≜UCL

Evidence Standards and Justifiable Evidence Claims

David Gough CEDIL – Centre for Evaluation Lecture Series London, 6th February 2019

> If you have questions please email: cedil@lshtm.ac.uk or tweet us on: @CEDILProgramme #CEDILPreLaunch



CENTRE



Department for International Development

www.cedilprogramme.org



Today

- 1. Evidence standards in different parts of an evidence ecosystem
- 2. How standards are specified / achieved
- 3. Evidence standards used by UK What Works Centres and evidence portals

Not discussing the specific methodological aspects required to make evidence claims.



Research evidence

Research provides us with information produced according to systematic rigorous accountable methods.

Evidence standards: criteria that must be met for preventive interventions to be judged 'tested and efficacious' or 'tested and effective'*

Though, of course, many different questions, methods driven by different theories and assumptions. Not just efficacy.

So define as: Standards for making justifiable evidence claims (whatever that claim may be)



Beyond effectiveness claims:

For example, from NICE Evidence Standards for Digital Health Technologies (2018):

- Relevance to current care pathways in the UK health and social care system
- Acceptability with users
- Appropriate economic analysis of economic impact

All types of justifiable evidence claims can help inform decision making (though today discuss impact studies).



EVIDENCE ECOSYSTEMS



Adapted from (Gough et al 2010, 2018)



Research on research use: increasing use



Framework from Science of Using Science (Langer, Tripney and Gough 2016)



Many terms used for knowledge use

- Knowledge Translation
- Knowledge Mobilization
- Knowledge Exchange
- K*

But what are the standards for such evidence claims?

• K Lite?



Evidence standards

Standards for making justifiable evidence claims.

Maybe be based on an:

Individual study - primary research

Evidence base – what we know and how we know it from the research that has been undertaken

Or in order to make evidence informed:

Guidance / recommendations – an efficient process to save every individual engaging with the evidence base



Primary research standards

Appraising the Weight of evidence = Suitability of method + Methodological standard + Relevance of focus*

For methods standards, for example, from prevention research**:

- Intervention description
- Measures and their properties
- Theory testing
- Valid causal inference
- Statistical analysis
- Efficacy claims
- Reporting



For valid causal inference the evidence concern is to avoid bias

Many rating scales. For example from the GRADE system:

- *Reduce rating of study if:*
 - Weak study design
 - Study limitations (with or without limitations in application such as attrition, non compliance, reliability and validity of outcome variables)
 - Indirectness of evidence / Relevance
- Increasing rating if:
 - Large effect
 - Confounding variables likely to reduce intervention effect
 - Dose-response gradient



Evidence base standards

- Strengths/limitations for individual included studies
 - As per previous slides
- Strengths/limitations *across* the included studies
 - Nature of totality of included studies + Extent of evidence

For example from GRADE: Inconsistency; Imprecision; Publication bias; Magnitude of treatment effect; Impact of confounders; Dose/response; Imprecisions (optimal information size)

• Strengths/limitations of the review method For example from ROBIS: *Concerns with the review process, Risk of bias, Relevance*



Review method

- Systematic review: rigorous explicit (research) process to identify and synthesize research evidence
- Narrative / expert review (? Basis and breadth of expertise and methods to identify and synthesize)
- Vote counting: Number of rigorous studies showing positive effect

| Domain | |
|-------------------------------------------------------------------------|----|
| 1. Concerns regarding specification of stud eligibility criteria | ły |
| 2. Concerns regarding methods used to identify and/or select studies | |
| 3. Concerns regarding used to collect data and appraise studies | 1 |
| 4. Concerns regarding the synthesis and findings | |



Guidance / recommendations

- Limitations for reviews
 - As per previous slides
- Guidance methods
 - Other data: how selected and their evidence standards
 - Values and priorities
 - Decision making processes
 - For example, the GRADE process:
 - the risks of using or not using the intervention;
 - the burden of experiencing the intervention;
 - the financial costs
 - other social preferences and values involved in making such decisions



2. How scientific standards are specified / achieved

- Methods manuals how to do the study
- Quality assurance processes how to enable and check quality
- Appraisal tools (and criteria) to assess and score quality achieved
- Reporting guidelines how to report
- Reporting appraisal tools how well reported



