Epidemiology and Economics

Two Sides of the Same Coin in Impact Evaluation Research

Introduction

Assessment of the efficacy and effectiveness of an intervention is termed ‘impact evaluation’ in development economics and it assesses the impact in terms of intended and unintended health, social, and economic outcomes. Epidemiology and Development Economics share a research interest in investigating and measuring the impact of interventions in international development.

When seeking to answer similar evaluation questions in global health and international development, Developmental Economists and Epidemiologists use different analytical tools to measure impact of the same practice and policy relevant interventions. In some cases, the same tools are used in different ways. In other instances, the same tools are used but are described using different terminology.

Thus, policy makers all too often get incomplete information to make policy decisions, when they need a synthesis of the totality of the evidence of adequate quality – both epidemiology and development economics each have thousands of adequate quality studies which are being wasted if the policies are only based on one discipline. Some argue that evaluations by epidemiologists and economists differ in approach and therefore are not strictly comparable. However, this often results in readers and decision makers being confused and unable to reconcile the differences between evaluations from the two disciplines.

In 2016 Spiegelman suggested that the distinctions between epidemiology and economics that have been made confound the underlying common ground (Spiegelman, 2016). We believe the same is true for the more focussed field of efficacy/effectiveness/impact studies of Low- and middle-Income Countries (LMIC)

interventions on which this paper focuses given the interest in international development research shared by the two disciplines.

Here we take six substantive issues where there are commonly understood concepts but language differs to a greater or lesser degree. This leads to an avoidable failure to produce truly comprehensive and high-quality syntheses of the totality of the evidence.

Methodology

The work for this paper was undertaken in three stages. First, we examined two works on differences in the evaluation practices between the disciplines: the proceedings from a workshop on the gaps between the research methods of Impact Evaluation in Economics and Epidemiology, and a paper written by economists and epidemiologists based at the London School of Hygiene and Tropical Medicine (Powell-Jackson *et al.*, 2018) on the practice of randomized trials by the two disciplines.

In the second stage we reviewed paired papers (one economic and one epidemiological) that evaluated similar public health interventions. We identified examples of primary studies (one economic and one epidemiological) that either evaluated similar health interventions, and/or discussed similar methodology issues.

We then used these examples in the third stage in a workshop held in November 2017 at the Centre of Excellence in Impact Evaluation and Learning with economists

and epidemiologists experienced in undertaking evaluation research[1]. The aim of the workshop was to discuss different methods and approaches in evaluation practice and to suggest how these differences could inspire future opportunities for interdisciplinary work. The studies were identified from the Cochrane Library review of each topic and matched based on the research question – the topics were: (1) effect of incentives on take up and use of bed nets, (2) impact of WASH (Water, Sanitation and Hygiene) interventions on diarrhoea, (3) anti-retroviral treatment (ART) for HIV, (4) conditional cash transfers (CCT) as monetary incentives given in return for fulfilling specific behavioural conditions, 5) Effectiveness of deworming of children. To be classified as epidemiological papers, the research had to be published in a medical or epidemiological journal. For economics papers, its authors had to be affiliated to an economics department and/or the paper had to be published in an economics journal.

Table A1 summarises the studies selected from each discipline which we use to illustrate examples of the issues discussed in this paper.

Comparing Economic and Epidemiological Approaches to Impact Evaluation

We recognise that a great deal of heterogeneity exists within the disciplines of economics and epidemiology and, therefore, this paper focuses on a specific category of each discipline. For economics, the focus chosen was development

[1] Participants in the CEDIL workshop: Orazio Attanasio (Yale University), Mike Clarke (Queen's University Belfast), Teresa Ejer (World Health Organization), Josephine Exley, Richard Hayes, and Tim Powell Jackson (all at the London School of Hygiene & Tropical Medicine), Peter Tugwell (University of Ottawa), Hugh Waddington (international Initiative for Impact Evaluation), Vivian Welch, and Howard White (both at Campbell Collaboration) and Marcella Vigneri (Centre of Excellence in Impact Evaluation and Learning, London School of Hygiene & Tropical Medicine).

economics applied to measuring the impact of socio-economic interventions. For epidemiology, the focus chosen was the effectiveness of interventions relevant to practice and policy in low- and middle-income countries.

