International Journal of Epidemiology, 2017, 1251–1276
doi: 10.1093/ije/dyx039
Advance Access Publication Date: 24 April 2017
Original article

Health Policy

Spillover effects on health outcomes in low- and middle-income countries: a systematic review

Jade Benjamin-Chung,[1]* Jaynal Abedin,[2] David Berger,[3] Ashley Clark,[4] Veronica Jimenez,[1] Eugene Konagaya,[1] Diana Tran,[1] Benjamin F Arnold,[1] Alan E Hubbard,[5] Stephen P Luby,[6] Edward Miguel[3] and John M Colford Jr[1]

[1]Division of Epidemiology, University of California, Berkeley, CA, USA, [2]Centre for Communicable Diseases, International Centre for Diarrhoeal Disease Research, Dhaka, Bangladesh, [3]Department of Economics, University of California, Berkeley, CA, USA, [4]Goldman School of Public Policy, University of California, Berkeley, CA, USA, [5]Division of Biostatistics, University of California, Berkeley, CA, USA and [6]Division of Infectious Disease and Geographic Medicine, Stanford University, Stanford, CA, USA

*Corresponding author. Division of Epidemiology, UC Berkeley School of Public Health, 101 Haviland Hall, Berkeley, CA 94720-7358, USA. E-mail: jadebc@berkeley.edu

Editorial decision 14 February 2017; accepted 24 February 2017

Abstract

Background: Many interventions delivered to improve health may benefit not only direct recipients but also people in close physical or social proximity. Our objective was to review all published literature about the spillover effects of interventions on health outcomes in low-middle income countries and to identify methods used in estimating these effects.

Methods: We searched 19 electronic databases for articles published before 2014 and hand-searched titles from 2010 to 2013 in five relevant journals. We adapted the Cochrane Collaboration's quality grading tool for spillover estimation and rated the quality of evidence.

Results: A total of 54 studies met inclusion criteria. We found a wide range of terminology used to describe spillovers, a lack of standardization among spillover methods and poor reporting of spillovers in many studies. We identified three primary mechanisms of spillovers: reduced disease transmission, social proximity and substitution of resources within households. We found the strongest evidence for spillovers through reduced disease transmission, particularly vaccines and mass drug administration. In general, the proportion of a population receiving an intervention was associated with improved health. Most studies were of moderate or low quality. We found evidence of publication bias for certain spillover estimates but not for total or direct effects. To facilitate improved reporting and standardization in future studies, we developed a reporting checklist adapted from the CONSORT framework specific to reporting spillover effects.

Conclusions: We found the strongest evidence for spillovers from vaccines and mass drug administration to control infectious disease. There was little high quality evidence of spillovers for other interventions.

Key words: Spillover effects; indirect effects; herd effects; herd immunity; diffusion; externalities; interference

Key Messages

- Spillovers are the effects of an intervention on individuals who did not receive the intervention but who are connected to recipients through physical or social proximity.
- Our systematic review found a wide range of terminology used to describe spillover effects, a lack of standardization among spillover measurement methods and poor reporting of spillover effects in many studies.
- The strongest evidence for spillover effects exists for studies of vaccines and mass drug administration to control infectious disease. The evidence of spillover effects for other interventions is of limited or poor quality.
- To facilitate improved reporting and standardization in future spillover studies, we developed a reporting checklist adapted from the CONSORT framework, specific to reporting of spillover effects.

Introduction

Interventions delivered to improve health are frequently targeted to specific populations. Such interventions may benefit not only direct recipients but also those who did not receive the intervention but are connected to recipients through physical or social proximity. Such effects, which we refer to as 'spillovers', are a component of the population-level impact of interventions. A wide range of terms has been used to describe spillovers in disciplines including economics, public health and political science: externalities,[1,2] interference,[3–7] contamination,[8] herd immunity,[9–11] stable unit treatment value assignment (SUTVA) violations,[12] stability violations[13] and indirect effects.[14,15]

A 'positive' spillover is an effect in the same direction as the treatment effect (on intervention recipients); conversely, a 'negative' spillover is an effect in the opposite direction of the treatment effect. If positive spillovers are present, studies that only estimate treatment effects without measuring spillover effects will underestimate the effectiveness of the intervention. In addition, cost-effectiveness calculations that exclude such spillovers may underestimate intervention benefits. Conversely, if negative spillovers are present, evaluations that do not measure spillover effects may overestimate health impacts and cost-effectiveness. Furthermore, negative spillovers could attenuate the effects of an otherwise beneficial intervention. For these reasons, when an intervention is capable of diffusing through a population, information about spillover effects is an important complement to estimates of treatment effects—when spillovers are found to be large and positive, such evidence may, for example, justify national scale-up of an intervention or a public subsidy.[16] The well-documented evidence of spillovers (i.e. 'herd effects') of many vaccines justifies the cost-effective scale-up of immunization efforts to a global level via programmes such as the Global Alliance for Vaccines and Immunization.[17,18]

In the epidemiological literature focusing on trials of interventions other than vaccines, spillovers have motivated randomizing clusters rather than individuals in order to minimize the chance of spillovers into control units (i.e. 'contamination').[8,19,20] Outside vaccine studies, epidemiologists consider spillovers in designing studies (e.g. the expectation of spillovers may motivate cluster-randomization instead of individual-randomization[8,19,20]), but they typically do not estimate spillovers explicitly alongside direct effects. Recently, spillovers have increasingly been framed as a quantity of interest themselves, particularly in economics where a growing literature describes spillovers of interventions including school-based deworming[2] and insecticide-treated bed nets.[21] On the whole, methods for estimating spillovers have developed independently within disciplines.

We conducted this systematic review in order to summarize the literature about spillover effects on health in low- and middle-income countries. We restrict our review to such countries because this review was supported by the International Initiative for Impact Evaluation (3ie), which focuses on low- and middle-income countries.[22] Our objective is to provide a broad summary of the types of spillover effects that have been measured to date.

Methods

Protocol and registration

We attempted to register our protocol with the Campbell Coordination International Development Coordinating Group (IDCG). However, because our protocol included a synthesis of methods in addition to a systematic review, the IDCG did not accept our protocol. Instead, the International Initiative for Impact Evaluation (3ie), which funded this endeavour, supported the development of the protocol and provided both internal and external review.

Eligibility criteria

A complete description of eligibility criteria is available in Supplement 1, as Supplementary data at *IJE* online. Briefly, we included studies that: (i) were conducted in low- or middle-income countries as defined by the World Bank[23] (as required by our funder); (ii) were quantitative studies evaluating an intervention; (iii) measured health outcomes; and (iv) included a comparison group with sufficient detail about the design and comparison group to determine whether there were serious threats to internal or external validity.

Information sources

We searched 19 electronic databases that contained articles on health, economics, social science, and other disciplines for articles published before 2014 (Supplement 2, available as Supplementary data at *IJE* online). In addition, we hand-searched all titles from 2010 through 2013 in the following journals, which we considered most likely to include relevant articles: *Health Economics*, *The Journal of Development Effectiveness*, *The Lancet*, *PLoS Medicine* and the *World Bank Policy Research Working Papers*.

Search

A detailed description of our search strategy is listed in Supplement 3, available as Supplementary data at *IJE* online. We searched reference lists of texts classified as eligible in the original search. We also identified records that cited included texts from the original search using Google Scholar. Following the search process, all records were merged, duplicates were removed and a unique ID was assigned to each record.

Study selection

At least one team member reviewed each record for relevance. Titles that were clearly not eligible for the review received no further review. We reviewed each available abstract that passed the title review for relevance. If an abstract was not available but a full text was, we reviewed the full text instead. Of the abstracts deemed relevant, we reviewed each full text for relevance. For records that were deemed not to be relevant, team members recorded the first reason for exclusion identified. If multiple versions of a paper were available, we included the most recent version of the paper.

Data collection process

We extracted data from included texts, and then a second team member independently checked all extracted data. In one case, spillover results were mentioned and disaggregated results were not listed in the publication, but the authors mentioned that results were available upon request.[24] We contacted the authors to request these results but did not receive a reply.

Data items

We extracted information about: interventions; outcomes measured; study site; primary study design; study design used to estimate spillovers; purported spillover mechanism; scale of spillover (e.g. household versus village); average cluster-level treatment coverage; whether or not spillover measurement was pre-specified; and direct effect, total effect, overall effect and spillover effects reported numerically in tables or text. If multiple effects or model specifications were used to estimate the direct, total or overall effects, we chose the estimate that appeared to be the primary finding reported by the author and that allowed the greatest comparability of the effect with the spillover estimates. We considered spillovers to be pre-specified if spillover estimation methods were included in the original study protocol.

Risk of bias in individual studies

We classified specific criteria related to risk of bias for each study using criteria compiled from relevant fields.[25–29] Duplicate assessment of risk of bias was performed for a 20% subsample. Classification was not blinded. Co-authors of this systematic review who had authored included studies did not participate in the classification of risk of bias criteria for any studies. For studies that performed secondary analyses, we attempted to obtain the original publication and incorporated information from the original publication(s) into our risk of bias assessment. We only assessed the risk of bias for the elements of the study that estimated effects on health outcomes. We also created an overall classification of risk of bias for individual studies by adapting the Cochrane GRADE approach[30,31] to spillover estimation (Supplement 4, available as Supplementary data at *IJE* online). We developed these criteria through an iterative process in which we revised our classification system after the initial risk of bias assessment for each study and discussion with multiple reviewers. We then classified each study's overall quality of evidence as 'very low', 'low', 'medium' or 'high'.

Summary measures

Due to the wide range of interventions and outcomes evaluated in included studies, we did not consider it reasonable to assume that the studies included were independent and that a common treatment effect existed across all included studies.[32] Thus, we did not calculate summary measures.

Synthesis of results

For spillover types for which a sufficient number of studies reported estimates, we standardized results for binary outcomes on the relative scale. We present results as the percentage reduction in outcomes attributable to intervention {[1- relative risk (RR)] x 100%}. For results reported only as risk differences, we calculated the percentage reduction by dividing the risk difference by the probability of the outcome in the control group. We chose the relative scale instead of the additive scale because the interpretation of risk differences depends upon the risk of the outcome among the untreated; given the wide range of interventions and outcomes included in this study, we consider the relative scale to be more appropriate because it facilitates direct comparison using a single measure. We did not synthesize results for continuous outcomes because very few studies measured the same continuous outcome. To generate forest plots, we converted estimates on the additive scale to the relative scale by dividing by the mean of the outcome among individuals not receiving treatment. In plots comparing estimates across studies, we presented 95% confidence intervals for the studies for which standard errors were reported or could be estimated on the relative scale. When possible, we used adjusted effect measures in these plots because many of the included studies used observational designs or used randomized designs that conditioned on a non-randomized variable (e.g. eligibility status) to measure spillovers. Thus, we consider adjusted estimates more appropriate because they are less likely to be biased than crude estimates. We excluded studies of low or very low quality from plots comparing results across studies.

Risk of bias across studies

To assess publication bias, we produced funnel plots. We produced separate plots for studies estimating risk ratios (or 1-RR) and risk differences because insufficient information was reported to standardize measures on a single scale. Funnel plots only included studies that estimated effects for binary outcomes. We did not produce funnel plots for estimates using continuous outcomes because the number of different outcomes measured would not have allowed for comparison across a useful number of studies.

Additional analyses

We searched each included text for terms commonly used to describe spillovers and noted whether the terms appeared in each text.