Potential problems: Individual studies or teams

- Media reporting of individual studies
 - ? Standards of the study
 - ? Relation to the wider evidence base
- REF impact: individual or team impact
 ? Relation to the wider evidence base
 2 False impact from initial structure and second second

? False impact from jointly attracted researchers and decision makers (not necessarily overt cherry picking)



Potential problems: Evidence base and guidelines not based on explicit rigorous processes

• Expert witnesses in court

Expert reviews of evidence 'Whether there is a reliable body of knowledge and experience to underpin the expert's evidence'* 'Reasonable scientific certainty'

• Health guidelines

Even WHO evidence based recommendations did not in the past use systematic reviews*

* Hodge 2017; ** Oxman, Lavis & Fretheim 2007



Two studies of evidence centres and portals



What Works Centres: https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3731 Web portal evidence standards: https://eppi.ioe.ac.uk/cms/Default.aspx?tabid=3743



UK What Works Centres plus web portals

- Centre for Ageing Better
- WWC Crime Reduction
- Early Intervention Foundation (EIF)
- Education Endowment Foundation (EEF)
- WWC Local economic growth
- National Institute of Excellence in Health and Social Care (NICE)
- WWC Wellbeing
- California Evidence-Based Clearinghouse for Child Welfare (CEBC), USA
- Clearinghouse for Labor Evaluation and Research (CLEAR), USA
- Conservation Evidence (CE), UK
- European Monitoring Centre for Drugs and Drug Addiction (EMCDDA), EU
- European Platform for Investing in Children (EPIC), EU
- Evidence Based Teen Pregnancy Programs (EBTP), USA
- Institute of Educational Sciences (IES) What Works Clearing House, USA
- What Works in Reentry Clearing House (WW RCH), USA



EVIDENCE ECOSYSTEMS





Centres / portals evidence breadth of questions N = 15

- Specific programmes: 5
- Broader issues or approaches: 10
- Guidance (as well as evidence base): 2
- Implementation: some relevant information only (e.g. costs; moderators and mediators; costs; implementation readiness



USA What Works Clearinghouse (Education)

STRENGTH OF EVIDENCE / BENEFIT

Effectiveness rating: 6 point scale based on the criteria of a certain number of studies: (i) Showing an effect; and (ii) Meeting evidence standards for each outcome domain

Highest rating = Positive effects: Strong evidence of a positive effect with no overriding contrary evidence with no overriding contrary evidence: (i) Two or more studies show statistically significant positive effects, at least one of which meets design standards without reservations; and (ii) No studies show statistically significant or substantively important negative effects.



Claims of effectiveness by evidence used

| Basis for evidence claim | Specific programmes | Issues/Intervention approaches |
|------------------------------|------------------------|--------------------------------|
| Systematic reviews | - | 6 |
| 'Narrative' / expert reviews | - | 2 |
| Listing studies and results | - | 2 |
| Vote counting | - | - |
| 1 or 2 good studies* | 5 | - |

*And maybe no evidence of harms (across the evidence base)

N = 15 topic based evidence centres and web portals (Gough et al 2018; Gough & White, 2018)



Methods used

- Methods manuals *common but vary in detail*
- Quality assurance processes common in expert processes
- Appraisal tools (and criteria/scales) *common* (often external such as GRADE or Maryland)
- Reporting guidelines *standardized within a portal*
- Reporting appraisal tools -

Level of detail of each varies considerably

• Which is better: detailed manual or simple criteria?



Questions raised

- Little consistency in how standards are:
 - Conceptualized (level of study, primary or evidence base)
 - Specified (Methods, QA, Criteria, Reporting)
 - Measured (Specific scales etc)
 - Considered good enough to make justifiable evidence claims
- Some evidence base evidence standards weak. '1 or 2 studies' approach common yet weaker than vote counting
- How useful are evidence claims about individual programmes rather than issues or approaches and evidence on theories of change?



Issues for users of evidence – avoiding K Lite

- Is the evidence from one study or the evidence base?
- If an evidence base, then was this systematic?
- What are the evidence standards for: Method of review; individual included studies; totality of evidence?
- What level of detail is provided in the standards from: method; internal QA; criteria; reporting?
- Is the research evidence really relevant? (As well as quality of method and execution and validity of result?)



www.cedilprogramme.org



Centre of Excellence for Development Impact and Learning





Department for International Development

http://eppi.ioe.ac.uk david.gough@ucl.ac.uk @ProfDavidGough