The workshop identified six areas of difference: (1) definition of the research question, and outcome measures of impact; 2) use of theory to underpin empirical models used in the evaluation; (3) study design; (4) publication of protocols and pre-specified outcomes; (5) generalisability and transferability of findings from an individual case study; and (6) replicability. We discuss each area below, with a reminder that any differences suggested are not intended to claim universal truths or to generalise across the spectrum. Counter examples no doubt exist but we believe the issues highlighted will resonate with readers and support constructive thinking in building interdisciplinary methods to generate stronger evidence from impact evaluation research.

1.Research questions and outcomes measures

The first difference identified between evaluations in economics and epidemiology is in the type of research question that each discipline seeks to answer. Impact evaluations in economics set out from an observable and recognised problem on the ground and typically use theoretical models of human behaviour to explain behaviour and phrase their research questions accordingly. Epidemiologists rarely use models or theory and take a pragmatic approach, focussing directly on measures that capture what happens when a certain intervention is implemented. These approaches are reflected not just in how research questions are framed in each discipline's impact evaluations (and in the type of outcome variables used to

measure impact), but also in the context and supporting information that is provided to justify the need to answer these questions. To illustrate this point, we looked at two studies that examined the impact of WASH interventions on diarrhoea-related outcomes among younger siblings of school-going children.

The clinical epidemiology paper by Dreibelbis et al. (2014) assessed the health and educational impacts of two WASH improvement interventions carried out in schools in Kenya on the prevalence of diarrhoea (Dreibelbis et al., 2014). The outcome indicator selected to measure impact was a 1-week prevalence rate of diarrhoea episodes among children of school age, measured from a sample of individuals who were interviewed twice in 26 months. The body of the paper focuses on explaining the study design and the methods adopted for the trial, while the section on results explains in detail different statistics for the odds of diarrhoea associated with the survey population in different arms of the trial.

The economic paper by Duflo et al. (2015) evaluated a village-level intervention promoting the adoption of household latrines and bathing facilities, a community water tank, and a distribution system that supplies piped water to household taps (Duflo et al., 2015). The authors detail the typical sequence of events for implementing the intervention, explaining in detail the procedures adopted by extension workers, the expected responses of village leaders, and a costing exercise to show the financial requirements for households constructing latrines and bathing facilities. The paper provides a large amount of detailed explanation of the economic study design, the sampling and evaluative method procedures, which takes the reader through the rationale for why and how the intervention is expected to change the behaviour of the targeted households.

The different framing of the research questions across disciplines is also mirrored in the type and number of outcome variables measured. For example, economic studies often tend to analyse several measures of impact that are correlated with potential externalities of the project, as well as unintended (positive and negative) effects that result from the intervention. In clinical epidemiology trials, for example, papers typically report a single primary outcome (complemented by power calculation to determine the optimal sample size calculation), which is investigated and gets most emphasis, and some secondary measures of impact, which may also be discussed and published in separate papers.

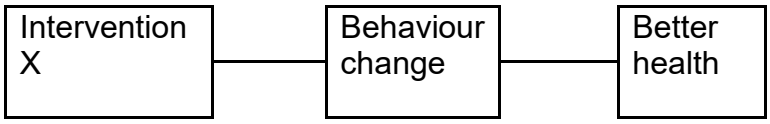
2. Using Theory to Underpin the Evaluation Problem

