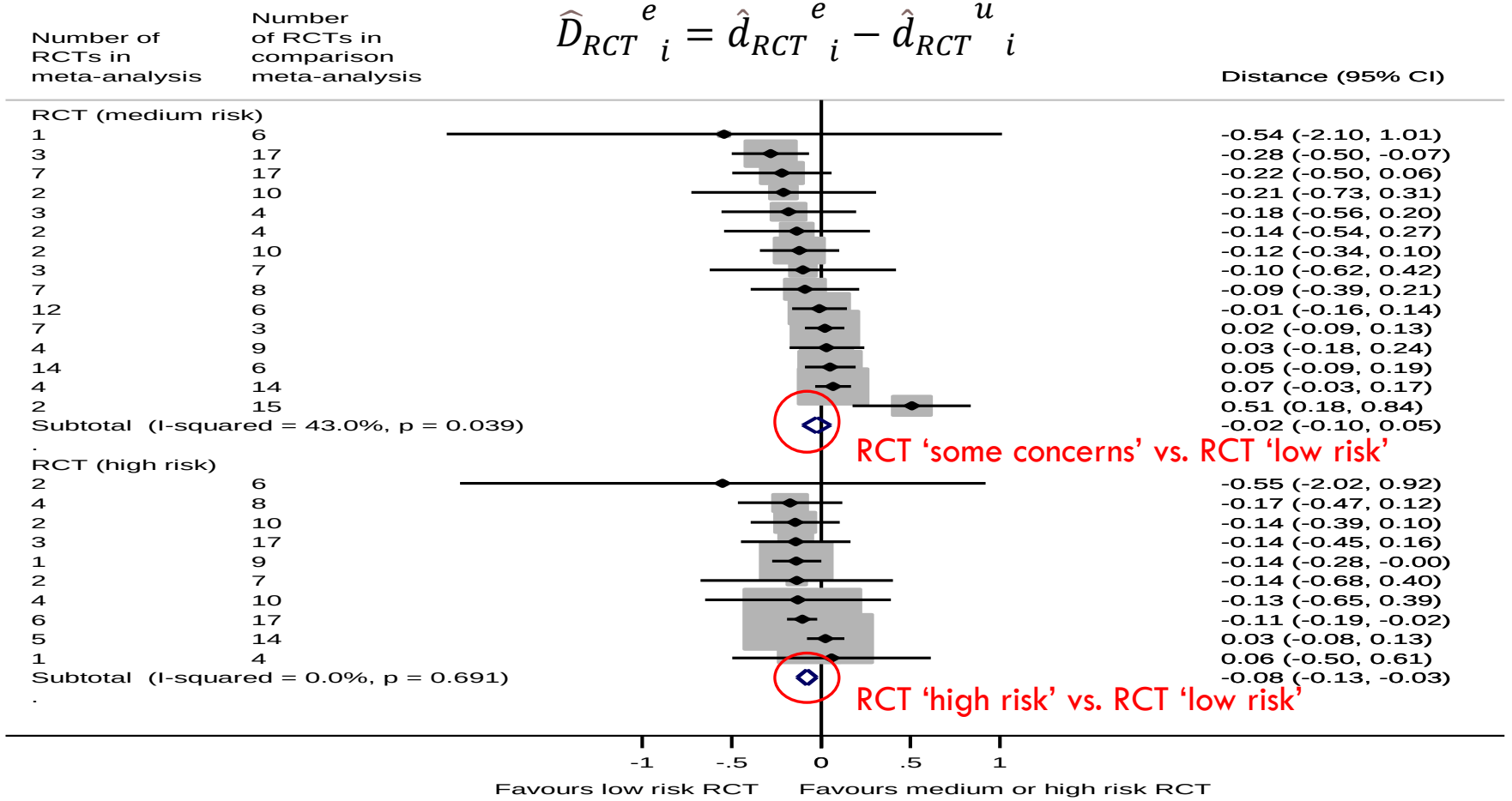
‘HIGH RISK’ NRSI OVERESTIMATED POOLED EFFECTS

$$\widehat{D}_{NRS_i} = \hat{d}_{NRS\ i} - \hat{d}_{RCT\ i} \qquad se(\widehat{D})_i = \sqrt{s^2_{NRS_i} + s^2_{RCT\ i}}$$