Results

Study selection

We retrieved 49 749 records through our search process (Supplement 7 Figure 1, available as Supplementary data at *IJE* online). Following removal of duplicate records and records from non-bibliographical sources, we reviewed 31 622 titles for relevance. We reviewed relevant abstracts and full texts and classified 28 studies from the original search as eligible. We performed title, abstract and full-text review on the reference lists of the 28 eligible texts ($n = 798$ records) and identified one additional eligible text. We also reviewed records that cited the 28 original included texts ($n = 1622$ records) and identified an additional 25 eligible texts. A total of 54 records were included in this systematic review. Reasons for exclusion of full texts are listed in Supplement 5, available as Supplementary data at *IJE* online. We extracted data from 51 studies. We could not extract data for two studies that only reported spillover effects graphically[33,34] or for one study that did not provide numerical results for spillover estimates.[24]

Study characteristics

Studies were conducted in 17 low- or middle-income countries. The most common study design was cluster-randomized trials ($n = 13$ studies; 24%) followed by re-analyses of cluster-randomized trials ($n = 9$; 17%) and re-analyses of individually randomized trials ($n = 7$; 13%). The most common interventions were vaccines ($n = 22$; 41%), mass drug administration for infectious disease control ($n = 7$; 13%) and health education ($n = 5$; 9%). Several programmes were commonly evaluated for spillovers: the maternal and child health programme in Matlab, Bangladesh;[35,36] the PROGRESA programme, which offered conditional cash transfers in Mexico;[37,38] cholera vaccines provided in Matlab Bangladesh;[39–44] and the Primary School Deworming Program in Busia, Kenya.[2,45–47] Studies estimated a variety of different statistical parameters to quantify spillovers; we define these parameters in Supplement 8, available as Supplementary data at *IJE* online.

Risk of bias within studies

Six studies (11%) had high quality evidence, 30 (56%) had moderate quality, 12 (22%) had low quality and six (11%) had very low quality evidence (Supplement 6, available as Supplementary data at *IJE* online). Of studies with high quality evidence, five used cluster-randomized[43,48–51] designs and one used a household secondary attack rate study design.[52] The proportion of studies with low or very low quality evidence was similar in studies that incorporated spillover measurement into the original design (35%) compared with those which did not pre-specify spillover estimation (36%). All high quality studies were peer-reviewed.

Spillover mechanisms

We identified three primary types of spillover mechanisms in included studies.

i. Reduced disease transmission ($n = 28$ studies): interventions may decrease the infectiousness of an intervention recipient, and in turn, the risk that non-recipients become ill may decrease.
ii. Social proximity ($n = 20$): interventions may create spillovers when individuals change their behaviour as a result of intervention and in turn influence the behaviour of non-recipients with whom they are in social proximity. Family members, neighbours, classmates or even residents of the same village or city could be considered socially proximate with varying degrees of closeness.
iii. Substitution ($n = 3$): when one household member receives additional resources as a result of intervention, spillovers may occur to other household members because additional resources are available to the household. For example, if one child receives free meals at school, more food may be available for siblings to eat at home.

Results by spillover mechanism

In this section, we summarize studies by spillover mechanisms because they influence spillover magnitude, scale of spillovers, and appropriate study designs for detecting spillovers. Within each mechanism we summarize studies by intervention type. Within each of these categories, we describe results by intervention types. We excluded very low quality studies ($n = 6$) from this summary.

Spillovers through reduced disease transmission

Studies of spillovers through reduced disease transmission included studies of vaccines ($n = 21$), mass drug administration to control infectious disease ($n = 6$), improved water and sanitation ($n = 3$) and insecticide treated nets ($n = 1$) (Tables 1–2; Supplement 7 Table 1, available as Supplementary data at *IJE* online). These studies evaluated spillovers through two approaches: analyses with group-level data (i.e. ecological analyses) and with individual-level data.

Eleven of the 13 studies that evaluated spillovers using group-level data found that the risk of illness declined as treatment coverage increased, suggesting that spillovers were present. Six studies evaluated spillovers of the cholera vaccine and found that cholera risk decreased among unvaccinated individuals as vaccine coverage increased (Supplement 7 Table 1 and Figure 2 Panel A).[39–42,44,53–60] No such pattern was evident among vaccinated individuals, suggesting that spillover effects did not yield additional protection beyond that conferred by the vaccine itself (Supplement 7 Figure 2 Panel B).[39,40,53,54] Five of these studies re-analysed data from the same trial, so their findings cannot be considered independent.[39–43] There are two significant limitations to this type of analysis in assessing spillovers. First, in observational studies or randomized trials without perfect compliance, this type of ecological comparison is likely to be confounded by factors associated with both treatment compliance and the outcome. For example, vaccination coverage may have been higher in high-income areas with better access to care, which may have partially explained lower illness levels in these areas. Second, spillover findings are likely to be highly sensitive to the definition of the area in which treatment coverage and outcomes were measured; groups of different sizes or composition may have produced different results.[61] Thus, overall, we consider these findings to be of lower quality than findings from studies analysing individual-level data.

Seventeen studies evaluated spillovers through reduced disease transmission using individual-level data (Tables 1–2). We separated these studies into two categories: those that measured spillovers within clusters (e.g. households, villages), and those that measured spillovers as a function of distance from treated individuals. Among the studies measuring spillovers in clusters, we expected that spillovers would be larger in smaller-sized clusters (e.g. households) because reductions in disease transmission are most likely to impact on individuals in close proximity. In general, we did not find this to be the case. Two out of four studies of spillovers in households found relatively large spillover effects;[52,62] both studies estimated the reduction in risk associated with living in households with individuals diagnosed with pertussis who were vaccinated versus unvaccinated for pertussis (the vaccine efficacy for infectiousness); the study in Senegal estimated an 85% [95% confidence interval (CI) 46%, 95%] risk reduction,[52] and the study in

Table 1. Spillover estimates from studies that estimated spillovers through reduced transmission using individual-level data measured within clusters

Reference / country / quality of evidence[a]	Parameter type (in bold) and parameter description	Intervention	Cluster size and treatment coverage[b]	Outcome / subgroup / time point	Estimate (95% CI)
Ali *et al.*, 2013 Country: India Quality: high	**Cluster-level spillover effect** Investigators compared cholera incidence per 100 000 person-days among unvaccinated individuals in clusters where the cholera vaccine was delivered to the rate in control clusters.	Cholera vaccine	Size: small group of households Coverage: 36%	Cholera	(1-RR) x 100%: 0% (−59%, 37%)
Baptista *et al.*, 2006 Country: Brazil Quality: high	**Vaccine efficacy for infectiousness**[c] Investigators compared rates of unvaccinated individuals in households with and without vaccinated, infected individuals.	Pertussis vaccine	Size: household Coverage: 31% received the vaccine in the past 10 years	Pertussis	(1-RR) x 100%: 61.6% (12.8%, 83.1%)
	Vaccine efficacy against illness Investigators compared rates of unvaccinated individuals with rates among vaccinated individuals in households with an infected individual.			Pertussis	(1-RR) x 100%: 12.5% (−5.3%, 27.3%)
Préziosi and Halloran, 2003 Country: Senegal Quality: high	**Vaccine efficacy for infectiousness** Investigators compared rates of unvaccinated individuals in households with and without vaccinated, infected individuals.	Pertussis vaccine	Size: household Coverage: 72% of infected index cases	Pertussis	(1-RR) x 100%: 85% (46%, 95%)
Ozier, 2014 Country: Kenya Quality: moderate	**Cluster-level spillover effect** Investigators compared outcomes of younger siblings of children who received school-based deworming with younger siblings of children who were not assigned to receive deworming. The authors presented other measures of cognitive function as well; we present a subset of these outcomes here.	School-based deworming	Size: household Coverage: not applicable	Raven's matrices score (measure of intelligence) Height Height-for-age Stunting (height-for-age z-score < -2)	Mean difference: 0.220 (0.067, 0.373) Mean difference: 0.204 (−0.378, 0.786) Mean difference: 0.029 (−0.057, 0.115) Risk difference: 0.007 (−0.024, 0.038)
Roca *et al.*, 2011 Country: The Gambia Quality: moderate	**Spillover effect conditional on exposure to treatment before and after treatment** In a population where the pneumococcal conjugate vaccine was not previously offered during routine immunization, investigators compared vaccine-type pneumococcus rates among unvaccinated individuals in villages randomized to receive partial vaccination coverage at baseline and follow-up. The authors presented results over a number of ages and time points. We only present the results of the final follow-up survey for vaccine-type pneumococcus (22 months post-vaccination).	Pneumococcal conjugate vaccine	Size: village (80–660 inhabitants) Coverage: 5–9%	Pneumococcal nasopharyngeal carriage among children 2 to <5 years Pneumococcal nasopharyngeal carriage among children 5 to < 15 years Pneumococcal nasopharyngeal carriage among children 15 years or older	Odds ratio: 0.28 (0.11, 0.70) Odds ratio: 0.25 (0.14, 0.46) Odds ratio: 0.43 (0.17, 1.10)

(continued)

Table 1. Continued

Reference / country / quality of evidence[a]	Parameter type (in bold) and parameter description	Intervention	Cluster size and treatment coverage[b]	Outcome / subgroup / time point	Estimate (95% CI)
	We consider vaccine-type pneumococcus most likely to be influenced by vaccination.				
Roca *et al.*, 2013 Country: The Gambia Quality: moderate	**Spillover effect conditional on exposure to treatment before and after treatment** In a population where the pneumococcal conjugate vaccine was not previously offered during routine immunization, investigators compared vaccine-type pneumococcus rates among unvaccinated individuals in villages randomized to receive partial vaccination coverage at baseline and at 4 years follow-up. The authors presented results over a number of ages and time points. We only present the results of the final follow-up survey for vaccine-type pneumococcus (22 months post-vaccination). We consider vaccine-type pneumococcus most likely to be influenced by vaccination.	Pneumococcal conjugate vaccine	Size: village (80–660 inhabitants) Coverage: 5–9%	Pneumococcal nasopharyngeal carriage among children 2.5 to < 5 years Pneumococcal nasopharyngeal carriage among children 5 to < 15 years Pneumococcal nasopharyngeal carriage among children 15 years or older	Odds ratio: 0.15 (0.07, 0.33) Odds ratio: 0.21 (0.10, 0.42) Odds ratio: 0.02 (0.003, 0.18)
Khan *et al.*, 2012 Country: Pakistan Quality: moderate	**Cluster-level spillover effect** Investigators compared typhoid among the unvaccinated in vaccinated vs control clusters.	Typhoid vaccine	Size: group with an average of 433 children Coverage: 38%	Typhoid incidence	(1-RR) x 100%: −10% (−116, 44)
Sur *et al.*, 2009 Country: India Quality: moderate	**Cluster-level spillover effect** Investigators compared typhoid among the unvaccinated in vaccinated vs control clusters.	Typhoid vaccine	Size: group with an average of 777 individuals Coverage: 60%	Typhoid incidence	(1-RR) x 100%: 44% (2%, 69%)
Miguel and Kremer, 2004 Country: Kenya Quality: moderate	**Cluster-level spillover effect** Investigators compared infection rates among students who did not receive deworming but attended schools with the deworming programme, with children in schools without the programme.[d]	School-based deworming	Size: schools with an average of ∼ 400 pupils Coverage: ∼ 70–80%	Moderate-heavy helminth infection	Risk difference: −0.18 (−0.32,−0.04)
Hammitt *et al.*, 2014 Country: Kenya Quality: moderate	**Spillover before and after treatment** Investigators compared vaccine-type pneumococcus among unvaccinated children (≥ 5 years of age) before and after a campaign was initiated.	Pneumococcal conjugate vaccine	Size: population of 260 000 (no clusters) Coverage: 79%	Pneumococcal nasopharyngeal carriage among children ≥ 5 years	Prevalence ratio: 0.34 (0.18, 0.62)
House *et al.*, 2009 Country: Ethiopia Quality: high	**Cluster-level spillover effect among ineligibles** We compared the prevalence of trachoma among untreated individuals in clusters randomly allocated to treatment 12 months after mass treatment with that in control clusters.	Mass azithromycin distribution	Size: administrative unit with ∼ 1400 people Coverage: 82%	Trachoma	(1-RR) x 100%: 35% (8%, 55%)[e]

(continued)

Table 1. Continued

Reference / country / quality of evidence[a]	Parameter type (in bold) and parameter description	Intervention	Cluster size and treatment coverage[b]	Outcome / subgroup / time point	Estimate (95% CI)
Chidambaram *et al.*, 2004 Country: Ethiopia Quality: low	**Cluster-level spillover effect** Investigators compared the odds of trachoma among ineligible individuals in programme areas with the odds in control areas.	Mass azithromycin distribution	Size: village Coverage: 91%	Trachoma	Odds ratio: 2.9 (1.1, 7.5)
Egere *et al.*, 2012 Country: The Gambia Quality: high	**Cluster-level spillover effect among ineligibles** Investigators compared pneumococcal carriage among infants too young to be vaccinated in fully vs partially vaccinated villages.	Pneumococcal conjugate vaccine	Size: village (80–660 inhabitants- Coverage: 100%	Pneumococcal carriage	Hazard ratio: 0.39 (0.26, 0.58).

