Institute of Education

EPPI CENTRE

UCL

Enhancing the portability of public health intervention review evidence for localised decision-making

a.k.a. It's a mean world for generalisability

Dylan Kneale, EPPI-Centre, UCL Institute of Education

ESMI Guest Lecture Series, University of Exeter

Generalisability in public health systematic reviews and **meta-analyses**

- How can we define it
- Why it matters
- How do systematic reviewers (meta-analysts) ‘treat’ it?
- What options are currently available/used?
- What could we do differently

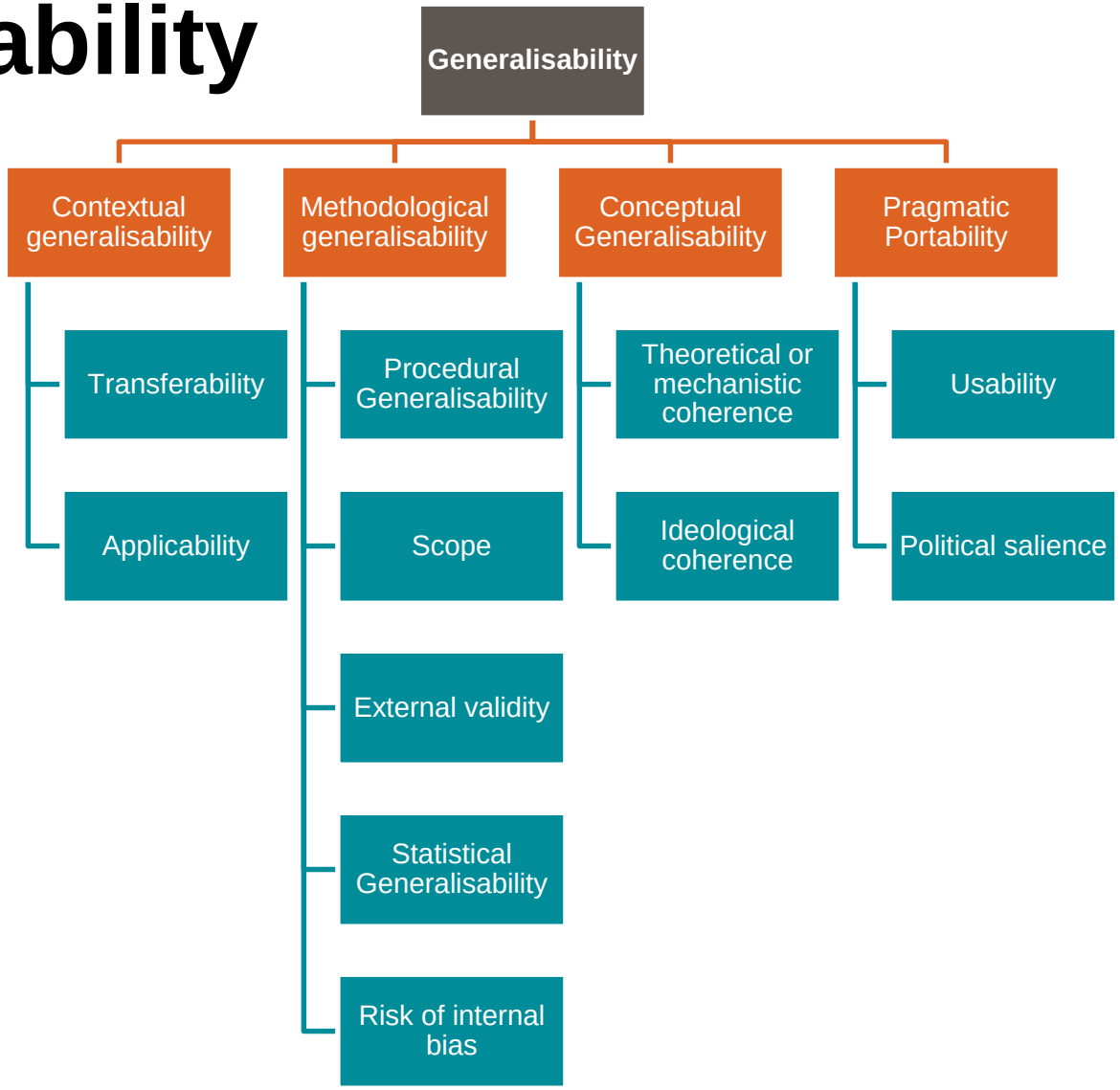
Part 1: How can we define generalisability?

Defining generalisability

- *'extent to which the results of a study conducted in a particular patient population and/or a specific context will apply for another population and/or in a different context'* (NICE 2013)
- Domains of generalisability (Wang et al 2006; Burchett et al 2011)
- Big distinction – applicability vs transferability
- (Even) more to it?