BUT 'HIGH RISK' RCTS MAY UNDERESTIMATE EFFECTS

$$\hat{D}_{RCT}{}^{e}{}_{i} = \hat{d}_{RCT}{}^{e}{}_{i} - \hat{d}_{RCT}{}^{u}{}_{i}$$

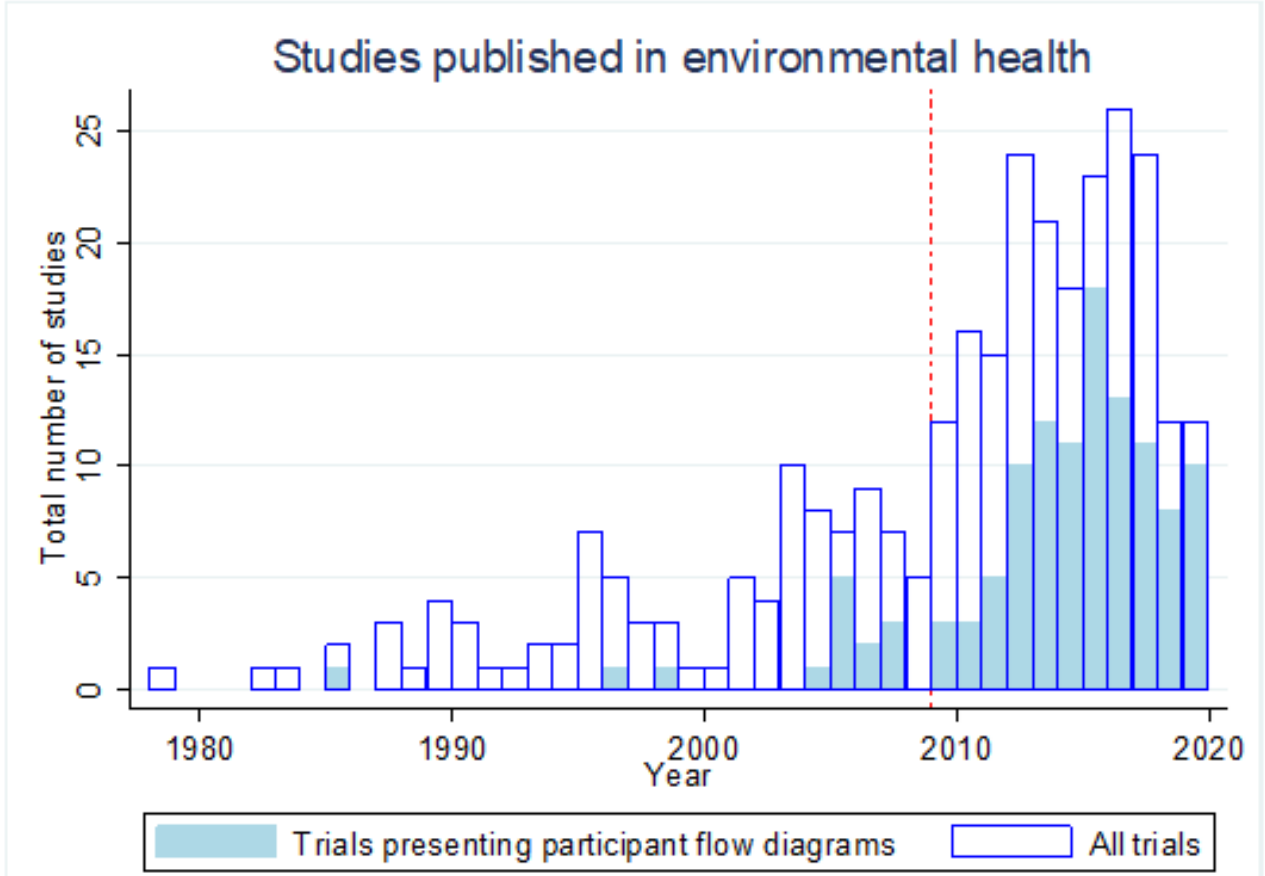


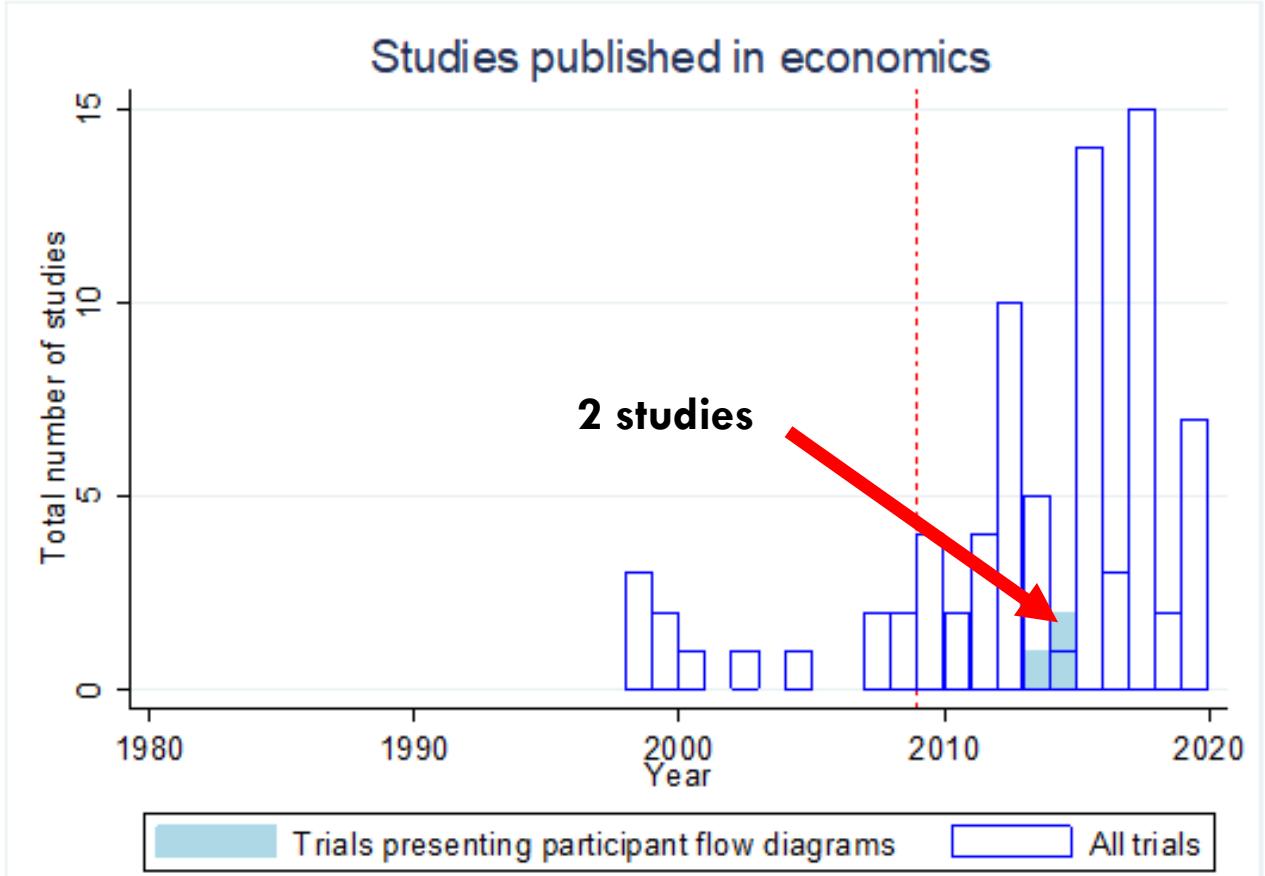
SUMMARY OF RANDOM EFFECTS META-ANALYSES

Comparison	D	95% confidence interval		I^2	Tau^2	N
NRS – RCT	**0.045**	**0.010**	**0.080**	68%	0.004	28
NRS (low risk) – RCT	0.002	-0.056	0.060	0%	0.000	6
NRS (medium risk) – RCT	0.010	-0.027	0.048	0%	0.000	15
NRS (high risk) – RCT	**0.171**	**0.065**	**0.278**	78%	0.033	18
RCT (medium) – RCT (low risk)	-0.024	-0.102	0.053	43%	0.008	15
RCT (high risk) – RCT (low risk)	**-0.080**	**-0.135**	**-0.026**	0%	0.000	10
NRS (low risk) – RCT (low risk)	-0.001	-0.044	0.042	0%	0.000	4
NRS (medium) – RCT (low risk)	-0.013	-0.060	0.034	0%	0.000	12
NRS (high risk) – RCT (low risk)	**0.130**	**0.008**	**0.253**	53%	0.021	13

Source: Waddington 2021 LSHTM

ON THE IMPORTANCE OF REPORTING PARTICIPANT FLOWS (CONSORT)