[a]The quality of evidence reported here applies to each study as a whole even if multiple types of spillovers were estimated.
[b]We estimated approximate cluster-treatment coverage using available information in each paper.
[c]The manuscript labels this parameter vaccine efficacy against transmission; however, we refer to it as vaccine efficacy for infectiousness based on the definition in Halloran E, Longini IM Jr, Struchiner CJ. *Design and Analysis of Vaccine Studies*. New York, NY: Springer, 2010.
[d]We used estimates from the replication study published in Aiken AM, Davey C, Hargreaves JR, Hayes RJ. Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a pure replication. *Int J Epidemiol* 2015;**44**:1572–80.
[e]Confidence intervals present a best-case scenario as they are not necessarily adjusted for clustering.

Brazil estimated a 61.6% (95% CI 12.8%, 83.1%) risk reduction.[62] Three of the eight studies measuring spillovers in larger clusters (e.g. schools, villages) found evidence of large spillovers.[51,63,64] We expected that spillovers would be larger at higher levels of treatment coverage, and we found this to be true in the relevant studies: of the four studies with cluster-level treatment coverage under 50%, two found no spillover effects.[53,65] Of those with treatment coverage over 50%, all found evidence of spillovers, and in four studies, spillover effects were relatively large.[51,52,63,64] Among studies that measured spillovers as a function of distance from treated individuals, the magnitude of spillovers was smaller than in studies evaluating spillovers within clusters. However, this finding may be explained by intervention type—none of these studies evaluated vaccines. Spillovers decayed with distance from treated individuals in two studies.[2,48]

Spillovers through social proximity

Seventeen studies evaluated spillovers through social proximity (Table 3). One found evidence of negative spillovers,[46] eight found no evidence of spillovers[66–73] and eight found evidence of spillovers for some but not all outcomes or conditions reported.[21,37,38,50,74–77] These studies measured spillovers through four mechanisms among untreated individuals who were: (i) in areas where cash transfers were offered; (ii) in or near areas where subsidies or microloans were offered to promote certain health products or behaviours (e.g. subsidies for vaccines); (iii) socially connected to treated individuals; or (iv) in the same schools or areas as treated individuals, regardless of social links.

We hypothesized that spillovers would be stronger for interventions involving incentives or cash transfers than for those that did not because intervention uptake might be higher and the intervention might receive more attention from untreated individuals than interventions with no transfers or incentives. We also hypothesized that studies considering spillovers through social proximity might be more likely to detect spillovers if they considered social connections between treated and untreated individuals. However, we found neither of these to be true among the studies in this review. Three of the studies measuring spillovers of cash transfers found no evidence of spillovers,[66–68] and two found evidence for some but not all outcomes measured.[37,38] Even among the outcomes for which there was evidence of spillovers, the effect sizes were small. Two of four studies evaluating spillovers of subsidies or microloans for health products found evidence of spillovers.[21,50,69,77] For example, in a study of incentives for immunization, Banerjee *et al.* estimated both total and spillover effects. The relative risk for the total effect on

Table 2. Spillover estimates from studies that estimated spillovers through reduced transmission using individual-level data and that measured spillovers as a function of distance to treated individuals

Reference / country / quality of evidence[a]	Parameter type (in bold) and parameter description	Intervention	Distance / area over which spillovers were measured	Outcome / subgroup / time point	Estimate (95% CI)
Perez-Heydrich *et al.*, 2014 Country: Bangladesh Quality: high	**Spillover effect conditional on treatment density** Investigators estimated the reduction in cholera risk cases per 1000 persons associated with varying levels of vaccination coverage using a counterfactual model. We report the maximum difference, which is the only quantitative estimate reported in the paper. This is the difference in risk between unvaccinated individuals in neighbourhoods with 60% vaccination coverage compared with those in neighbourhoods with 32% coverage.	Cholera vaccine	Neighbourhood of 64 individuals	Cholera risk per 1000 persons	Risk difference: 5.29 (2.61, 7.96)
Hawley *et al.*, 2003 Country: Kenya Quality: high	**Spillover effect conditional on distance among individuals** The authors conducted a test of trend by comparing individuals who did not receive ITNs who lived ≥ 900 m from compounds that did receive ITNs (the reference) with individuals who lived at decreasing distances from compounds that received ITNs.	Insecticide-treated nets	0–900 m	Clinical malaria	Odds ratio: 0.92 (0.75, 1.12)
				High-density parasitaemia	Odds ratio: 0.89 (0.78, 1.01)
				Moderate anaemia	Odds ratio: 0.78 (0.69, 0.89)
				Haemoglobin level	Mean difference for one unit increase: 0.18 (0.06, 0.31)
				Child mortality	Hazard ratio: 0.94 (0.90, 0.98)
Ziegelhöfer *et al.*, 2012 Country: Guinea Quality: moderate	**Spillover effect conditional on treatment density** Investigators estimated the association between diarrhoea and the fraction of people within 3 km who received water supply points.	Improved water supply	3 km	Diarrhoea	Mean difference: −0.59 (−1.32, 0.14) [c]
Miguel and Kremer, 2004 Country: Kenya Quality: moderate	**Spillover effect conditional on treatment density** Investigators estimated the association between the proportion of schoolchildren within 0–3 km and 3–6 km receiving deworming medication at schools and worm infections among untreated children. The effects reported here are for a one-unit increase in the number of children treated, controlling for the number of schoolchildren within a specific distance.[d]	School-based deworming	0–3 km	Moderate-heavy helminth infection (association with treatment density within 0–3 km)	Mean difference: −0.21 (−0.41, −0.01)
			3–6 km	Moderate-heavy helminth infection (association with treatment density within 3–6 km)	Mean difference: −0.05 (−0.21, 0.11)

(continued)

Table 2. Continued

Reference / country / quality of evidence[a]	Parameter type (in bold) and parameter description	Intervention	Distance / area over which spillovers were measured	Outcome / subgroup / time point	Estimate (95% CI)
Baird *et al.*, 2013 Country: Kenya Quality: low	**Spillover effect conditional on treatment density** Investigators estimated the association between the density of deworming treatment within 6 km of each school during childhood and self-reported outcomes in adulthood, 10 years after a school-based deworming programme.	School-based deworming	6 km	Self-reported health is 'very good' Height Body mass index Number of pregnancies Any miscarriages	Probability difference: 0.128 (−0.097, 0.353) Mean difference: −1.891 (−5.158, 1.376) Mean difference: 0.317 (−0.210, 0.844) Mean difference: −0.335 (−0.958, 0.288) Probability difference: −0.078 (−0.151, −0.005)
Shekhawat *et al.*, 2014 Country: Tanzania Quality: low	**Cluster-level spillover effect**[e] Investigators compared trachoma prevalence in 4 birth cohorts who received multiple rounds of mass azithromycin distribution. To measure spillovers, they assessed whether the prevalence of trachoma was lower among the youngest children in later birth cohorts before their first round of mass azithromycin distribution.	Mass azithromycin distribution	Not applicable	Trachoma	No quantitative estimates were reported, but they concluded that spillovers were present
Paul *et al.*, 1962 Country: Costa Rica Quality: low	**Vaccine efficacy** Investigators compared the rate of poliovirus vaccine strain excretion over time in the unvaccinated family contacts of children participating in a vaccine trial.	Polio vaccine	Not applicable	Poliovirus	No quantitative estimates were reported, but they found that the rate of excretion among household contacts increased following vaccination

[a]The quality of evidence reported here applies to each study as a whole even if multiple types of spillovers were estimated.

[b]We estimated approximate cluster-treatment coverage using available information in each paper.

[c]Confidence intervals present a best-case scenario as they are not necessarily adjusted for clustering.

[d]We report findings from a replication study of the original study, which revised estimates after correcting for coding errors in the original study. Aiken AM, Davey C, Hargreaves JR, Hayes RJ. Re-analysis of health and educational impacts of a school-based deworming programme in western Kenya: a pure replication. *Int J Epidemiol* 2015;**44**:1572–80.

[e]This parameter was not explicitly estimated, but it could have been using the data collected in the study.

Table 3. Spillover estimates from studies that estimated spillovers through social proximity

Reference / country / quality of evidence[a] / scale at which spillovers were measured	Parameter type (in bold) and parameter description	Intervention	Outcome / subgroup / time point	Estimate (95% CI)
Spillovers of cash transfer interventions				
Avitabile, 2012 Country: Mexico Quality: low Scale: village	**Cluster-level spillover effect among ineligibles** Investigators compared the difference in the difference (DID) in outcomes before and during the programme among ineligible individuals in clusters where cash transfers were offered to those of ineligibles in the control clusters.	Conditional cash transfers	Cervical cancer screening Blood sugar screening Blood pressure screening	Mean difference: 0.061 (0.022, 0.100) Mean difference: 0.010 (−0.025, 0.045) Mean difference: 0.025 (−0.010, 0.060)
Handa *et al.*, 2001 Country: Mexico Quality: moderate Scale: localities	**Cluster-level spillover effect among ineligibles** Investigators compared the difference in the difference (DID) in outcomes before and during the programme among ineligible individuals in clusters where cash transfers were offered to those of ineligibles in clusters without cash transfers.	Conditional cash transfers	Child nutrition surveillance 6 months after programme initiation Child nutrition surveillance 12 months after programme initiation	Mean difference: 2.307 (0.817, 3.797) Mean difference: 6.846 (2.632, 11.060)
Ribas *et al.*, 2011 Country: Paraguay Quality: moderate Scale: district	**Cluster-level spillover effect among ineligibles** Investigators compared outcomes among ineligible individuals in clusters where cash transfers were offered to those of ineligibles in clusters without cash transfers.	Conditional cash transfers	Child growth monitoring visits	Mean difference: −0.014 (−0.169, 0.141)
Baird *et al.*, 2013 Country: Malawi Quality: very low Scale: ∼250 households	**Cluster-level spillover effect** Investigators compared psychological distress among girls who did not receive cash transfers in areas where cash transfers were offered to those in comparison areas. They also stratified by whether girls who did not receive cash transfers girls lived in a household with another treated girl.	Conditional and unconditional cash transfers	Psychological distress among all untreated girls during the intervention Psychological distress among all untreated girls after the intervention Psychological distress among untreated girls in households without treated girls during the intervention Psychological distress among untreated girls in households without treated girls after the intervention Psychological distress among untreated girls in households with treated girls during the intervention	Mean difference: 0.064 (0.007, 0.121) Mean difference: 0.007 (−0.056, 0.070) Mean difference: 0.099 (0.038, 0.160) Mean difference: 0.001 (−0.058, 0.060) Mean difference: −0.086 (−0.188, 0.016)