Burchett H, Umoquit M, Dobrow M: **How do we know when research from one setting can be useful in another? A review of external validity, applicability and transferability frameworks**. *Journal of health services research & policy* 2011, **16**(4):238-244.
NICE: **Guide to the methods of technology appraisal**. In. London: National Institute for Health and Care Excellence; 2013.
Wang S, Moss JR, Hiller JE: **Applicability and transferability of interventions in evidence-based public health**. *Health promotion international* 2006, **21**(1):76-83

Broad understandings of generalisability



Defining generalisability ii

- Viewing generalisability as a series of questions to be asked
- Are we all asking the same questions?
- Questions are of differential interest to different stakeholders
  - Contextual generalisability – Evidence users
  - Methodological generalisability – Evidence generators
- Ahmad (2010) – country of study (46%); applicability/generalisability (13%)
- **Terminology matters**
- **Contextual generalisability = low priority**

AHMAD, N., BOUTRON, I., DECHARTRES, A., DURIEUX, P. & RAVAUD, P. 2010. Applicability and generalisability of the results of systematic reviews to public health practice and policy: a systematic review. *Trials,* 11**,** 20

Part 2:
Why does generalisability of review evidence matter?

Why does generalisability of review evidence matter?

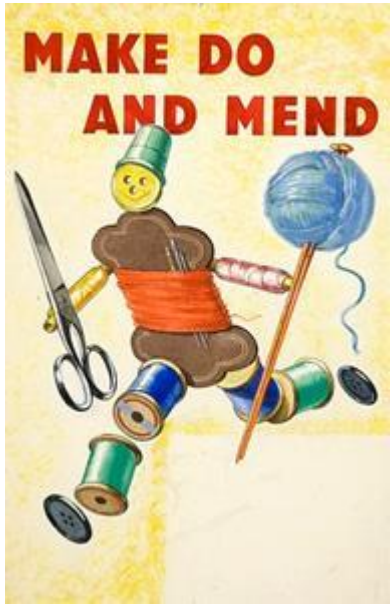
- **Findings from systematic review of evidence use in public health**
  - Emphasis on heterogeneity among LA
  - Public health decision-making more politicised
  - Experts (who?) and local evaluations (quality?) are an important part of the mix
  - Decision-makers trade quality for (local) relevance

Kneale, D., Rojas-Garcia, A., Raine, R., Thomas, J. (2017). The use of evidence in English local public health decision-making: a systematic scoping review. *Implementation Science*, 12 (53), doi:10.1186/s13012-017-0577-9

Institute of Education

Findings from documentary analysis of evidence use in public health

- Evidence use culture is flourishing BUT:
- (Academic) Research evidence use – limited and confined to trusted sources
- LAs/HWBs conducting own research
- Little appetite for systematic reviews
- Areas with the highest needs and lowest budgets are less likely to be engaging with research evidence



Kneale, D., Rojas-Garcia, A., Thomas, J. (in press). The use of evidence in English local public health decision-making: a systematic scoping review. *Journal of Public Health*, doi: 10.1093/pubmed/fdx152

Findings from qualitative analysis of evidence use in public health

- Greater role for evidence in the commissioning process?
- Timescales – no ongoing support from evidence producers. Model of Publish and Run in evidence generation!
- Hard to develop ongoing relationships
  - Funding and time constraints
- Decision-makers challenged by lack of:
  - Information about small groups
  - Information about small areas
  - Support in applying evidence across diverse areas

“I mean it's partly also because a lot of times we will look at systematic reviews and they will say there really isn't evidence on ***how effective these intervention models are in contexts*** *and in deprived contexts or among ethnic minority groups or others.*

…And then they say that there should be more research done on how this might work in more deprived communities and then you think great...

But then this there isn't this research done or if there isn't more systematic reviews of that type of research i.e. in deprived areas. Then we have little alternative but to go to a smaller scale. Not that we ever necessarily categorise it as less rigorous but it just means that we're not able to vouch with confidence about how things will work in this type of setting.”

Why does generalisability of review evidence matter?

- Low levels of systematic review usage
- Strong preference for localised estimates
- Strong preference for quantitative data (concerning!?!)
- Valued but low level of use of evidence on effectiveness
- Impact on decision-making?
- **What is current practice and what tools do we have at our disposal?**

Part 3: How do systematic reviewers (meta-analysts) ‘treat’ generalisability?

