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:

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

Electronic repositories, bibliographic searches author tracking (n=3,271)contacts (n=951)(n=320) Records after duplicates removed (n=3,904)Excluded Titles and abstracts screened (n=3,328)(Stage 1) Full-text articles screened (Stage 1) Excluded (n=576)(n=443)Records assessed for eligibility Excluded (Stage 2) (n=133) (n = 125)Included L&MIC primary studies in meta-analysis (n=8): 600 effect size estimates

Source: Sharma Waddington, Villar and Valentine 2022 Evaluation Review

RESULTS FROM FIXED EFFECT META-ANALYSIS

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

Estimator	Standardised bias	Mean squared	Percent of bias	Num. of estimates
		error	removed	
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

HUGH SHARMA WADDINGTON 5

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 HUGH SHARMA WADDINGTON 7

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

Ĺ	$\hat{D}_{NRS_i} = \hat{d}_{NRS_i} - \hat{d}_{RCT_i}$	$se(\widehat{D})_i = \sqrt{s_{NRS_i}^2 + s_{RCT_i}^2}$
Number	Number of RCTs in	
of NRS in meta-analysis	comparison meta-analysis	Distance (95% CI)
NRS (low risk)		
4	2	-0.10 (-0.33, 0.13)
4	15	-0.02 (-0.24, 0.20)
5	6	-0.02 (-0.17, 0.13)
3	3	0.00 (-0.09, 0.09)
5	6	0.00 (-0.12, 0.12)
5		0.12 (-0.06, 0.30)
Subtotal (I-squar	ed = 0.0%, p = 0.775)	
NRS (medium ris	k)	
1	6	-0.19 (-0.57, 0.19)
2	17	-0.12 (-0.32, 0.07)
3	4	-0.05 (-0.31, 0.21)
7	15	-0.04 (-0.13, 0.04)
12	6	-0.03 (-0.16, 0.10)
2	4	
7	3	
2	17	
5	4	
14		
2	8	
- 1	9	
1	1	0.26 (-0.57, 1.09)
1	10	0.47 (0.06, 0.89)
Subtotal (I-squar	red = 0.0%, p = 0.593)	(Φ) NRSI (some concerns' vs RCTs 0.01 (-0.03, 0.05)
NRS (high risk)		
1	1	-0.18 (-0.29, -0.07)
12	15	-0.07 (-0.18, 0.04)
3	1	0.01 (-0.08, 0.10)
11	4	0.09 (-0.01, 0.19)
9	4	
∠ 7	∠ 10	
/ 11	4	
3	-+ 17	
3	8	
2	- 6	0.21 (-0.30, 0.72)
4	3	0.22 (-0.29, 0.73)
1	9	0.29 (-0.01, 0.59)
4	14	0.31 (0.03, 0.59)
5	17	0.34 (-0.07, 0.75)
з	4	0.43 (0.03, 0.83)
3	10	0.58 (-0.10, 1.26)
1	7	1.05 (0.73, 1.38)
Subtotal (I-squar	red = 77.6%, p = 0.000)	NRSI 'high risk' vs RCTs 0.17 (0.06, 0.28)
-		

ο

-1.38

HUGH SHARMA WADDINGTON

Т

1.38

10

BUT 'HIGH RISK' RCTS MAY UNDERESTIMATE EFFECTS



SUMMARY OF RANDOM EFFECTS META-ANALYSES

Comparison	D	95% confidence interval		 2	Tαυ²	Ν
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)



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

MEDICINE

CONSORT (Moher et al., 1998)