(continued)

Table 3. Continued

Reference / country / quality of evidence[a] / scale at which spillovers were measured	Parameter type (in bold) and parameter description	Intervention	Outcome / subgroup / time point	Estimate (95% CI)
			Psychological distress among untreated girls in households with treated girls after the intervention	Mean difference: 0.015 (−0.142, 0.172)
Contreras and Maitra, 2013[b] Country: Columbia Quality: very low Scale: household	**Cluster-level spillover effect among ineligibles** Investigators compared the difference in the difference (DID) in outcomes before and during the programme among ineligible individuals in households where cash transfers were offered to those of ineligibles in households without cash transfers. The authors presented results stratified by age and gender, but here we only present pooled results.	Conditional cash transfers	Self-reported to be ill at 1 year Self-reported to be ill at 4 years In bed as a result of illness at 1 year In bed as a result of illness at 4 years Hospitalized in the previous year at 1 year Hospitalized in the previous year at 4 years	Mean difference: −0.030 (−0.060, 0.000) Mean difference: −0.018 (−0.055, 0.019) Mean difference: 0.017 (−0.046, 0.080) Mean difference: 0.042 (−0.028, 0.112) Mean difference: −0.010 (−0.020, 0.009) Mean difference: −0.016 (−0.031, −0.002)
Spillovers of interventions with subsidies or microloans				
Bhattacharya *et al.*, 2013 Country: Kenya Quality: low Scale: neighbourhood	**Total effect conditional on treatment density** Investigators estimated the association between the proportion of nearby households (within 250 m, 500 m and 1000 m) receiving subsidies and the probability of ITN use at different income thresholds of eligibility for the subsidy. The authors presented results over a large number of subsidy thresholds; we excluded some to avoid redundancy.	Subsidized insecticide-treated nets	Probability of ITN use when < 50% eligible for the subsidy Probability of ITN use when ≥ 50% eligible for the subsidy	Up to a 5% decrease At most thresholds, the 95% confidence intervals did not span 0%. Up to 4.8% increase All confidence intervals did not include 0%.
Banerjee *et al.*, 2010 Country: India Quality: high Scale: 6 km between villages	**Spillover effect conditional on distance to clusters** Investigators compared immunization rates in villages within 6 km of villages randomized to either an immunization campaign or an immunization campaign with incentives, with rates in villages randomized to the control group.	Immunization campaign without incentives Immunization campaign with incentives Immunization campaign without incentives Immunization campaign with incentives Immunization campaign without incentives	Number of immunizations Number of immunizations Child received ≥ 1 immunization Child received ≥ 1 immunization Child has BCG scar	Relative risk: 1.18 (0.92, 1.43) Relative risk: 1.48 (1.18, 1.77) Relative risk: 1.00 (0.80, 1.19) Relative risk: 1.05 (0.86, 1.24) Relative risk: 1.00 (0.73, 1.28)

(continued)

Table 3. Continued

Reference / country / quality of evidence[a] / scale at which spillovers were measured	Parameter type (in bold) and parameter description	Intervention	Outcome / subgroup / time point	Estimate (95% CI)
		Immunization campaign with incentives	Child has BCG scar	Relative risk: 1.05 (0.78, 1.32)
		Immunization campaign without incentives	Child was completely immunized	Relative risk: 1.83 (0.93, 2.73)
		Immunization campaign with incentives	Child was completely immunized	Relative risk: 3.47 (2.18, 4.77)
Tontarawongsa *et al.*, 2011 Country: India Quality: moderate Scale: village	**Spillover effect conditional on number of social network links** Investigators measured whether insecticide-treated net (ITN) acquisition and use among untreated individuals in areas where free ITNs were offered was associated with the percentage of respondents' social ties in a programme offering free and subsidized ITNs. The authors presented results stratified by the type of social links, but we do not present those results here.	Free insecticide-treated nets	Recently acquired at least one ITN	Mean difference: −0.008 (−0.141, 0.125)
			Fraction of household members slept under ITN last night	Mean difference: 0.071 (−0.011, 0.153)
			Recently acquired at least one bed net	Mean difference: 0.183 (−0.091, 0.457)
			Fraction of household members slept under bed net last night	Mean difference: 0.157 (−0.014, 0.328)
	Spillover effect conditional on social network links' ITN acquisition and use Investigators measured whether insecticide-treated net (ITN) acquisition and use among untreated individuals was associated with the acquisition and use among social ties in a programme offering free and subsidized ITNs.	Average per capita bed nets owned by peers	Recently acquired at least one ITN	Mean difference: −0.050 (−0.320, 0.220)
			Fraction of household members slept under ITN last night	Mean difference: 0.020 (−0.145, 0.185)
			Recently acquired at least one bed net	Mean difference: 0.042 (−0.350, 0.434)
			Fraction of household members slept under bed net last night	Mean difference: 0.056 (−0.216, 0.328)
		Average ITN usage the previous night among peers	Recently acquired at least one ITN	Mean difference: −0.107 (−0.283, 0.069)
			Fraction of household members slept under ITN last night	Mean difference: 0.007 (−0.115, 0.129)
			Recently acquired at least one bed net	Mean difference: −0.036 (−0.305, 0.233)
			Fraction of household members slept under bed net last night	Mean difference: 0.016 (−0.180, 0.212)
Godlonton and Thornton, 2012 Country: Malawi Quality: moderate Scale: 1 km	**Overall effect conditional on outcome density** Investigators assessed whether a 1% increase in the proportion of neighbours within 0–0.5 km who received HIV test results (regardless of incentives received) was associated with choosing to learn one's HIV test results.	Incentives for voluntary counselling and testing for HIV	Probability of learning HIV test results	Mean difference: 0.106 (0.014, 0.198)

(continued)

Table 3. Continued

Reference / country / quality of evidence[a] / scale at which spillovers were measured	Parameter type (in bold) and parameter description	Intervention	Outcome / subgroup / time point	Estimate (95% CI)
	Investigators assessed whether a 1% increase in the proportion of neighbours within 0–0.5 km who received incentives for learning their HIV test results was associated with choosing to learn one's HIV test results. The authors stratified results by gender, the proportion of neighbours over increasing distances, distance to HIV testing centres and other variables. See the paper for the full set of results.		Probability of learning HIV test results	Mean difference: −0.064 (−0.227, 0.099)
Spillovers through social connections to treated individuals				
Chong *et al.*, 2013 Country: Colombia Quality: moderate Scale: classroom / school	**Cluster-level spillover effect** Investigators compared outcomes among students in classrooms that did not receive the education programme but were in schools where other classrooms received it, with outcomes in schools where the programme was not offered. The authors presented other outcomes as well. See the paper for the full set of results.	Online sexual health education	Knowledge index at 1 week	Mean difference: 0.015 (−0.073, 0.103)
			Knowledge index at 6 months	Mean difference: 0.013 (−0.148, 0.174)
			Attitude index at 1 week	Mean difference: 0.026 (−0.066, 0.118)
			Attitude index at 6 months	Mean difference: 0.023 (−0.077, 0.123)
			Condom voucher redemption at 6 months	Mean difference: 0.040 (−0.031, 0.111)
	Spillover effect conditional on treatment density Investigators estimated the difference in outcomes among students who did not participate in the programme if all vs none of their friends participated in the programme. The authors presented other outcomes as well. See the paper for the full set of results.		Knowledge index at 6 months	Mean difference: −0.074 (−0.480, 0.332)
			Attitude index at 6 months	Mean difference: −0.046 (−0.326, 0.234)
			Condom voucher redemption at 6 months	Mean difference: −0.156 (−0.274, −0.038)
German *et al.*, 2012 Country: Thailand Quality: moderate Scale: one social network node	**Spillover effect among social network members** Investigators estimated the mean depression score for peers of individuals randomized to treatment intervention participants with that of peers of individuals randomized to control.	Peer support intervention	Depression score of peers of intervention recipients	Mean difference: −0.095 (−0.18, −0.01)
			Depression score of peers of controls recipients	Mean difference: −0.092 (−0.18, −0.01)
Kremer and Miguel, 2007	**Overall effect conditional on number of social network links**	School-based deworming	Deworming consumption	Probability: −0.031 (−0.058, −0.004)

(continued)

Table 3. Continued

Reference / country / quality of evidence[a] / scale at which spillovers were measured	Parameter type (in bold) and parameter description	Intervention	Outcome / subgroup / time point	Estimate (95% CI)
Country: Kenya Quality: moderate Scale: 3–6 km between schools	Investigators measured whether the number of social links with parents whose children previously received deworming was associated with the probability of taking deworming.			
Spillovers through residence in the same area as treated individuals, regardless of social links				
Janssens *et al.*, 2006 Country: India Quality: low Scale: village	**Cluster-level spillover effect** Investigators compared vaccination rates among non-participants in villages with the programme with rates in the control group.	Women's empowerment programme	Tuberculosis vaccine	Mean difference: 0.149 (0.053, 0.245)
			DTP vaccine	Mean difference: 0.122 (0.024, 0.220)
			Measles vaccine	Mean difference: 0.268 (0.170, 0.366)
	Cluster-level spillover effect in which controls are matched to the untreated Investigators compared vaccination rates among non-participants in villages with the programme, with rates among individuals in the control group who were matched to non-participants in the treatment group.		Tuberculosis vaccine	Mean difference: 0.108 (0.059, 0.157)
			DTP vaccine	Mean difference: 0.090 (0.045, 0.135)
			Measles vaccine	Mean difference: 0.089 (0.042, 0.136)
Dupas 2006 Country: Kenya Quality: moderate Scale: school	**Spillover effect conditional on treatment density** Investigators compared condom use among students in schools with different proportions of students who previously participated in a health education programme. The authors presented other outcomes as well; we excluded some to avoid redundancy.	Proportion of girls who received information about HIV transmission	Probability of using a condom during sex for girls	Mean difference: 0.476 (0.111, 0.841)
			Probability of using a condom during sex for boys	Mean difference: 0.109 (−0.254, 0.472)
		Proportion of boys who received information about HIV transmission	Probability of using a condom during sex for girls	Mean difference: −0.388 (−0.798, −0.022)
			Probability of using a condom during sex for boys	Mean difference: 0.042 (−0.387, 0.471)
Azad *et al.*, 2010 Country: Bangladesh Quality: low	**Cluster-level spillover effect conditional on exposure to treatment** Investigators measured outcomes among women exposed to women's groups who did not participate in	Women's groups and health service strengthening	Neonatal mortality among women exposed to women's groups who did not participate in them (cluster-	Mean difference: −4.8 per 1000 live births [c]

(continued)

Table 3. Continued

Reference / country / quality of evidence[a] / scale at which spillovers were measured	Parameter type (in bold) and parameter description	Intervention	Outcome / subgroup / time point	Estimate (95% CI)
Scale: union (population ~15 000–35 000)	them and women in the same villages where the groups took place who had not heard of them. We compared these rates with those in the control group (the authors did not explicitly measure spillovers). The authors assessed numerous other outcomes as well, such as the stillbirth rate and number of antenatal visits, and there was no evidence of spillovers for these outcomes.		level spillover effect conditional on exposure to treatment) Neonatal mortality among women in the same villages where the groups took place, who had not heard of them (cluster-level spillover effect)	 Mean difference: −4.1 per 1000 live births [c]
Singh 2011 Country: India Quality: moderate Scale: city block	**Cluster-level spillover effect** Investigators compared the difference in weight-for-age z-score among children of mothers in the intervention clusters that did not participate in the intervention between baseline and follow-up, with the difference for among children of mothers in a control group, between baseline and follow-up.	Nutrition education	Weight-for-age z-score	Difference in difference: 0.013 (−0.067, 0.093)
Björkman and Svensson, 2009 Country: Uganda Quality: moderate Scale: ~30 km between villages	**Overall effect conditional on distance** Investigators assessed whether being within 10 km of a treatment clinic was associated with increased health care utilization in control areas. The authors also measured spillovers on child death. See the manuscript appendix for results.	Community monitoring and provision of health services	Rate of outpatient visits Mean deliveries per facility per month	Mean difference: 68.1 (−51.1, 187.3) Mean difference: 2.6 (−5.4, 10.6)

[a]The quality of evidence reported here applies to each study as a whole even if multiple types of spillovers were estimated.
[b]Substitution is another possible mechanism of spillover in this paper.
[c]95% Confidence intervals could not be calculated due to insufficient information in the paper.

complete child immunization was 6.66 (95% CI 4.53, 8.80); for the spillover effect the relative risk was 3.47 (95% CI 2.18, 4.77).[50] Two of the three studies that measured spillovers through social links to treatment recipients found no evidence of spillovers,[70,71] and one found evidence of negative spillovers.[46] Of the five studies measuring spillovers among untreated individuals in the same schools, villages or areas as treated individuals, one found evidence of spillovers,[76] two found evidence for some but not all outcomes[74,75] and two found no evidence of spillovers.[72,73] Because the number of studies measuring spillovers through social proximity is relatively small and the types of interventions and outcomes measured varied widely, it is likely that the patterns we observed in this review do not necessarily generalize to the same interventions implemented in other contexts.