Generalisability claims of meta-analysis

- Donaldson (2001) [through synthesizing] *'different participants in different situations, and using different research procedures, one is able to get a better estimate of the robustness or the external validity of a given finding or effect'*
- Booth, Rees et al. 2015 p336 *'the inclusion of multiple studies' is described as increasing 'the generalisability of findings and prompts the practitioner to examine conflicting results'*

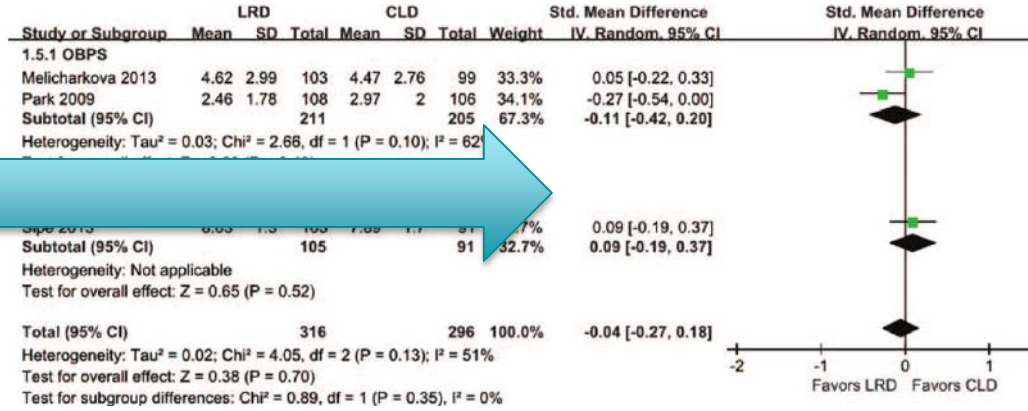
Donaldson SI, Street G, Sussman S, Tobler N: **Using meta-analyses to improve the design of interventions**. In: *Handbook of program development for health behavior research and practice.* edn. Edited by Sussman S. Thousand Oaks, California: Sage; 2001: 449-466.
Booth A, Rees A, Beecroft C: **The research process in nursing**. In., edn. Edited by Gerrish K, Lacey A. London: John Wiley & Sons; 2015.

Not atypical!

Not naïve – don't claim generalisability is fixed but…

- Unpacking them:
  - Terminology
  - Aggregative logic - more studies = more generalizable?
    - PICO represented?
  - A sample of activity, and a sample of effect sizes
    - Not a census – is sample representative?
  - Generalisable to who, where, when?
- **Potential for misinterpretation**
  - Large number of studies, low heterogeneity = generalisable to settings falling in 'P'ICOS??

Two main approaches available (to the 13%!)



The problem with tools and checklists (i)

- We often don't use them when they are available
  - Ahmad (2010)
- We don't know which to use
  - Burchett et al (2011) 25+; 2017 +++
- We don't agree on how to use; integrated with other QA

The problem with tools and checklists (ii)

- Unwieldy – use for ‘configurative’ meta-analysis?
- Lack theory
- Look for difference rather than ‘generalisability’
- Flourishing industry

Using subgroup analyses and meta-regression to explore generalisability

- Several issues exist and identified (e.g. Petticrew; Oxman & Guyatt)
- Additional issues for understanding generalisability:
- **1. Terminology**
- **2. Poor/misspecification**
- **3. Confounding and 'collinearity'**
- **4. Lack theory**

Oxman, A. D. and G. H. Guyatt (1992). "A consumer's guide to subgroup analyses." Ann Intern Med **116**(1): 78-84.
Petticrew, M., P. Tugwell, E. Kristjansson, S. Oliver, E. Ueffing and V. Welch (2011). "Damned if you do, damned if you don't: subgroup analysis and equity." Journal of epidemiology and community health

- **1. Same terminology – Sub-group** analysis/Meta-regression is used to describe:
  - A. Study level characteristics
  - B. Participant level characteristics (rarer; IPD)
    - "Stratification"?

- **Issues in interpretation:**
  - **A interpreted as B**

- Meta-analysis of interventions to increase physical activity among adults
- *"Participant characteristics [including gender] were unrelated to physical activity effect sizes"* (Conn, Hafdahl et al. 2011)
- Meta-regression model included % of women
- Not incorrect in its conclusion BUT
- Does this tell that PA interventions work equally as well for women as men? Stratified approach?
- Potential for ecological fallacy?; Helpful for generalisability?

- **2. Misspecification and misinterpretation**
- Collaborative care interventions to improve the management of depressive disorders (Thota, Sipe et al. 2012)
- Establishing the ‘applicability or generalizability of the intervention to a comprehensive range of populations and settings’ was a key aim
- Realised in quantitative terms through the conduct of several subgroup analysis
- Grouped trials - majority female or male

Table 4. Subgroup analyses from studies that reported depression symptom improvement

Variable	Category	Study arms, *n*	Stratified estimate (Hedges' *g*)
Age	Adult	19	0.31
	Older adult	6	0.46
	Teen	1	0.25
Gender	Majority female	22	0.36
	Majority male	4	0.24
Race/ethnicity	Majority white	8	0.3
	Majority African-American	1	0.56
	Majority Latino	2	0.26
SES*	Low	1	0.27
	Majority low	1	1.02
	Majority mid-high	3	0.09
Mental health specialist	Psychiatrist and/or psychologist and others	15	0.34
	Psychiatrist/psychologist	10	0.36
	MD and/or RN	2	0.25