In economics, theory plays an important role to ensure rigorous identification of the causal pathway to impact by looking at the most plausible models of human behaviour and deriving from this a set of testable assumptions. Theory is used to provide a rationale for the analysis, interpret the experimental results of an evaluation, and expand the usefulness of the experiment. Economists also use theory to generalise beyond the experiment, for example by making predictions about the effect of a future change in the design of a policy. External validity is the term used by Economists in impact evaluation research to refer to the transportability of causal relationships across setting. Causal effects are often presented in economic models as sets of equations, which may include non-linear relationships, lags and feedback loops (known as simultaneity). Although the practice of grounding research in theory is not universal among economists, using theory to inform the design of a study is generally considered the gold standard; it guides empirical

testing of research hypotheses and it strengthens the analysis underlying impact evaluations. Papers in economic journals will usually have a section titled 'The Model' which outlines the theory being tested. In the language of Economics, the model provides the assumptions underlying the data analysis. One example of this in Economics is *an evaluation of the PROGRESA Conditional Cash Transfer (CCT) programme* (Attanasio et al., 2012). The intent of this paper is to estimate a structural model of education choices using data from the PROGRESA randomised experiment, and to use the model to simulate the effect of changes to some of the parameters of the programme (the use of monetary incentives) to incentivise school enrolment. In this case, the use of experimental variation in the data allows estimation of a structural model that offers a conceptual framework for a richer policy analysis, and a better understanding of the mechanisms driving the effects.

In epidemiology, while structural equation modelling is commonly used for complex interventions, the common practice is to use visual logic models (also called causal chain or analytic frameworks) to show how the intervention leads to outcomes that occur under different scenarios without making explicit assumptions about individual behaviour and social interactions. Figure 1, for example, provides a simplified version –carrying out the intervention brings about behaviour change, which in turn leads to better health.

Figure 1: Logic Model /Causal Chain /Analytic framework



Applying this to the same issue of CCT, epidemiologists used this logic model to examine the impact of *Bolsa Alimentação*, a national health related CCT programme in Brazil on growth among children of beneficiary households (Morris et al., 2004). The paper opens by explaining how poverty-related factors such as lack of access to nutritionally rich diets, inadequate infant feeding practices, and repeated illness, are contributing factors to stunting. Therefore, based on the assumption that children from poorer families would benefit from significant improvements in living conditions, for example through food supplements, the authors test the hypothesis that offering direct transfers of money to very poor families through the programme ('intervention' in figure 1) leads to an improvement in the children's growth ('better health' in figure 1) as a result of behaviour change brought about by improving living standards. In this example, logic is used to explain the order of causation that follows from introducing a specific incentive that is expected to change behaviour.

There are a few epidemiology examples where theory is developed and used to test the assumptions behind behavioural interventions such as hand washing practices in a similar way to in economics (Eccles et al., 2005). An example of this is the study by Dawood et al of modelling of estimates of global mortality associated with pandemic influenza A H1N1 virus circulation. This translates to looking at transmission mechanisms capturing the dynamic nature and spread of diseases and to incorporating positive and negative feedback characteristics of infectious processes (Dawood et al., 2012).

3. Study Designs

Evaluation studies conducted by economists frequently use quasi-experimental methods, although randomised trials have become accepted as the golden standard (R. Khandker, Gayatri B., & Hussain A., 2010). Methods used in economic studies such as instrumental variable models (IV), local average treatment effect (LATE) models, regression discontinuity design (RDD) are now starting to be considered by epidemiologists (Bor et al 2014, Bärnighausen et al 2017). Interrupted-time series designs and cohort studies, which are used by epidemiologists, are practically absent in evaluations conducted by economists.

Economists and epidemiologists employ different methods to address selection bias. In randomised trials, epidemiologists will often focus on an intention-to-treat analyses (where those allocated to the intervention and control groups are compared regardless of whether they received their allocated intervention) but will sometimes use either a per-protocol analysis (where those who adhere to the allocated intervention are compared to those in the control group who do not receive it), or an "as treated" analysis (where those who adhere to the intervention are compared to those who do not adhere to it), in order to identify the impact of an intervention on those who actually receive it. In observational studies, which are more often used by Economists than by Epidemiologists, the problem of selection bias concerns more generally the ability to identify a suitable counterfactual. The strategy used to address the problem involves choosing among a variety of matching procedures (White et al. 2014), as well as using instrumental variables, or 'encouragement design' randomised trials, where participants are randomised to the offering of the

intervention, and then proceed to estimate a local-average treatment effect for people encouraged to take-up the treatment (White et al. 2017).