As for studies of spillovers through disease transmission, we also assessed whether spillover presence and effect sizes were associated with the size of the area in which spillovers were measured (e.g. household versus city). We did not find evidence of any patterns associated with area size.

Spillovers through substitution

Four studies measured spillovers through substitution (Table 4).[24,78–80] Three studies measured whether siblings of children participating in school nutrition programmes

Table 4. Spillover estimates from studies that estimated spillovers through substitution

Reference / country / quality of evidence[a]	Parameter type (in bold) and parameter description	Intervention	Outcome	Mean difference (95% CI)
Fitzsimons *et al.*, 2012	**Cluster-level spillover effect among ineligibles**	Information on infant nutrition and health	Height-for-age	−2.66 (−0.540, 0.008)
			Weight-for-age	−0.142 (−0.456, 0.172)
Country: Malawi	Investigators measured several outcomes among older children who were not targeted by the programme who lived in the same households as program beneficiaries.		Weight-for-height	−0.038 (−0.332, 0.256)
			Diarrhoea	0.004 (−0.055, 0.063)
Quality: moderate			Vomiting	−0.042 (−0.134, 0.050)
			Fast breathing	−0.008 (−0.110, 0.094)
			Fever	−0.018 (−0.130, 0.094)
			Chills	−0.033 (−0.170, 0.104)
Kazianga *et al.*, 2014	**Cluster-level spillover effect among ineligibles**	School feeding programme	Weight-for-age	0.031 (−0.230, 0.292)
			Height-for-age	0.094 (−0.218, 0.406)
Country: Burkina Faso	This study estimated the mean difference in the difference (DID) in weight-for-age and height-for-age z-scores between baseline and follow-up. They estimated spillovers among pre-school-aged children who lived in households where school-aged children received a school feeding programme or a take-home rations programme compared with those where school-aged children received neither.	Take-home rations	Weight-for-age	0.445 (0.159, 0.731)
			Height-for-age	0.079 (−0.262, 0.420)
Quality: moderate				
Zivin *et al.*, 2009	**Cluster-level spillover effect**	HIV/AIDS treatment	Weight-for-height	0.374 (−1.163, 1.911)
Country: Kenya	Investigators compared weight-for-height z-scores of children whose parents were HIV-positive and had received more than 100 days of antiretroviral therapy, with those whose parents had received fewer than 100 days of therapy.			
Quality: moderate				
Buttenheim *et al.*, 2011	**Cluster-level spillover effect among ineligibles**	School feeding and take home rations programme	Child growth and anaemia	The authors report that they found evidence of spillovers, but they did not present disaggregated spillover results.
Country: Laos	Investigators compared outcomes of younger and older siblings of children participating in a school feeding and take-home rations programme with those in a control group.			
Quality: low				

[a]The quality of evidence reported here applies to each study as a whole even if multiple types of spillovers were estimated.

or whose mothers participated in nutrition education programmes experienced improved growth as a result. None of these studies found evidence of spillovers.

Results pooled across intervention

Because studies estimated a wide variety of types of spillovers, results from many studies were not directly comparable. However, a sufficient number of studies estimated within-cluster spillovers to allow for comparison of results across interventions. This type of spillover compares outcomes among untreated individuals in clusters with different proportions of treatment; most commonly, the effect compares untreated individuals in clusters in which some proportion of the cluster receives treatment with those in which no one receives treatment. The three interventions for which investigators reported positive spillover estimates were mass administration of azithromycin to control trachoma (35% decrease in trachoma),[49] the typhoid vaccine in India (44% decrease in typhoid)[81] and the pneumococcal conjugate vaccine (70% decrease in vaccine-type pneumococcus) (Figure 1).[51] These spillovers were measured in studies with moderate and high quality. The remaining studies could not distinguish spillover estimates from the null. Within-cluster spillovers were stronger in studies in which the proportion of individuals treated within clusters was higher (Figure 2). The largest spillovers were present for a study evaluating spillovers of the pneumococcal vaccine by comparing outcomes in villages where 100% of individuals were vaccinated, with those in which only infants were vaccinated.[51] Both studies with average cluster-level treatment coverage below 40% did not find evidence of spillovers.[53,65]

Risk of bias across studies

The funnel plots for total and direct effects suggest that publication bias was not present (Supplement 7 Figure 3). The funnel plot for spillover effects estimated with risk differences was balanced around 0, indicating minimal publication bias (Figure 3). For spillovers estimated with risk ratios, the plot was asymmetrical, with few studies producing estimates of negative spillover effects at any level of precision, indicating strong publication bias. Risk ratios were more common in public health studies of interventions that were unlikely to result in negative spillovers (e.g. vaccines). Conversely, risk differences were more common in economics studies of interventions for which the expected direction of spillover effects is less clear.

Additional analysis

We identified 15 terms commonly used to describe the concept of spillovers (Table 5). The most common terms were 'indirect effect' and 'spillover', followed by 'externality/externalities'. 'Indirect protection' and 'herd protection' were other common terms.

Discussion

Summary of evidence

To our knowledge, this is the first systematic review of health-related spillovers of interventions in low- and middle-income countries. Evidence of spillovers was strongest for spillovers through reduced disease transmission, and in particular for vaccines and mass drug

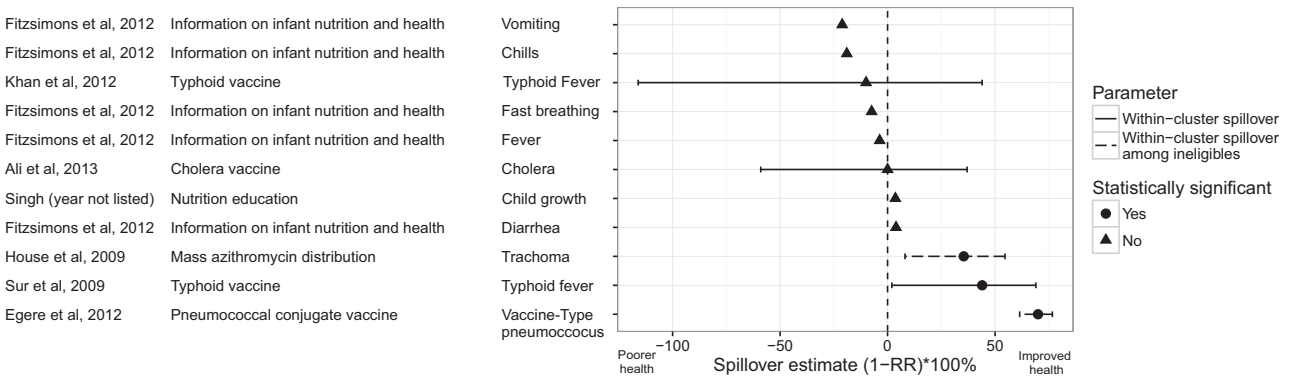

Figure 1. Cluster-level spillover effects. On the x-axis, the cluster-level spillover effect is shown as the % change in outcome among the untreated in the treated cluster from the mean in the control group [i.e., (1-RR) x 100%, where RR is the relative risk]. Outcomes were recoded so that a greater value of the spillover effect indicates an improvement in health (e.g., higher vaccination coverage, lower mortality) and a smaller value indicates poorer health (e.g., lower vaccination coverage, higher mortality). This figure excludes studies of low or very low quality and studies that did not report information that allowed for standardization. Statistical significance was determined based on the measures presented in the paper for the parameter on its original scale. (a) Information required to convert standard errors for risk differences to standard errors for (1-RR) x 100% was not reported, thus 95% confidence intervals are not presented. (b) These studies were conducted in the same country (India) and are subject to dependence.

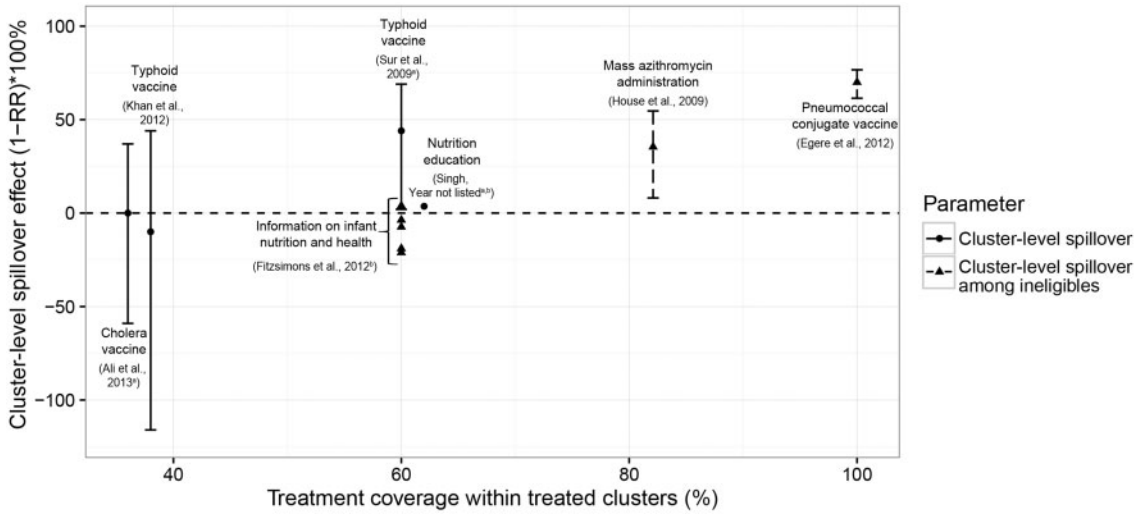

Figure 2. Cluster-level spillover effects by treatment coverage level. This figure plots cluster-level spillover estimates by the level of treatment coverage within treated clusters. We estimated treatment coverage using information available in each paper. On the y-axis, the cluster-level spillover effect is shown as the % change in outcome among the untreated in the treated cluster from the mean in the control group [i.e., (1-RR) x 100%, where RR is the relative risk]. Outcomes were recoded so that a greater value of the spillover effect indicates an improvement in health (e.g., higher vaccination coverage, lower mortality) and a smaller value indicates worse health (e.g., lower vaccination coverage, higher mortality). This figure excludes studies of low or very low quality and studies that did not report information that allowed for standardization. (a) These studies were conducted in the same country (India) and are subject to dependence. (b) Information required to convert standard errors for risk differences to standard errors for (1-RR) x 100% was not reported, thus 95% confidence intervals are not presented.