Note: Data were not available for all variables for all studies.
*$p<0.05$
CBO, community-based organization; RN, registered nurse; VA, Veterans Affairs

- **3. A form of 'collinearity'**
- Confounding established risk of SGA (e.g. Sun 2010)
- Also a risk of 'collinearity'
- Same trials falling in one set of SGA and another
- Treated independently in interpretation
  - Age
  - Mode of intervention
  - Not only confounded but 'collinear'

Sun, X., M. Briel, S. D. Walter and G. H. Guyatt (2010). "Is a subgroup effect believable? Updating criteria to evaluate the credibility of subgroup analyses." Bmj **340**: c117

- **4. Lack theory**
- [Not new but unchanged]
- Reviewers are still tempted:
- *Below, we pre-specify a list of a priori moderator variables assumed to moderate the true effect of the interventions: Geographical region; Place of residence; Ethnicity; Occupation; Religion; Education; Social capital; Socio-economic position; Age; Disability; Sexual orientation; Young vs. older; Non-poor vs. poor vs. very poor; Rural vs. urban; Collectives vs. individuals*
- Greater likelihood of T1 error but also
- **Perfunctory subgroup analysis**

- **Disaggregative analysis**
- **Checklists and tools not always informative**
- **Analytical options available are used in perfunctory way**

Need to respond to the need for greater attention to contextual generalisability

- A. Improve on the use the methods we already have
- B. Develop new methods

Using the options available more effectively i

- Tools and checklists…

Improving on current disaggregative analyses:

1. Ground in theory
2. Distinguish between study-level and participant-level characteristics (‘stratification’)
3. Report and interpret level of heterogeneity within subgroups

Using the options available more effectively ii

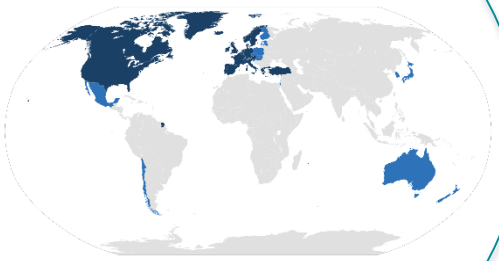
4. Potential issues of confounding and collinearity are explicitly considered
5. Operationalised through covariates and indicators that make conceptual sense
6. Discuss extent to which the sample is representative of the scope of the PICO criteria

- BUT…is there more?

Part 4: New approaches?

Systematic Reviews and Public Health Decision-Making

Systematic Reviewers (Meta-analysts)



PH Decision-makers in England

Solutions proposed

a. ~~Do nothing – synthesising evidence from diverse contexts enhances generalisability in itself~~ ☒☒☒

b. Assess generalisability

   Yes…but no agreed framework that reflects decision-maker input ☒☑

c. Use existing meta-analytic approaches

   Yes…but explain study-level heterogeneity (limited for generalisability) ☒☑

d. Use other resources to hand???

   Sources of secondary data from the settings to which we want to generalise

A missing piece of the puzzle?

- UK large longitudinal panel, cohort and cross-sectional surveys + burgeoning real-world data landscape
- Epidemiological patterns, patterns of 'usual care' – factors reflecting applicability and transferability
- Context rich
- Potentially provide natural controls
- Epistemological differences: are these insurmountable?

	Meta-analyses	Large secondary datasets
Availability	Dependent on publication of study result	Dependent on deposit of secondary data for re-analyses
Research focus	Narrow analytic intent/research question	Usually multipurpose with user-defined analytic intent
Purpose	Often based on studies aiming to establish causal inference	Often used for studies aiming to classify/predict

Solutions proposed

If survey data can be used with meta-analysis, then what are some of the questions that could be asked?

Trialling one or a combination of three main approaches:

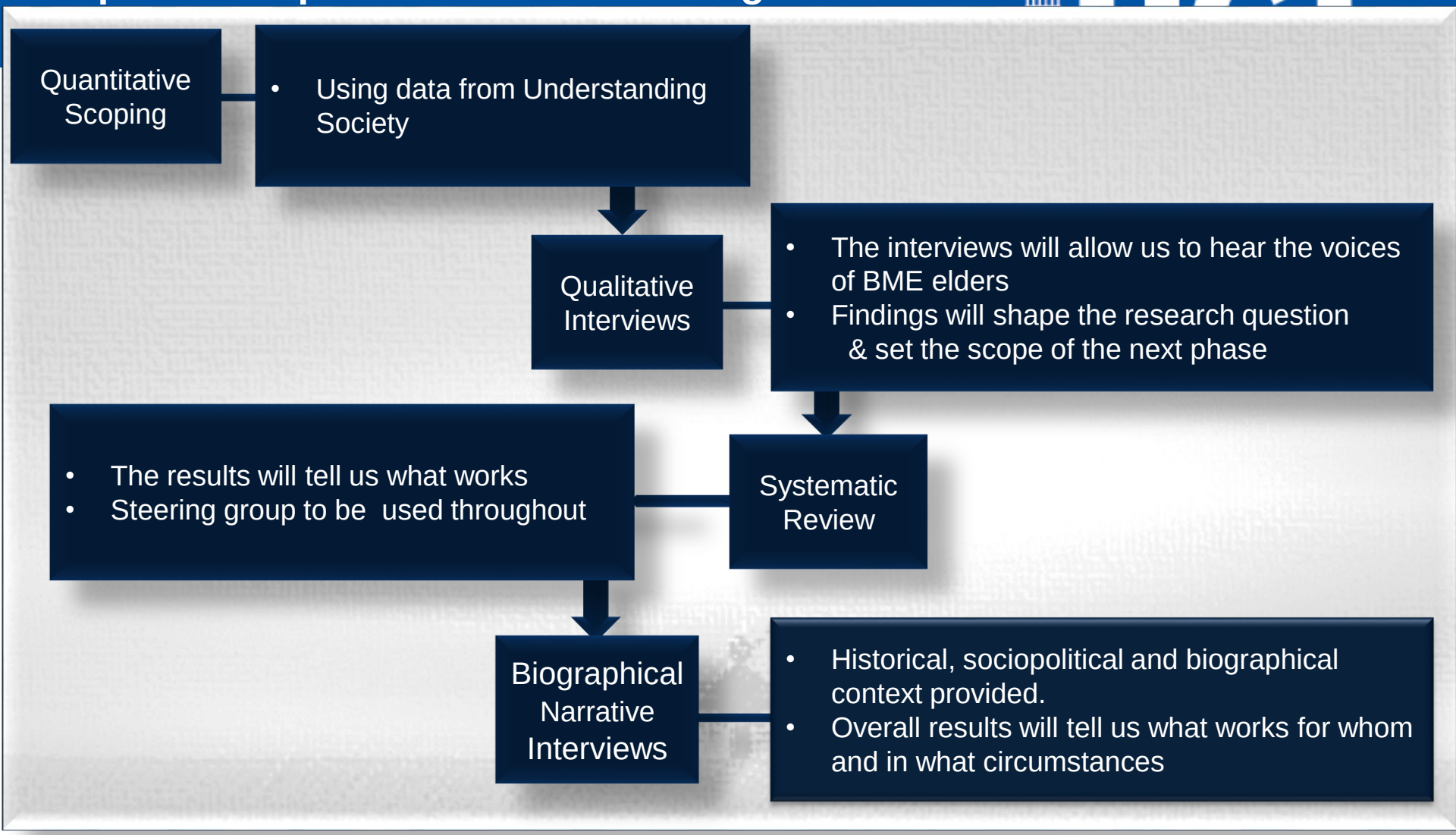
1. Purposeful exploration before starting a review: using the results of preliminary SDA to guide SR/MA

2. Purposeful exploration after conducting a review: using SR/MA to structure SDA to explore *applicability* of findings

3. Purposeful exploration during a review: using SDA to enhance generalisability of estimates

Institute of Education

Purposeful exploration before starting a SR



Purposeful exploration after SR

- Katz et al. (2008) - obesity in school settings
- 8 different countries; SMD 0.29 (CI: -0.45 to -0.14) in children's body weight.
- Parental involvement (either alongside children or separately) SMD -0.20 (CI: -0.41 to 0.00) in children's body weight.
- Hypothetical decision-maker – London GLA
- Ensure family-based, multi-component lifestyle weight management services for children and young people are available
- Is involving parents viable in London?

Purposeful exploration after SR

- Use Millennium Cohort Study to answer two questions:
  - (i) are parents of overweight/obese children more/less likely to attend after school activities?
  - (ii) are overweight/obese children more/less likely to attend after school activities?
- Parents not attending parents' evening 2* as high among parents of obese in London (but still low)
- Parents who regularly work evenings lower; no relationship with obesity
- Among children in London, just over a quarter of children aged 7 (26%) were attending breakfast or after-school clubs.
- Also ¼ in London (23%) reported getting tired during the school day.
- After-school not feasible for children; but could deliver to parents?