4. Publishing Protocols

The publication of protocols and pre-registration of research has become the norm for randomised trials done by epidemiologists and is slowly becoming so for economists. This is also the case for systematic reviews in epidemiology (Booth *et al.*, 2011, 2013). When the evaluation method adopted for a study is not a randomised trial, then the practice of publishing a protocol is less frequent in both disciplines. This is likely to be because randomised trials have defined guidelines about pre-registration for publication of the final report in many journals, whereas the same is not true for other study designs.

Protocols are an important insurance against data mining (Ioannidis, 2005), p-hacking (Head et al., 2015), and selective reporting (Kirkham et al., 2018). Data mining is the practice of analysing and manipulating the data until it produces the desired signal and statistical significance level on the variable of interest. An associated problem is p-hacking, which occurs when researchers refine their analyses until they reach statistical significance. Many argue that current scientific publication practices create strong incentives to publish statistically significant (usually positive) results, and this in turn pushes researchers to selectively seek publication of only statistically significant research findings. The introduction of clear protocols and statistical analysis plans for randomised trials makes this practice

easier to identify. Instances of p-hacking are more common in economics but there are some examples in clinical epidemiology, especially for diagnostic and prognostics (Albarqouni et al., 2017). Another known problem for both disciplines is selective reporting of only positive results, omitting negative ones (Brodeur et al, 2019, and Dwan et al., 2014) which often arises from pressure to report from a research study. Prevention of data mining, p-hacking, and selective reporting needs funding organisations and ethics committees to require the publication of protocols and full data analysis from all studies they fund or approve.

5. Replicability

In economics replication is an important component of impact evaluation work. It is standard practice that when a new result is found, other scientists around the world may attempt to replicate the results. The replication of results with new data is referred to as external replication, whereas internal replication is the attempt to reproduce the results of the original study using the same data. In this respect, the International Initiative for Impact Evaluation (3ie) has gone a step further by funding internal replications of high profile studies in international development (3ie).

In clinical epidemiology, there is an active debate about how the discipline needs to address the trade-off between the importance of replication as a basic tenet of the scientific method, versus the concern about research waste (Glasziou and Chalmers, 2017), a major contributor being the remarkably widespread inappropriate replication of primary studies long after the answer is known (Fergusson, 2005). A well-known example of this is the contrasting recommendations on global policies around deworming from the development economics discipline and the clinical

epidemiologists that required an independent replication to resolve the differences (Welch et al. 2017).

6. Generalisability and Transferability

Both developmental economics and epidemiology insist on internal validity (i.e. design and conduct should attempt to minimize the possibility of bias) but to be useful, whether assessing clinical, global health, or social policy interventions, the results must also be relevant in similar settings. The latter requirement is generally referred to as external validity or generalisability.

Economists are often explicit about how their findings may extrapolate to other contexts using *structures speculation* (Banerjee et al 2016). For example, Dupas examined whether demand for insecticide treated bednets (ITNs) changed following the introduction of price subsidises and exposure to different marketing messages among rural households in Kenya (Dupas, 2009). The study found significant price elasticity to the cost of ITNs and no variation in uptake due to the marketing messages used. The findings of the impact evaluation are discussed in relation to those reported in other economic case studies and are used to explain more widely the reason why poor households, who have generally limited access to savings, systematically underinvest in health services.

Epidemiologists are divided on this issue. On the one hand epidemiology articles usually express caution about translating results from one study to other settings. Rothwell, for example, has rightly suggested that generalizability is a “slippery concept” in clinical epidemiology and medical science (Rothwell, 2005). On the other hand, given the increasing importance of practice (e.g., Clinical Guidelines produced

by innumerable clinical specialties) and policy guidelines (of which the WHO for example produces over 100 every year https://www.who.int/publications/who-guidelines), clinical epidemiology groups such as AHRQ (Gartlehner et al, 2006) and GRADE's Adolopment Evidence to Decision framework (Schünemann. HJ, 2017) have developed useful and widely adopted frameworks to ensure external validity.