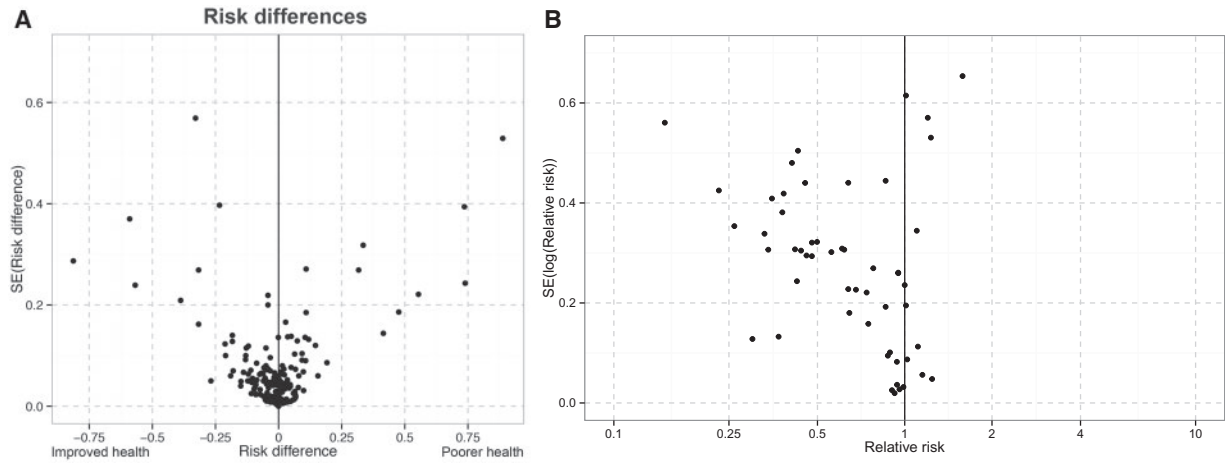

Figure 3. Funnel plots for spillover effects. Panel A: This plot includes spillover estimates from 19 studies that reported risk differences for binary outcomes, of which all but one were from studies in the economics literature. These studies evaluated a wide range of interventions including women's empowerment programs, mass drug administration for infectious disease control, peer group interventions, and nutrition programs. Panel B: This plot includes spillover estimates from 14 studies that reported risk ratios or protective efficacy ((1-RR) x 100%) for binary outcomes, all of which were from studies in the public health literature. These studies evaluated vaccines and mass drug administration for infectious disease control.

administration for infectious disease control. There was also strong evidence of spillovers for insecticide-treated net use on health outcomes, but only one study evaluated this association. In studies of spillovers through social proximity, there was weak evidence of spillovers in most studies with a few exceptions: for example, there was evidence that an immunization campaign with incentives increased immunization coverage among non-participants in nearby villages. There was no evidence of spillovers through substitution effects in the three relevant studies.

There are several reasons why we believe we found the strongest evidence for spillovers through reduced disease transmission. First, spillovers through reduced disease transmission are mostly a function of physical proximity. Infectious disease theory suggests that spillovers occur through reduced disease transmission when susceptible and

Table 5. Search terms related to spillover effects in included texts by academic field[b]

	Economics	Geography	Public health	Total
Indirect effect*[a]	12	2	13	27
Spillover*[a]	23	0	1	24
Externalit*[a]	19	0	0	19
Seconda*[a]	3	3	10	16
Indirect protection	0	4	11	15
Herd protect*[a]	0	2	12	14
Diffusion	7	1	3	11
Herd immunity	1	4	5	10
Herd effect*[a]	0	0	10	10
Peer effect*[a]	9	0	0	9
Unexpected	2	0	3	5
Interference	2	0	2	4
Indirect protective	0	0	4	4
Contagion	3	0	0	3
Unexpected benefit*[a]	0	0	1	1

[a]Asterisks at the end of search terms indicate wild-card characters allowed at the end of the search term. For example, 'externalit*' would retrieve search results for 'externality' and 'externalities'.

[b]Counts allow for multiple terms per included text.

infected individuals come into contact;[82] such contact can modify disease transmission across different populations and pathogens. Indeed, we found evidence of spillovers through reduced disease transmission across interventions, outcomes and populations. On the other hand, spillovers through social proximity may be a function of physical proximity as well as social dynamics that are highly dependent on culture and context, and these factors may vary by population, intervention and health outcome. The relative complexity of spillovers through social proximity may make spillovers less likely to occur through this mechanism than through reduced disease transmission. Similarly, for spillovers through substitution, although an intervention may free up a fixed amount of resources in a household, whether those resources support the health of non-intervention recipients may depend on complex factors, such as education level and culture, which vary across populations.

Second, study designs may have been more appropriate for detecting spillovers through reduced disease transmission than through social proximity or substitution. A rich literature has refined study designs to estimate spillovers of vaccines;[4,5,7,13,15,82–86] these methods can easily be extended to studies of other interventions that produce spillovers through reduced disease transmission. For spillovers through social proximity or substitution, there is no equivalent methodological literature focused on empirical measurement. As a result, in this review, study designs for detecting spillovers through these mechanisms may have been suboptimal or biased. Indeed, the proportion of studies that we classified as moderate, low, or very low quality was greater among studies measuring spillovers through social proximity or substitution than among studies of spillovers through reduced disease transmission.

Finally, it is also possible that rigorous studies to measure spillovers through social proximity or substitution simply have not been conducted yet or were missed in our search. There were five high quality studies of spillovers through reduced disease transmission compared with only one for spillovers through social proximity and none for substitution. Thus, our findings do not necessarily reflect a lack of spillovers of any particular intervention. Rather, with the exception of vaccines and mass drug administration to control infectious disease, our findings show little evidence for health-related spillovers from currently published intervention studies conducted in low- and middle-income countries.

Quality of evidence

Most studies reported moderate or low quality evidence of spillovers, and there were two overarching sources of bias. First, many study designs did not adequately minimize unmeasured confounding of spillover estimates. Only two out of the 23 studies estimating within-cluster spillover effects used double-randomized designs[67,70] which allow for the strongest inference for this type of spillover by minimizing selection bias and unmeasured confounding. In the 21 other such studies, untreated individuals in treated clusters may have been systematically different from individuals in control clusters, possibly because they were not eligible to receive the intervention or chose not to receive the intervention. Such systematic differences between the populations used to measure total effects versus spillover effects could result in biased spillover estimates relative to the estimates that would be obtained in a double-randomized design, in which measured and unmeasured confounders are balanced across both populations.

Second, in 33 out of 54 studies, spillover measurement was not pre-specified, which may have increased the chance that a study's results were biased. We found evidence of publication bias for spillover estimates reported as risk ratios but not for total or direct effects. Pre-specification helps prevent publication bias. Without pre-specification, spillover parameters may be defined in a way that increases the chance of detecting positive spillovers, whether intentionally or not. For example, studies estimating spillovers conditional on treatment density within fixed areas may define areas in a way that increases the magnitude of spillover effects. In addition, when spillover measurement is not pre-specified, investigators may fail to measure spillovers altogether or they may be less likely to report null spillover findings.

Table 6. Reporting checklist for studies estimating spillovers

Section/topic	No.	Checklist item
Title and abstract		
Title and abstract	1	If spillovers were measured as a primary outcome of a study, mention them in the title and/or abstract. Use the term 'spillovers' or 'indirect effects' to refer to spillovers
Introduction		
Background	2	Use the term 'spillovers' or 'indirect effects' to refer to spillovers
Methods		
Study design	3	Indicate whether spillover estimation was pre-specified
	4	Describe whether buffers existed between treatment and control units, whether in physical or social distance
	5	If treatment or outcome density was measured within areas, describe the rationale for and method of defining these areas
	6	Describe the scale on which spillovers are expected (e.g. household, village etc.)
	7	For study designs used to estimate spillovers other than the double-randomized or the cluster-randomized design, provide a clear description of the assumptions required to estimate valid statistical parameters if SUTVA is violated
Participants	8	Provide a clear description of treatment eligibility criteria
	9	State whether individuals enrolled to measure spillovers were eligible for the treatment or not
Interventions	10	Provide a clear description of how treatment was allocated to groups and individuals
	11	Describe whether untreated individuals in treated areas were randomly assigned to not receive treatment, if they opted out of treatment, if they were ineligible for treatment or if there were other reasons they were not treated
	12	State whether the level of treatment allocation was chosen in order to measure spillovers
	13	Describe the mechanism of spillovers hypothesized and assessed for each treatment
	14	Describe whether a buffer zone was created between treatment and control units
Outcomes	15	If outcomes measured to estimate direct, total or overall effects differed from outcomes measured to estimate spillover effects, provide a rationale for the difference
Study size	16	Describe any calculations conducted to determine the sample size needed to estimate spillover parameters. If none, state that none were conducted
Statistical methods	17	Define the specific spillover parameter(s) estimated for each intervention
	18	Describe the statistical analysis methods used to estimate spillover effects
	19	Indicate whether spillovers were estimated among individuals allocated to not receive treatment vs those that chose not to take treatment (i.e. indicate whether the spillover analysis was intention-to-treat)
Results		
Participant flow	20	If using a clustered design to measure spillovers, provide the number of clusters allocated to treatment and control that were included in the assessment of spillovers
	21	If using a clustered design to measure spillovers, provide the number of individuals that received and did not receive treatment within treatment and control clusters
	22	If using a clustered design to measure spillovers conditional on eligibility status, provide the number of individuals eligible to receive treatment in treated clusters and the total number of individuals in treated clusters
	23	If using a clustered design to measure spillovers, provide the number of individuals allocated to treatment within treatment clusters, allocated to not receive treatment within treated clusters, and allocated to control clusters
	24	If using a clustered design to measure spillovers, provide information about the proportion of individuals receiving treatment within each cluster
	25	If measurement occurred in buffer zones between treatment and control clusters, provide the number of individuals who did and did not receive treatment in buffer zones
	26	Describe whether loss to follow-up rates were similar among individuals measured for spillover vs direct/total/overall effects and whether the characteristics of those lost to follow-up for spillover measurement differed from those who were not lost to follow-up
Recruitment	27	If dates of data collection for spillover measures differed from dates for direct, total or overall effect measures, explain the discrepancy

(continued)

Table 6. Continued

Section/topic	No.	Checklist item
Main results	28	Clearly label which results estimate each spillover parameter
	29	If multiple spillover mechanisms were hypothesized, label results according to the hypothesized spillover mechanism
	30	Present direct, total, overall and spillover effects in the same population subgroups to allow for assessment of the proportion of the total and overall effects attributable to spillovers
	31	Report whether there was any evidence that untreated individuals in the treatment or control group were exposed to treatment (e.g. if untreated individuals had heard of the intervention or knew individuals who received it)
	32	Describe any evidence of contamination of the control group
Discussion		
Summary of findings / key results	33	Present theory or evidence supporting the proposed mechanism of spillover.
Limitations	34	Discuss any potential biases that may be present for spillover parameters Discuss whether these biases may also be present for direct or total effect parameters. This includes contamination of the control group
	35	Articulate whether any analyses conducted to estimate spillovers were not pre-specified
Generalizability	36	Comment on external validity of findings and whether any methods used to estimate spillover effects may have compromised external validity (e.g. matching of untreated in the treatment group to untreated in the control group)

SUTVA, stable unit treatment value assignment.

Finally, because spillover effects are likely to be smaller than treatment effects in most cases, studies that do not pre-specify spillover measurement and incorporate them into sample size calculations may be underpowered to detect spillovers. Because spillovers tend to have smaller effect sizes relative to total or overall effects, typically larger sample sizes are required to detect them. As a result, they are more subject to selective reporting than direct effects.