Purposeful exploration during a review

- Moving from applicability to transferability
- Limitations of SGA and M-R is that they are both usually extensions of bivariate analyses
- Areas rural/urban, diverse/homogenous, deprived/advantaged simultaneously
- Look matrix of difference and ways of developing scores
- Incremental extension
  - A form of enhanced SGA
  - Recalibration

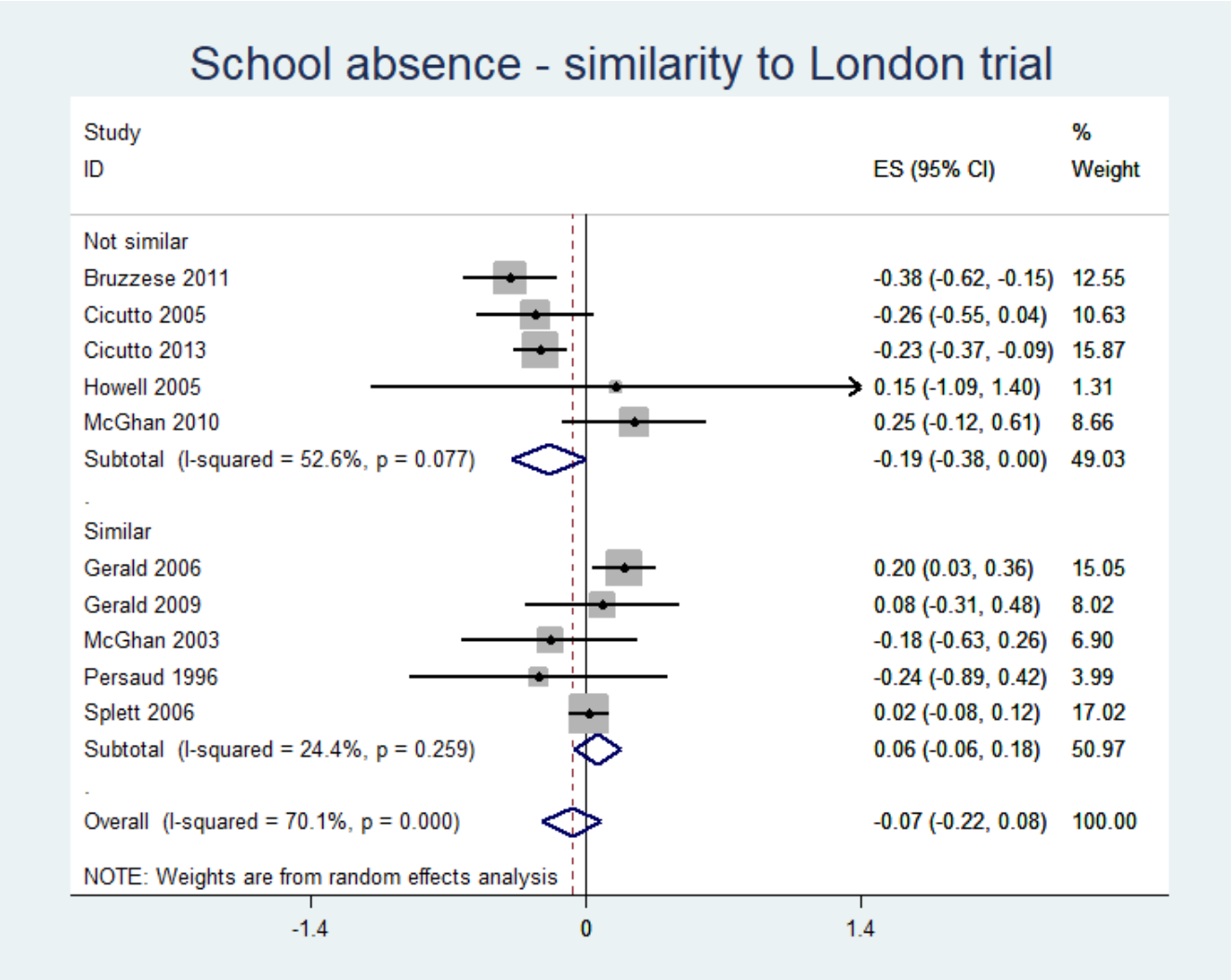
Example of Cochrane Asthma Review

- School-based asthma interventions to improve self-management
- Cochrane review to inform design of a trial
- Applicability questions
- Extend to transferability question

Common Data

study	es	se	bme	School absence	male
Bruzzese 2011	-0.38207	0.120627	95%	26.0	29.60%
Cicutto 2005	-0.25559	0.150825	47.500%	4.3	59%
Cicutto 2013	-0.22931	0.071145	19.10%	3.6	57%
Gerald 2006	0.198715	0.08385	97%	3.3	54%
Gerald 2009	0.084988	0.200481	91%	4.2	57%
Howell 2005	0.151706	0.63494	75%	31.1	65%
McGhan 2003	-0.18135	0.227497	77.80%	6.6	58.80%
McGhan 2010	0.24628	0.18688	11.20%	3.6	62%
Persaud 1996	-0.23645	0.334639	69%	7.6	64%
Splett 2006	0.019	0.051402	80%	13.5	58%
London Schools			85%	9.0	58.20%