4. Conclusions

In this paper we analysed some of the differences between efficacy/effectiveness/impact studies of practice and policy relevant interventions in LMIC. The rich mix of evaluation methods available in economics and epidemiology suggests that it is indeed desirable to build opportunities for experimenting beyond common practices in each discipline, with much scope for cross-discipline learning and adaption (e.g., the use of the regression discontinuity design in epidemiology and increased emphasis on systematic reviews in economics to advance learning).

First, the use theory in Economics to test and predict behavioural change in response to an intervention is an element of enrichment in evaluation research. This approach could be combined usefully with the epidemiological practice of using a general theory to frame the evaluation of a specific intervention to predict the outcome under different scenarios and enhance the transferability of the research findings.

Second, the generalisability and transferability of evaluation outcomes is recognised as difficult, and one where both disciplines could benefit from with a better description of the settings of the intervention to assess the external validity of the findings and to guide and refine the design of interventions in other settings.

Third, the rich mix of evaluation methods available in economics and epidemiology suggest that it may be possible and desirable to build opportunities for experimenting beyond common practices in each discipline, with much scope for cross-discipline learning and adaption (e.g., the use of the regression discontinuity design and increased emphasis on systematic reviews to advance learning).

Fourth, the practice of setting the findings of evaluations in the context of the totality of the related evidence (e.g., in systematic reviews) is something that both economists and epidemiologists could substantially improve on. According to research published in the Lancet and an accompanying editorial (Clark & Horton, 2010; Clarke et al., 2010), the lack of integration of results of new trials into existing systematic reviews greatly diminishes both the scientific and ethical value of the trials.

Evaluations conducted in economics and epidemiology on similar topics offer unique opportunities for advancing interdisciplinary science work of policy relevance. But *what makes interdisciplinary science work*? In this paper we suggest two clear avenues. Firstly, the identification of an important question in international development, for example in the public health domain, where both disciplines share a common research interest and a willingness to conduct the necessary research. Secondly, offering research grants for collaborative projects across disciplines (where a team combining economists and epidemiologist would be one example) that are underpinned by discussion and agreement on the best combination of methods from the participating disciplines to evaluate what works and how (and its possible adaptability to other contexts).

References

3ie Replication. *International Initiative for Impact Evaluation,* Online: http://www.3ieimpact.org/en/evaluation/impact-evaluation-replication-programme/ (accessed on 5 April 2019).

Albarqouni, L.N., Lopez-Lopez, J.A., & Higgins, J.P.T. (2017). Indirect evidence of reporting biases was found in a survey of medical research studies. *Journal of Clinical Clinical epidemiology ,* 83, 57-64.

Attanasio, O.P., Meghir, C., & Santiago, A.N.A. (2012). Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to Evaluate PROGRESA. *The Review of Economic Studies,* 79, 37-66.

Bärnighausen, T. *et al.* (2017) 'Quasi-experimental study designs series—paper 1: introduction: two historical lineages', *Journal of Clinical Clinical epidemiology* . doi: 10.1016/j.jclinepi.2017.02.020.

Banerjee, A.V., Chassang, S., & Snowberg, E. (2016). Decision Theoretic Approaches to Experiment Design and External Validity. *National Bureau of Economic Research Working Paper Series,* No. 22167.

Bor, J., Moscoe, E., Mutevedzi, P., Newell, M.L., & Barnighausen, T. (2014). Regression discontinuity designs in clinical epidemiology : causal inference without randomized trials. *Clinical epidemiology ,* 25, 729-737.

Brodeur, A., Cook, N., & Heyes, A. (2020). Methods Matter: P-Hacking and Causal Inference in Economics. *American Economic Review*.

Clarke, M., Hopewell, S., & Chalmers, I. (2010). Clinical trials should begin and end with systematic reviews of relevant evidence: 12 years and waiting. *Lancet,* 376, 20-21.