The overall quality of evidence was lower for economics studies than public health studies. Most public health studies evaluated spillovers of vaccines, whereas many economics studies measured spillovers of complex interventions such as conditional cash transfer programmes. Our finding that studies of more complex interventions typically had lower quality ratings is consistent with other studies.[87,88] There are several reasons for this pattern. First, complex, realistic interventions often cannot feasibly or ethically be randomized. As a result, many observational studies, some which employed innovative methods for measuring spillover effects in realistic settings, received lower quality ratings in our adapted GRADE framework.[89] Second, evaluations of complex interventions often cannot blind participants and/or investigators, resulting in lower quality ratings. Third, many public health studies measured spillovers of outcomes directly targeted by an intervention (e.g. the impact of the cholera vaccine on cholera risk). Studies that measured outcomes indirectly affected by the intervention (e.g. the impact of childhood deworming on later miscarriage[47]) were more common in economics and received a lower quality rating due to concerns about indirectness of evidence.[90] Finally, economics and other social science studies have different reporting norms compared with public health studies and do not always report information required to receive a high quality rating, such as whether randomized treatment allocation was concealed.

Reporting recommendations

We found a wide range of terminology used to describe spillovers, a lack of standardization among spillover methods, and poor reporting of spillovers in many studies. Very few studies clearly defined the specific spillover effect estimated, and in many studies insufficient information was available to compare spillover effects with direct effects or with spillover effects in other studies. More standardized, systematic reporting across disciplines, particularly in the social sciences, would increase comparability across studies and allow for more careful assessment of risk of bias in studies.[91] To facilitate such standardization, we propose a checklist specific to reporting of spillover effects, adapted from the CONSORT and STROBE frameworks[92,93] (Table 6). This checklist is focused only on reporting spillover effects and is meant to complement the CONSORT[92] and STROBE[94] checklists. We provide an explanation and examples for each item in the checklist in Supplement 9, available as Supplementary data at *IJE* online. By including

items in the checklist that apply to both randomized and observational studies, our objective is to foster more consistent reporting of spillovers across academic disciplines in future studies.

Limitations

Our search and review process was subject to several limitations. Although we made every effort to conduct a comprehensive search, since the concept of spillovers is poorly indexed, it is possible that we missed relevant articles. Greater consistency in the use of terms that describe spillovers would improve future efforts to identify relevant papers by searching electronic databases. We excluded studies from high-income countries from this review since our focus was on interventions relevant to populations in low- and middle-income countries. This focus was a requirement of our funder. However, there are relevant papers measuring health spillovers from high-income countries, many of which evaluate vaccines.[95–97] Some relevant papers which may have been eligible came to our attention after we completed our search process, so we did not include them.[98–101] In addition, some of the databases we searched (e.g. Google Scholar) do not allow for repeatable searches, so our complete search results cannot be fully replicated. During the review process, some titles and abstracts could only be reviewed by one team member, and duplicate risk of bias assessment was only possible in a subset of studies. It is possible that there was misclassification that would have been prevented by complete duplicate review.

Our synthesis of results was also subject to several limitations. The information needed to convert standard errors from the additive to the relative scale was not available in the included studies, so our comparison of estimates across studies did not take precision into account. Since there were very few studies measuring spillovers of the same intervention, our ability to summarize results by intervention type was limited. For papers on vaccines and mass drug administration for infectious disease control, results from studies included in the review may have been dependent because many studies re-analysed data from the same study populations or from the same country. Evidence of spillovers for these interventions in other populations would strengthen the generalizability of these findings.

Conclusions

This review of spillover effects on health outcomes in low- and middle-income countries found a wide range of terminology used to describe spillovers, a lack of standardization among spillover methods and poor reporting of spillovers in many studies. The strongest evidence for spillover effects was found in studies evaluating vaccines and mass drug administration to control infectious disease. There was little evidence available for other types of interventions, and the quality of evidence was moderate or poor in most studies. Future studies would benefit from incorporation of spillover measurement in the design phase and standardized reporting of spillover estimation methods and spillover findings.

Supplementary Data

Supplementary data are available at *IJE* online.

Funding

Funding for this study was provided by the International Initiative for Impact Evaluation (3ie) (SR4-1084).

Acknowledgements

We are grateful to John Eyers at the International Initiative for Impact Evaluation (3ie) for his thoughtful input on our search strategy.

Conflict of interest: None declared.

References

1. Duflo E, Glennerster R, Kremer M. Using randomization in development economics research: A toolkit. *Handb Dev Econ* 2007;**4**:3895–62.
2. Miguel E, Kremer M. Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities. *Econometrica* 2004;**72**:159–217.
3. Cox DR. *Planning of Experiments*. Oxford, UK: Wiley, 1958.
4. Hudgens MG, Halloran ME. Toward Causal Inference With Interference. *J Am Stat Assoc* 2008;**103**:832–42.
5. Tchetgen EJT, VanderWeele TJ. On causal inference in the presence of interference. *Stat Methods Med Res* 2012;**21**:55–75.
6. Rosenbaum PR. Interference Between Units in Randomized Experiments. *J Am Stat Assoc* 2007;**102**:191–200.
7. VanderWeele TJ, Tchetgen Tchetgen EJ. Effect partitioning under interference in two-stage randomized vaccine trials. *Stat Probab Lett* 2011;**81**:861–69.
8. Hayes R, Alexander ND, Bennett S, Cousens SN. Design and analysis issues in cluster-randomized trials of interventions against infectious diseases. *Stat Methods Med Res* 2000;**9**:95–116.
9. John TJ, Samuel R. Herd immunity and herd effect: new insights and definitions. *Eur J Epidemiol*. 2000;**16**:601–06.
10. Fine PE. Herd immunity: history, theory, practice. *Epidemiol Rev* 1993;**15**:265–302.
11. Fox JP, Elveback L, Scott W, Gatewood L, Ackerman E. Herd immunity: basic concept and relevance to public health immunization practices; 1971. *Am J Epidemiol* 1995;**141**:187–97.

12. Rubin DB. Formal mode of statistical inference for causal effects. *J Stat Plan Inference* 1990;**25**:279–92.
13. Halloran ME, Struchiner CJ. Causal Inference in Infectious Diseases. *Epidemiology* 1995;**6**:142–51.
14. Halloran ME, Haber M, Longini IM Jr, Struchiner CJ. Direct and indirect effects in vaccine efficacy and effectiveness. *Am J Epidemiol* 1991;**133**:323–31.
15. VanderWeele T, Tchetgen Tchetgen E, Halloran M. Components of the Indirect Effect in Vaccine Trials: Identification of Contagion and Infectiousness Effects. *Epidemiology* 2012;**23**:751–61.
16. Dybvig PH, Spatt CS. Adoption externalities as public goods. *J Public Econ* 1983;**20**:231–47.
17. Ozawa S, Mirelman A, Stack ML, Walker DG, Levine OS. Cost-effectiveness and economic benefits of vaccines in low- and middle-income countries: A systematic review. *Vaccine* 2012;**31**:96–108.
18. Edejer TT-T, Aikins M, Black R, Wolfson L, Hutubessy R, Evans DB. Cost effectiveness analysis of strategies for child health in developing countries. *BMJ* 2005;**331**:1177.
19. Campbell MJ, Donner A, Klar N. Developments in cluster randomized trials and Statistics in Medicine. *Stat Med* 2007;**26**:2–19.
20. Donner A, Klar N. *Design and Analysis of Cluster Randomization Trials in Health Research*. Oxford, UK: Wiley, 2010.
21. Bhattacharya D, Dupas P, Kanaya S. *Estimating the Impact of Means-tested Subsidies under Treatment Externalities with Application to Anti-Malarial Bednets*. Cambridge, MA: National Bureau of Economic Research, 2013.
22. Benjamin-Chung J, Abedin J, Berger D *et al*. The identification and measurement of health-related spillovers in impact evaluations: a systematic review.London: International Initiative for Impact Evaluation (3ie), 2015.
23. World Bank. *How We Classify Countries*. 2012. http://data.worldbank.org/about/country-classifications (1 March 2012, date last accessed).
24. Buttenheim A, Alderman H, Friedman J. Impact evaluation of school feeding programmes in Lao People's Democratic Republic. *J Dev Eff* 2011;**3**:520–42.
25. Higgins J, Greene S. *Cochrane Handbook for Systematic Reviews of Interventions*. Lonon: Cochrane Collaboration, 2011.
26. Coalition for Evidence-Based Policy. Checklist For Reviewing a Randomized Controlled Trial of a Social Program or Project, To Assess Whether It Produced Valid Evidence.Washington, DC: Coalition for Evidence-Based Policy, 2010.
27. Effective Practice and Organisation of Care (EPOC) Group. Risk of bias for studies with a separate control group.London: Cochrane Review Group, 2009.
28. Gertler PJ. *Impact Evaluation in Practice*. New York, NY: World Bank Publications, 2011.
29. Lee DS, Lemieux T. Regression Discontinuity Designs in Economics. *J Econ Lit* 2010;**48**:281–355.
30. GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ* 2004;**328**:1490.
31. Guyatt GH, Oxman AD, Vist GE *et al*. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ* 2008;**336**:924–26.
32. Berk R, Freedman D. Statistical assumptions as empirical commitments. In: Collier D, Sekhon J, Stark P (eds). *Statistical Models and Causal Inference*. Cambridge, UK: Cambridge University Press, 2010.
33. Shekhawat N, Mkocha H, Munoz B *et al*. Cohort and Age Effects of Mass Drug Administration on Prevalence of Trachoma: A Longitudinal Study in Rural Tanzania. *Invest Ophthalmol Vis Sci* 2014;**55**:2307–14.
34. Paul JR, Horstmann DM, Riordan JT *et al*. An oral poliovirus vaccine trial in Costa Rica. *Bull World Health Organ* 1962;**26**:311–29.
35. Chaudhuri A. Intra-household spillover effects of a maternal and child health program: evidence from rural Bangladesh. Dissertation. Department of Economics, San Francisco State University, 2005.
36. Joshi S, Schultz TP. Family Planning and Women's and Children's Health: Long-Term Consequences of an Outreach Program in Matlab, Bangladesh. *Demography* 2013;**50**:149–80.
37. Avitabile C. *Spillover Effects in Healthcare Programs: Evidence on Social Norms and Information Sharing*. Washington, DC: Inter-American Development Bank, 2012.
38. Handa S, Huerta M-C, Perez R, Straffon B. *Poverty, Inequality, and Spillover in Mexico's Education, Health, and Nutrition Program*. Washington, DC: International Food Policy Research Institute, 2001.
39. Ali M, Emch M, von Seidlein L *et al*. Herd immunity conferred by killed oral cholera vaccines in Bangladesh: a reanalysis. *Lancet* 2005;**366**:44–49.
40. Emch M, Ali M, Root ED, Yunus M. Spatial and environmental connectivity analysis in a cholera vaccine trial. *Soc Sci Med* 2009;**68**:631–37.
41. Ali M, Emch M, Yunus M *et al*. Vaccine Protection of Bangladeshi Infants and Young Children Against Cholera: Implications for Vaccine Deployment and Person-to-Person Transmission. *Pediatr Infect Dis J* 2008;**27**:33–37.
42. Emch M, Ali M, Park J-K, Yunus M, Sack DA, Clemens JD. Relationship between neighbourhood-level killed oral cholera vaccine coverage and protective efficacy: evidence for herd immunity. *Int J Epidemiol* 2006;**35**:1044–50.
43. Perez-Heydrich C, Hudgens MG, Halloran ME, Clemens JD, Ali M, Emch ME. Assessing effects of cholera vaccination in the presence of interference. *Biometrics* 2014;**70**:731–41.
44. Root ED, Giebultowicz S, Ali M, Yunus M, Emch M. The Role of Vaccine Coverage within Social Networks in Cholera Vaccine Efficacy. *PLoS One*. 2011;**6**:e22971.
45. Ozier O. *Exploiting Externalities to Estimate the Long-term Benefits of Early Childhood Deworming*. Working Paper Series. New York, NY: World Bank, 2014.
46. Kremer M, Miguel E. The illusion of sustainability. *Q J Econ* 2007;**122**:1007–65.
47. Baird S, Hicks JH, Kremer M, Miguel E. Worms at Work: Public Finance Implications of a Child Health Investment. Cambridge, MA: Poverty Action Lab, 2013.
48. Hawley WA, Phillips-Howard PA, Kuile FOT *et al*. Community-Wide Effects of Permethrin-Treated Bed Nets on Child Mortality and Malaria Morbidity in Western Kenya. *Am J Trop Med Hyg* 2003;**68**(**Suppl 4**):121–27.
49. House JI, Ayele B, Porco TC *et al*. Assessment of herd protection against trachoma due to repeated mass antibiotic

distributions: a cluster-randomised trial. *Lancet* 2009;**373**: 1111–18.