Enhanced SGA

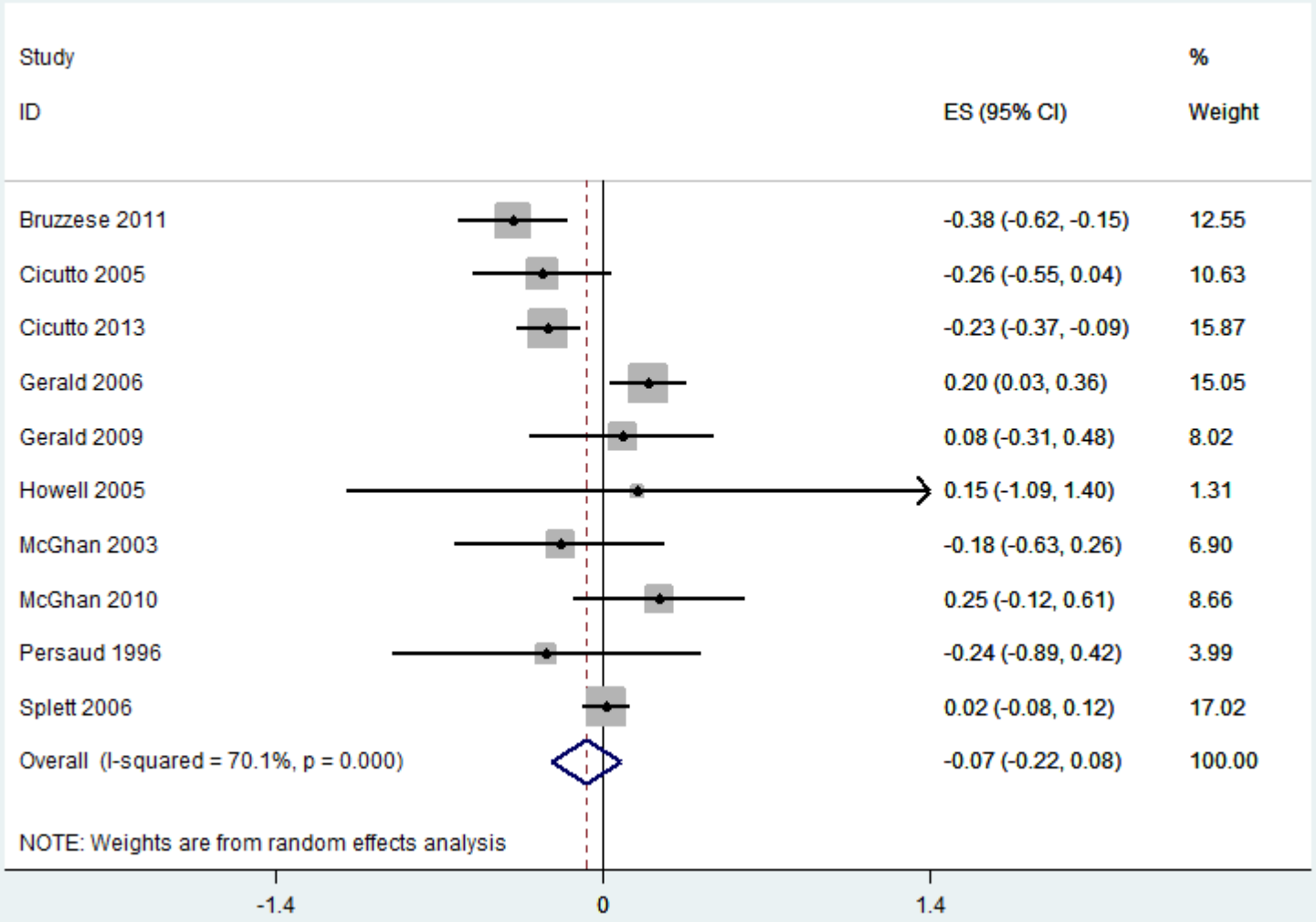


Recalibration

- Study weighting = size of study
- Recalibration = recalculate pooled effect size to more closely resemble expected effect for the inference population.
- Observed factors may include population, setting, and conditions of usual care or intervention implementation.
- Analogous in principle to reweighting secondary survey data to match key marginal distributions in the population
- Worked example