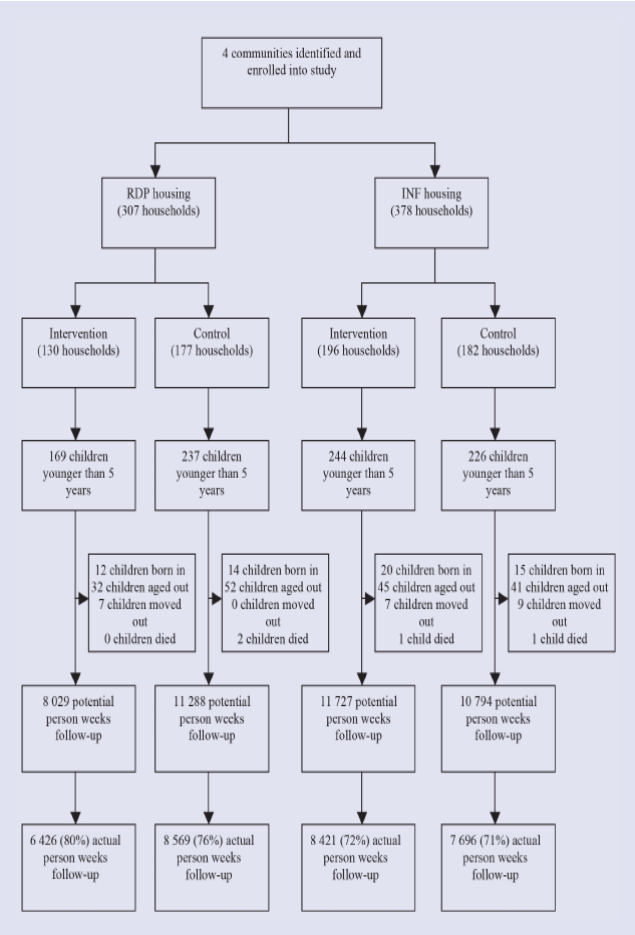


Source: data from Chirgwin et al 2021 Campbell Sys Rev

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

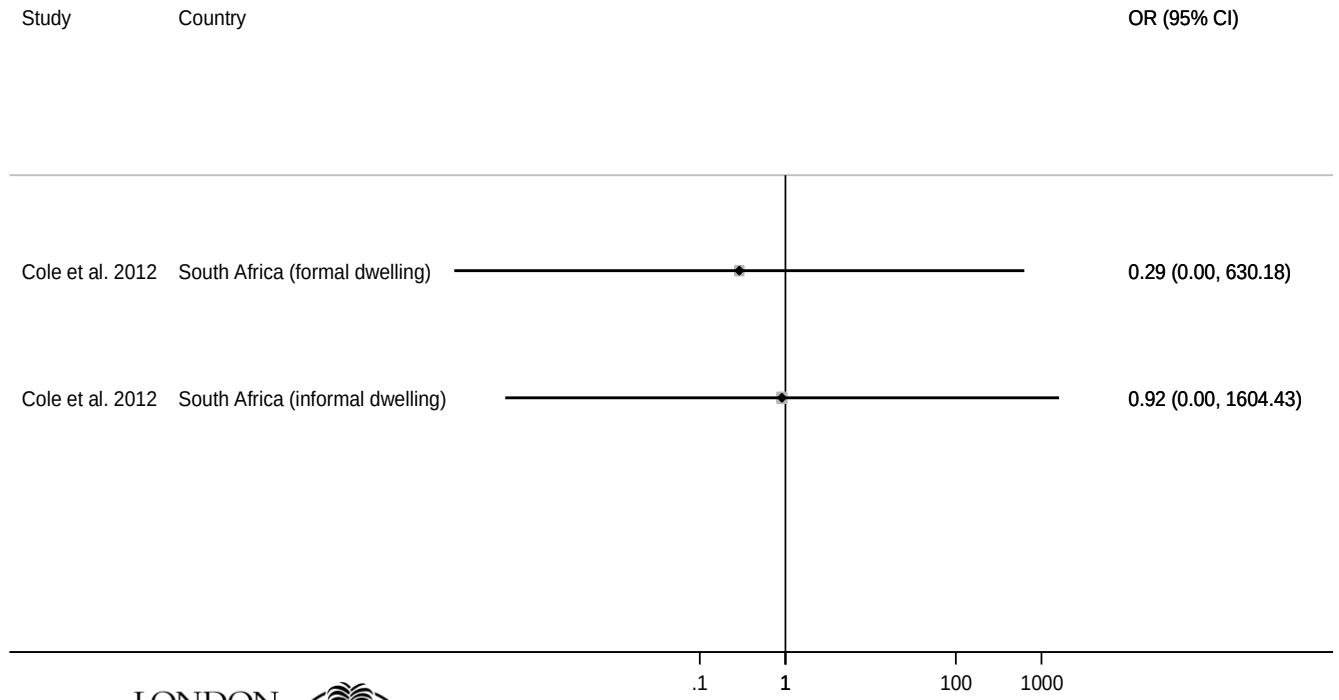