50. Banerjee AV, Duflo E, Glennerster R, Kothari D. Improving immunisation coverage in rural India: clustered randomised controlled evaluation of immunisation campaigns with and without incentives. *BMJ* 2010;**340**:c2220.
51. Egere U, Townend J, Roca A *et al*. Indirect effect of 7-valent pneumococcal conjugate vaccine on pneumococcal carriage in newborns in rural Gambia: a randomised controlled trial. *PloS One* 2012;**7**:e49143.
52. Préziosi M-P, Halloran ME. Effects of pertussis vaccination on transmission: vaccine efficacy for infectiousness. *Vaccine* 2003;**21**:1853–61.
53. Ali M, Sur D, You YA *et al*. Herd protection by a bivalent-killed-whole-cell oral cholera vaccine in the slums of Kolkata, India. *Clin Infect Dis* 2013;**56**:1123–31.
54. Khatib AM, Ali M, von Seidlein L *et al*. Effectiveness of an oral cholera vaccine in Zanzibar: findings from a mass vaccination campaign and observational cohort study. *Lancet Infect Dis* 2012;**12**:837–44.
55. Huq A, Yunus M, Sohel SS *et al*. Simple sari cloth filtration of water is sustainable and continues to protect villagers from cholera in Matlab, Bangladesh. *mBio* 2010;**1**: e00034–10.
56. Haile M, Tadesse Z, Gebreselassie S *et al*. The Association between Latrine Use and Trachoma: A Secondary Cohort Analysis from a Randomized Clinical Trial. *Am J Trop Med Hyg* 2013;**89**:717–20.
57. Cooper E, Fitch L. Pertussis: Herd Immunity and Vaccination Coverage in St. Lucia. *Lancet* 1983;**322**:1129–32.
58. Forleo-Neto E, de Oliveira CF, Maluf EMCP *et al*. Decreased point prevalence of Haemophilus influenzae Type b (Hib) oropharyngeal colonization by mass immunization of Brazilian children less than 5 years old with Hib polyribosylribitol phosphate polysaccharide—tetanus toxoid conjugate vaccine in combination with diphtheria-tetanus toxoids—pertussis vaccine. *J Infect Dis* 1999;**180**:1153–58.
59. Chen W-J, Moulton LH, Saha SK, Mahmud AA, Arifeen SE, Baqui AH. Estimation of the herd protection of Haemophilus influenzae type b conjugate vaccine against radiologically confirmed pneumonia in children under 2 years old in Dhaka, Bangladesh. *Vaccine* 2014;**32**:944–48.
60. Root ED, Lucero M, Nohynek H *et al*. Distance to health services affects local-level vaccine efficacy for pneumococcal conjugate vaccine (PCV) among rural Filipino children. *Proc Natl Acad Sci U S A* 2014;**111**:3520–25.
61. Openshaw S. Ecological fallacies and the analysis of areal census data. *Environ Plan A* 1984;**16**:17–31.
62. Baptista PN, Magalhães V, Rodrigues LC, Rocha MW, Pimentel AM. Pertussis vaccine effectiveness in reducing clinical disease, transmissibility and proportion of cases with a positive culture after household exposure in Brazil. *Pediatr Infect Dis J* 2006;**25**:844–46.
63. Chidambaram JD, Melese M, Alemayehu W *et al*. Mass Antibiotic Treatment and Community Protection in Trachoma Control Programs. *Clin Infect Dis* 2004;**39**:e95–e97.
64. Hammitt LL, Akech DO, Morpeth SC *et al*. Population effect of 10-valent pneumococcal conjugate vaccine on nasopharyngeal carriage of Streptococcus pneumoniae and non-typeable Haemophilus influenzae in Kilifi, Kenya: findings from cross-sectional carriage studies. *Lancet Glob Health* 2014;**2**:e397–e405.
65. Khan MI, Soofi SB, Ochiai RL *et al*. Effectiveness of Vi capsular polysaccharide typhoid vaccine among children: A cluster randomized trial in Karachi, Pakistan. *Vaccine* 2012;**30**:5389–95.
66. Ribas RP, Soares FV, Teixeira CG, Silva E, Hirata GI. Externality and behavioural change effects of a non-randomised CCT programme: Heterogeneous impact on the demand for health and education. Working Paper No. 2011-1. Brasilia: International Policy Centre for Inclusive Growth, 2011.
67. Baird S, De Hoop J, Özler B. Income shocks and adolescent mental health. *J Hum Resour* 2013;**48**:370–403.
68. Contreras D, Maitra P. Health spillover effects of a conditional cash transfer program [Internet]. Monash University Working Paper, 2012.
69. Tontarawongsa C, Mahajan A, Tarozzi A. (Limited) diffusion of health-protecting behaviors: evidence from non-beneficiaries of a public health program in Orissa (India). Department of Economics, Duke University, 2011.
70. Chong A, Gonzalez-Navarro M, Karlan D, Valdivia M. *Effectiveness and Spillovers of Online Sex Education: Evidence from a Randomized Evaluation in Colombian Public Schools*. Cambridge, MA: National Bureau of Economic Research, 2013. F
71. German D, Sutcliffe CG, Sirirojn B *et al*. Unanticipated Effect of a Randomized Peer Network Intervention on Depressive Symptoms Among Young Methamphetamine Users in Thailand. *J Community Psychol* 2012;**40**:799–813.
72. Singh P. Spillovers in learning and behavior: Evidence from a nutritional information campaign in urban slums. Munich Personal RePEc Archive 33362, 2011.
73. Björkman M, Svensson J. Power to the people: evidence from a randomized field experiment on community-based monitoring in Uganda. *Q J Econ* 2009;**124**:735–69.
74. Dupas P. Relative risks and the market for sex: Teenagers, sugar daddies and HIV in Kenya. Munich Personal RePEc Archive 248, 2005
75. Azad K, Barnett S, Banerjee B *et al*. Effect of scaling up women's groups on birth outcomes in three rural districts in Bangladesh: a cluster-randomised controlled trial. *Lancet* 2010;**375**:1193–202.
76. Janssens W. Measuring externalities in program evaluation. Discussion paper. Tinbergen Institute, 2005.
77. Godlonton S, Thornton R. Peer effects in learning HIV results. *J Dev Econ* 2012;**97**:118–29.
78. Fitzsimons E, Malde B, Mesnard A, Vera-Hernández M. Household responses to information on child nutrition: experimental evidence from Malawi. Institute for Fiscal Studies, 2012.
79. Kazianga H, Levy D, Linden LL, Sloan M. *The Effects of 'Girl-Friendly' Schools: Evidence From the BRIGHT School Construction Program in Burkina Faso*. Cambridge, MA: National Bureau of Economic Research, 2012.
80. Zivin JG, Thirumurthy H, Goldstein M. AIDS treatment and intrahousehold resource allocation: Children's nutrition and schooling in Kenya. *J Public Econ* 2009;**93**:1008–15.

81. Sur D, Ochiai RL, Bhattacharya SK *et al*. A Cluster-Randomized Effectiveness Trial of Vi Typhoid Vaccine in India. *N Engl J Med* 2009;**361**:335–44.
82. Halloran E, Longini IM Jr, Struchiner CJ. *Design and Analysis of Vaccine Studies*. New York, NY: Springer, 2010.
83. Halloran ME, Struchiner CJ. Study Designs for Dependent Happenings. *Epidemiology* 1991;**2**:331–38.
84. Longini IM, Sagatelian K, Rida WN, Halloran ME. Optimal vaccine trial design when estimating vaccine efficacy for susceptibility and infectiousness from multiple populations. *Stat Med* 1998;**17**:1121–36.
85. Clemens J, Shin S, Ali M. New approaches to the assessment of vaccine herd protection in clinical trials. *Lancet Infect Dis* 2011;**11**:482–87.
86. Halloran ME. The Minicommunity Design to Assess Indirect Effects of Vaccination. *Epidemiol Methods* 2012;**1**:83–105.
87. Moversusisyan A, Melendez-Torres GJ, Montgomery P. Outcomes in systematic reviews of complex interventions never reached 'high' GRADE ratings when compared with those of simple interventions. *J Clin Epidemiol* 2016;**78**:22–33.
88. Moversusisyan A, Melendez-Torres GJ, Montgomery P. Users identified challenges in applying GRADE to complex interventions and suggested an extension to GRADE. *J Clin Epidemiol* 2016;**70**:191–99.
89. Durrheim DN, Reingold A. Modifying the GRADE framework could benefit public health. *J Epidemiol Community Health* 2010;**64**:387.
90. Guyatt GH, Oxman AD, Kunz R *et al*. GRADE guidelines: 8. Rating the quality of evidence—indirectness. *J Clin Epidemiol* 2011;**64**:1303.
91. Miguel E, Camerer C, Casey K *et al*. Promoting Transparency in Social Science Research. *Science* 2014;**343**:30–31.
92. Schulz K, Altman D, Moher D; the CONSORT Group. CONSORT 2010 Statement: updated guidelines for reporting parallel group randomised trials. *BMC Med* 2010;**8**:18.
93. von Elm E, Altman DG, Egger M *et al*. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med* 2007;**4**:e296.
94. Vandenbroucke JP, von Elm E, Altman DG *et al*. Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration. *PLoS Med* 2007;**4**:e297.
95. Metlay JP, Fishman NO, Joffe M, Edelstein PH. Impact of pediatric vaccination with pneumococcal conjugate vaccine on the risk of bacteremic pneumococcal pneumonia in adults. *Vaccine* 2006;**24**:468–75.
96. Piedra PA, Gaglani MJ, Kozinetz CA *et al*. Herd immunity in adults against influenza-related illnesses with use of the trivalent-live attenuated influenza vaccine (CAIV-T) in children. *Vaccine* 2005;**23**:1540–48.
97. de Heer HD, Koehly L, Pederson R, Morera O. Effectiveness and Spillover of an After-School Health Promotion Program for Hispanic Elementary School Children. *Am J Public Health* 2011;**101**:1907–13.
98. Spears D. Essays in the Economics of Sanitation and Human Capital in Developing Countries PhD thesis. Department of Economics, Princeton University, 2013.
99. Baird S, Bohren A, McIntosh C, Ozler B. Designing experiments to measure spillover and threshold effects. PIER Working Paper No. 14–006. Department of Economics, University of Pennsylvania, 2014.
100. Barham T. Enhancing Cognitive Functioning: Medium-Term Effects of a Health and Family Planning Program in Matlab. *Am Econ J Appl Econ* 2012;**4**:245–73.
101. Duflo E. Child Health and Household Resources in South Africa: Evidence from the Old Age Pension Program. *Am Econ Rev* 2000;**90**:393–98.