Clark, S., & Horton, R. (2010). Putting research into context--revisited. *Lancet,* 376, 10-11.

Dawood, F.S., Iuliano, A.D., Reed, C., Meltzer, M.I., Shay, D.K., Cheng, P.Y., et al. (2012). *Lancet Infectious Diseases,* 12, 687-695.

Dreibelbis, R., Freeman, M.C., Greene, L.E., Saboori, S., & Rheingans, R. (2014). The impact of school water, sanitation, and hygiene interventions on the health of younger siblings of pupils: a cluster-randomized trial in Kenya. *American Journal of Public Health,* 104, e91-97.

Duflo, E., Greenstone, M., Guiteras, R., & Clasen, T. (2015). Toilets can work: Short and medium run health impacts of addressing complementarities and externalities in water and sanitation. *National Bureau of Economic Research Working Paper Series,* No. 21521.

Dupas, P. (2009). What matters (and what does not) in households' decision to invest in malaria prevention? *American Economic Review,* 99, 224-230.

Dwan, K., Altman, D. G., Clarke, M., Gamble, C., Higgins, J. P. T., Sterne, J. A. C., Kirkham, J. J. (2014). Evidence for the Selective Reporting of Analyses and Discrepancies in Clinical Trials: A Systematic Review of Cohort Studies of Clinical Trials. *PLoS Medicine*. https://doi.org/10.1371/journal.pmed.1001666

Eccles, M., Grimshaw, J., Walker, A., Johnston, M., & Pitts, N. (2005). Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings. *Journal of Clinical Clinical epidemiology ,* 58, 107-112.

18-29; discussion 229-32. doi: 10.1191/1740774505cn085oa. PMID: 16279145.

Gartlehner G, Hansen RA, Nissman D, et al. (2006) Criteria for Distinguishing Effectiveness From Efficacy Trials in Systematic Reviews. Rockville (MD): Agency

for Healthcare Research and Quality (US); 2006 Apr. (Technical Reviews, No. 12.) Available from: https://www.ncbi.nlm.nih.gov/books/NBK44029/

Glasziou P, Chalmers, I. (2017) Ill informed replications will increase our avoidable waste of research. The BMJ Opinion: https://blogs.bmj.com/bmj/2017/03/20/ill-informed-replications-will-increase-our-avoidable-waste-of-research/

Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD (2015) The Extent and Consequences of P-Hacking in Science. PLoS Biol 13(3): e1002106. https://doi.org/10.1371/journal.pbio.1002106

Ioannidis JPA (2005) Why Most Published Research Findings Are False. PLoS Med 2(8): e124. https://doi.org/10.1371/journal.pmed.0020124

Khandker, S., Gayatri B., S. and Hussain A., K. (2010) *Handbook on Impact*, *Learning*.

Kirkham J J, Altman D G, Chan A, Gamble C, Dwan K M, Williamson P R et al. (2018) Outcome reporting bias in trials: a methodological approach for assessment and adjustment in systematic reviews *BMJ* ; 362 :k3802 doi:10.1136/bmj.k3802

Morris, S.S., Olinto, P., Flores, R., Nilson, E.A., & Figueiro, A.C. (2004). Conditional cash transfers are associated with a small reduction in the rate of weight gain of preschool children in northeast Brazil. *Journal of Nutrition,* 134, 2336-2341.

Powell-Jackson, T., Davey, C., Masset, E., Krishnaratne, S., Hayes, R., Hanson, K., & Hargreaves, J. R. (2018) Trials and tribulations: cross-learning from the practices of epidemiologists and economists in the evaluation of public health interventions, *Health Policy and Planning*, 33, 5, 702–706, https://doi.org/10.1093/heapol/czy028.

Rothwell, P.M. (2005). External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet,* 365, 82-93.

Sackett DL, Haynes RB, Tugwell P. (1985) Clinical Epidemiology: A Basic Science for Clinical Medicine. First Edition. Boston: Little, Brown.