MEASUREMENT OF ALL-CAUSE MORTALITY IN TRIALS IS RELIANT ON REPORTING OF PARTICIPANT FLOWS

CONSORT (Moher et al., 1998)



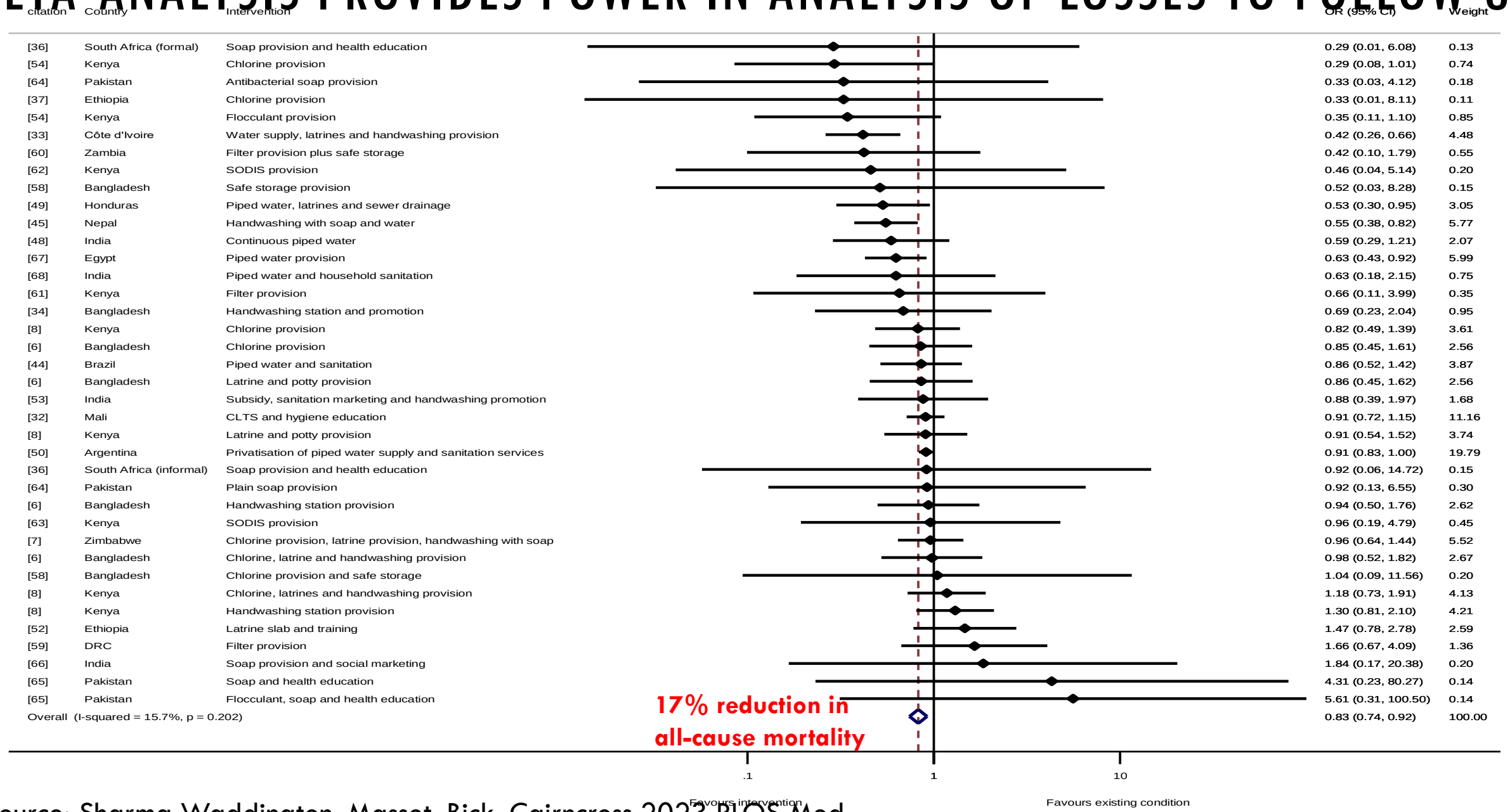
Source: Cole et al. (2012)