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



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

[36]	South Africa (formal)	Soap provision and health education		0.29 (0.01, 6.08)	0.13
[54]	Kenya	Chlorine provision	_	0.29 (0.08, 1.01)	0.74
[64]	Pakistan	Antibacterial soap provision	_	0.33 (0.03, 4.12)	0.18
[37]	Ethiopia	Chlorine provision	• <u> </u>	0.33 (0.01, 8.11)	0.11
[54]	Kenya	Flocculant provision		0.35 (0.11, 1.10)	0.85
[33]	Côte d'Ivoire	Water supply, latrines and handwashing provision	i	0.42 (0.26, 0.66)	4.48
[60]	Zambia	Filter provision plus safe storage		0.42 (0.10, 1.79)	0.55
[62]	Kenya	SODIS provision	_	0.46 (0.04, 5.14)	0.20
[58]	Bangladesh	Safe storage provision		0.52 (0.03, 8.28)	0.15
[49]	Honduras	Piped water, latrines and sewer drainage	_	- 0.53 (0.30, 0.95)	3.05
[45]	Nepal	Handwashing with soap and water		0.55 (0.38, 0.82)	5.77
[48]	India	Continuous piped water		0.59 (0.29, 1.21)	2.07
[67]	Egypt	Piped water provision		0.63 (0.43, 0.92)	5.99
[68]	India	Piped water and household sanitation	_	0.63 (0.18, 2.15)	0.75
[61]	Kenya	Filter provision		0.66 (0.11, 3.99)	0.35
[34]	Bangladesh	Handwashing station and promotion		0.69 (0.23, 2.04)	0.95
[8]	Kenya	Chlorine provision		0.82 (0.49, 1.39)	3.61
[6]	Bangladesh	Chlorine provision		0.85 (0.45, 1.61)	2.56
[44]	Brazil	Piped water and sanitation		0.86 (0.52, 1.42)	3.87
[6]	Bangladesh	Latrine and potty provision		0.86 (0.45, 1.62)	2.56
[53]	India	Subsidy, sanitation marketing and handwashing promotion		0.88 (0.39, 1.97)	1.68
[32]	Mali	CLTS and hygiene education		0.91 (0.72, 1.15)	11.16
[8]	Kenya	Latrine and potty provision		0.91 (0.54, 1.52)	3.74
[50]	Argentina	Privatisation of piped water supply and sanitation services	<u>k</u>	• 0.91 (0.83, 1.00)	19.79
[36]	South Africa (informal)	Soap provision and health education		0.92 (0.06, 14.72) 0.15
[64]	Pakistan	Plain soap provision		0.92 (0.13, 6.55)	0.30
[6]	Bangladesh	Handwashing station provision	ł,	0.94 (0.50, 1.76)	2.62
[63]	Kenya	SODIS provision		0.96 (0.19, 4.79)	0.45
[7]	Zimbabwe	Chlorine provision, latrine provision, handwashing with soap	÷	0.96 (0.64, 1.44)	5.52
[6]	Bangladesh	Chlorine, latrine and handwashing provision		0.98 (0.52, 1.82)	2.67
[58]	Bangladesh	Chlorine provision and safe storage		1.04 (0.09, 11.56) 0.20
[8]	Kenya	Chlorine, latrines and handwashing provision	÷	1.18 (0.73, 1.91)	4.13
[8]	Kenya	Handwashing station provision	+	1.30 (0.81, 2.10)	4.21
[52]	Ethiopia	Latrine slab and training	<u>+</u>	1.47 (0.78, 2.78)	2.59
[59]	DRC	Filter provision	+	1.66 (0.67, 4.09)	1.36
[66]	India	Soap provision and social marketing		1.84 (0.17, 20.38) 0.20
[65]	Pakistan	Soap and health education	+	4.31 (0.23, 80.27) 0.14
[65]	Pakistan	Flocculant, soap and health education	17% reduction in	\$ 5.61 (0.31, 100.5	0) 0.14
Overall	(I-squared = 15.7%, p = 0	.202)		0.83 (0.74, 0.92)	100.00
			all-cause mortality		
			.1	1 10	

Source: Sharma Waddington, Masset, Bick, Cairncross 2025 PLOS Med.

Favours existing condition

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



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

HUGH SHARMA WADDINGTON