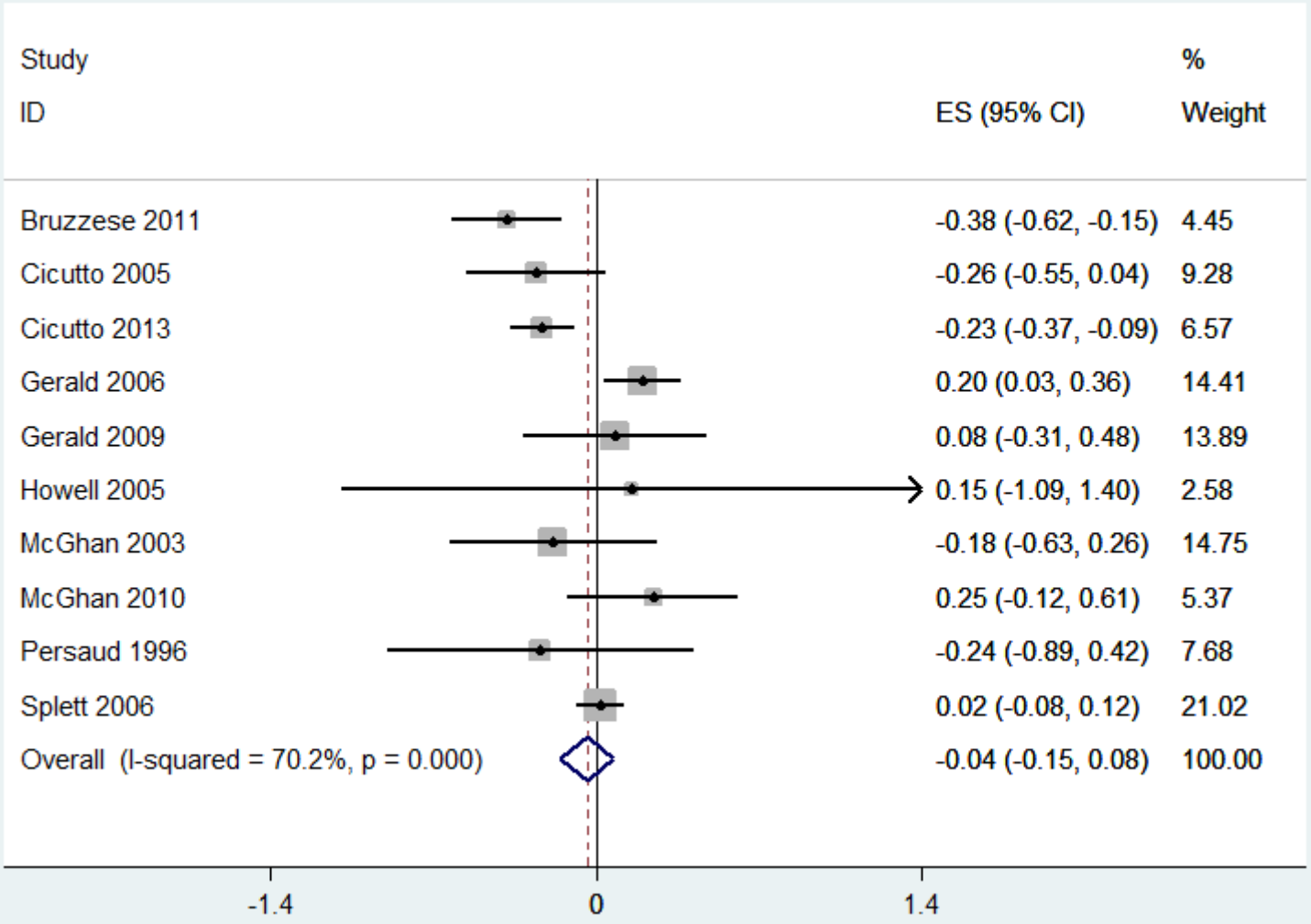
Recalibration

Global effect



Recalibration

London weighted



We have not cracked it!...yet?!?

- Issues, both conceptual and statistical, need to be explored in assessing viability
  - Choice of indicators, weights etc in recalibration
  - Choice of indicators, weights etc in ESGA
  - Greater use of data on usual care conditions
  - Potential Matching
- Not intended to supersede, present alongside
- Developments in the availability of data should also be matched with developments in the methods to utilise data

Institute of Education

EPPI CENTRE

UCL

Part 5: Conclusions and discussion points

Summary and Conclusions

- Are we negligent in the way in which we approach generalisability?
- Research wastage – reviews being produced but usable
- Many reasons why there is a gap between evidence and decision
- BUT little evidence of I.D.
- Instead critical look at evidence could help

Summary and Conclusions

- ‘Meta-analysis can give the big fact, but will struggle to provide anything more sophisticated’ (Glass)
- Has become prophetic?
- From a decision-maker perspective – utility of big fact from global settings?

Glass GV: **Meta-analysis at middle age: a personal history**. *Research synthesis methods* 2015, **6**(3):221-231.

Potential steps to address generalisability challenge

- 1. More extensive structured summary of the generalisability of evidence that reflects the theoretical concerns
- 2. Use the methods we have more effectively
  - Assessment and exploration
- 3. Exploit availability of new data sources
  - New opportunities
  - *Mixed methods?*

Thank you for listening!

D.Kneale@ucl.ac.uk

Acknowledgements:

- James Thomas, Alison O'Mara-Eves, Ian Shemilt, Antonio Rojas-Garcia, Brenda Hayanga, Dick Wiggins