Forest plot of all-cause mortality from trial arms in Cole et al. (2012)



LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

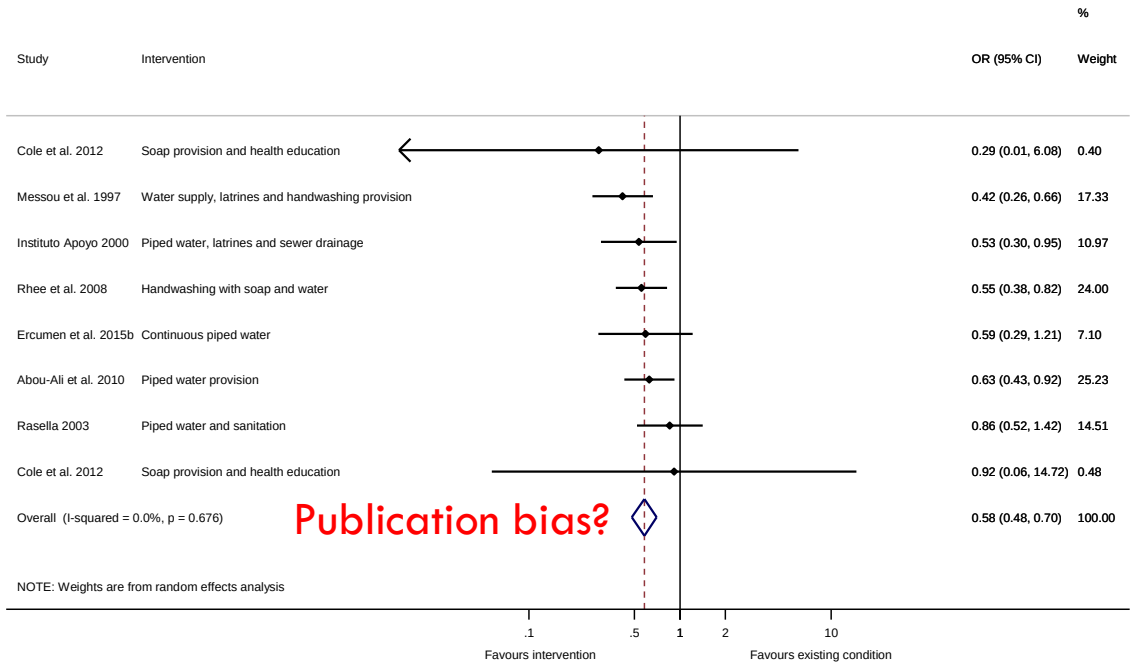
META-ANALYSIS PROVIDES POWER IN ANALYSIS OF LOSSES TO FOLLOW UP



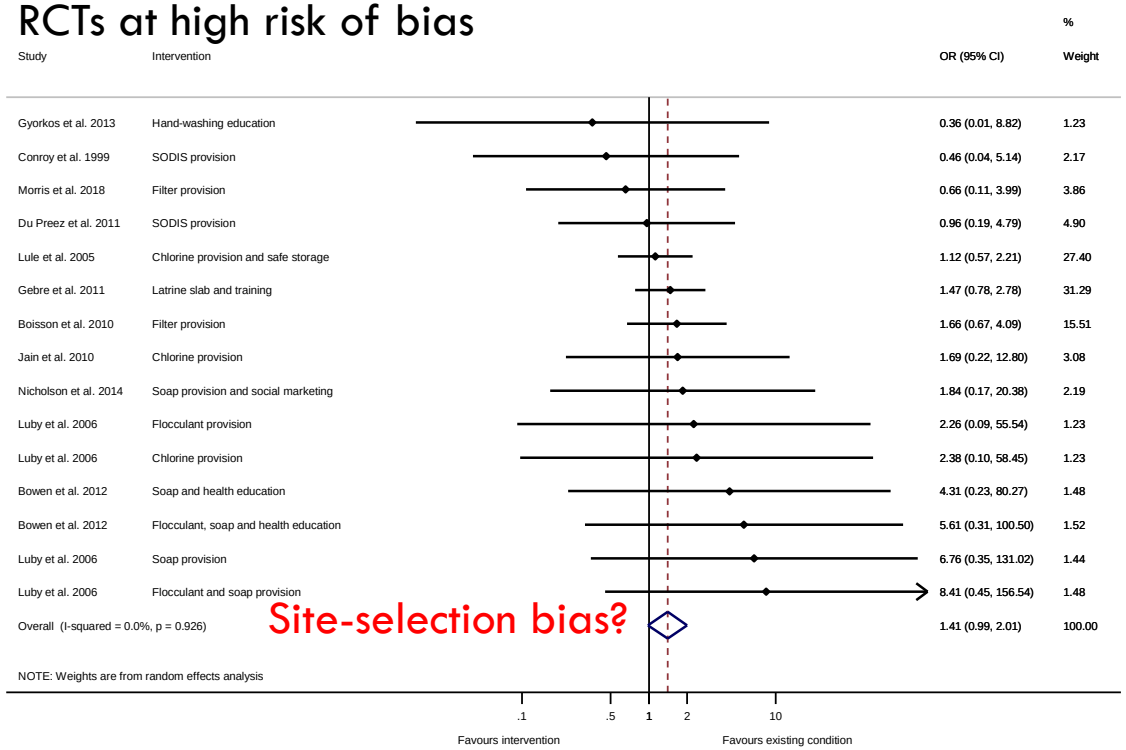
Source: Sharma Waddington, Masset, Bick, Cairncross 2023 PLOS Med.

DICHOTOMOUS RELATIONSHIP BETWEEN ROB AND EFFECT SIZES OBSERVED IN META-ANALYSIS OF PRIMARY STUDIES

NRSI at high risk of bias



RCTs at high risk of bias



Source: Sharma Waddington, Masset, Bick, Cairncross 2023 PLOS Med.

CONCLUSIONS

Study design is probably the most important factor in determining bias (Cook, Shadish and Wong, 2008)

Site-selection effects (intervention design and implementation fidelity) may explain dichotomy in relationship between RoB and effect sizes for RCTs and NRSI

Health data science can usefully conduct internal replication studies and reviews of these studies (e.g. ITS) to empirically validate risk-of-bias tools

Standards for reporting need to be urgently improved in social science studies

We plan a collaborative project on reporting standards in NRSI/QEDs - please contact me if interested

THANKS

Hugh.Waddington@lshtm.ac.uk