Schünemann, Holger J. et al. (2017) GRADE Evidence to Decision (EtD) frameworks for adoption, adaptation, and de novo development of trustworthy recommendations: GRADE-ADOLOPMENT**.** *Journal of Clinical Epidemiology*, Volume 81, 101 - 110.

Spiegelman, D. (2016). Evaluating Public Health Interventions: 1. Examples, Definitions, and a Personal Note. *American Journal of Public Health,* 106, 70-73.

Tugwell P, Welch, VA, Karunananthan S, Maxwell LJ, Akl EA, Avey MT et al. (2020) When to replicate systematic reviews of interventions: consensus checklist *BMJ*; 370 :m2864

Welch VA, Ghogomu E, Hossain A, Awasthi S, Bhutta ZA, Cumberbatch C, Fletcher R, McGowan J, Krishnaratne S, Kristjansson E, Sohani S, Suresh S, Tugwell P, White H, Wells GA. (2017) Mass deworming to improve developmental health and wellbeing of children in low-income and middle-income countries: a systematic review and network meta-analysis. *Lancet Glob Health*. Jan;5(1):e40-e50. doi: 10.1016/S2214-109X(16)30242-X.

White H, Raitzer D, editors. (2017) Impact Evaluation of Development Interventions: A Practical Guide. *Asian Development Bank*.

White H, Sabarwal S. (2014) Quasi-experimental Design and Methods, Methodological Briefs: Impact Evaluation 8. Florence, Italy: UNICEF Office of Research.

Table A1: Economic and Clinical epidemiology Papers Consulted

IE areas Disciplines	Research question	Use of theory	Study design	Publication of protocols	Generalisability/ Transferability	Replicability
Epidemiology	Dreibelbis, R. et al. (2014) **The Impact of School Water, Sanitation, and Hygiene Interventions on the Health of Younger Siblings of Pupils: a Cluster-Randomized Trial in Kenya.** American Journal of Public Health; 104 (1)	Eccles, M., Grimshaw, J., Walker, A., Johnston, M., & Pitts, N. (2005). **Changing the behavior of healthcare professionals: the use of theory in promoting the uptake of research findings**. *Journal of Clinical Clinical epidemiology* , 58, 107-112	Bor, J., Moscoe, E., Mutevedzi, P., Newell, M. L., and Bärnighausen, T. (2014) **Regression Discontinuity Designs in Clinical epidemiology. Causal Inference Without Randomized Trials** Clinical epidemiology 25: 729–737	Dwan, K., Altman, D. G., Clarke, M., Gamble, C., Higgins, J. P. T., Sterne, J. A. C., Kirkham, J. J. (2014). **Evidence for the Selective Reporting of Analyses and Discrepancies in Clinical Trials: A Systematic Review of Cohort Studies of Clinical Trials**. *PLoS Medicine*.	Rothwell, P.M. (2005). **External validity of randomised controlled trials: "to whom do the results of this trial apply?**". *Lancet,* 365, 82-93.	Glasziou and Chalmers (2017) **https://blogs.bmj.com/bmj/2017/03/20/ill-informed-replications-will-increase-our-avoidable-waste-of-research/** I
Economics	Duflo E., et al. **Toilets Can Work: Short and Medium Run Health Impacts of Addressing Complementarities and Externalities in Water and Sanitation**. NBER Working Paper No. 21521, 2015	Attanasio, O., Meghir, C., and Santiago A. (2012) **Education Choices in Mexico: Using a Structural Model and a Randomized Experiment to evaluate Progresa** The Review of Economic Studies, Vol. 79, No. 1, pp. 37-66	R. Khandker, S., Gayatri B., S. and Hussain A., K. (2010) ***Handbook on Impact, Learning***.	Brodeur, A., Cook, N. and Heyes, A. (2020) '**Methods Matter: P-Hacking and Causal Inference in Economics**', *American Economic Review*.	Dupas, P. (2009) What **Matters (and What Does Not) in Households' Decision to Invest in Malaria Prevention?** American Economic Review 99(2):224-30	3ie Replication studies **https://www.3ieimpact.org/evidence-hub/replication-studies-status**