Nguồn: https://egap.org/methods-guides/10-things-you-need-know-about-spillovers

10 Things to Know about Spillovers

Abstract

This guide[1] helps you think through how to design and analyze experiments when there is a risk of "interference" between units. This has been an important area of research in recent years and there have been real gains in our understanding of how to detect spillover effects. Spillovers arise whenever one unit is affected by the treatment status of another unit. Spillovers make it difficult to work out causal effects (we say why below). Experimentalists worry a lot about them, but the complications that spillovers create are not unique to randomized experiments.

1. What they are

Spillovers refer to a broad class of instances in which a given subject is influenced by whether other subjects are treated.

Here are some examples of how spillovers (or "interference") might occur:

- **Public Health:** Providing an infectious disease vaccine to some individuals may decrease the probability that nearby individuals become ill.
- **Criminology:** Increased enforcement may displace crime to nearby areas.
- **Education:** Students may share newly acquired knowledge with friends.
- **Marketing:** Advertisements displayed to one person may increase product recognition among her work colleagues.
- **Politics:** Election monitoring at some polling stations may displace fraud to neighboring polling stations.
- **Economics:** Lowering the cost of production for one firm may change the market price faced by other firms.
- **Within-subjects experiments across many domains:** the possibility that treatment effects persist or that treatments are anticipated can be modeled as a kind of spillover.

These examples share some features:

- An *intervention*: the vaccine, increased enforcement, election monitoring;
- An *outcome*: incidence of disease, crime rates, electoral fraud; and
- A "*network*" that links units together: face-to-face social interaction, geographic proximity within a city, road distance between polling stations.
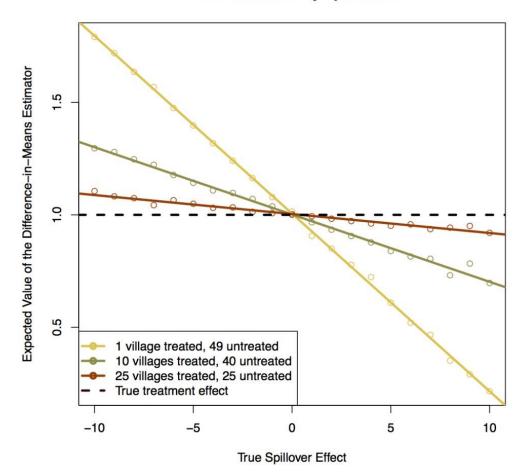
The network is a crucial feature of any spillover analysis. For each unit, it describes the set of other units whose treatment assignments "matter." To take the education example: it may matter to me if you treat another student in my classroom, but it probably doesn't matter if

you treat a student in a different city. I'm connected to the other students in my classroom but not to students in other cities.

2. If ignored, spillovers may "bias" treatment effect estimates

If unaddressed, spillovers "bias" standard estimates of treatment effects (e.g., differences-in-means). "Bias" is in scare quotes because those estimators will return unbiased estimates of causal effects, just not the causal effects that most researchers are interested in.

Imagine an experiment in which there are 50 villages. A treatment (such as a vaccination program) is randomly assigned to some villages but not others. Let's assume that a village receives spillovers if another village within a 5km radius is treated. Imagine the outcome is some measure of health (such as the prevalence of an infectious disease). If we naively compare treated villages to untreated villages, we may not recover an unbiased estimate of the direct effect of treating a village. The reason is that each village's outcome is affected not only by whether that village is treated, but also by whether neighboring villages are treated.

In order to see how spillovers can distort estimated treatment effects, consider the graph below:

Bias Introduced by Spillovers

The graph considers a situation in which the true direct effect of treating a village is 1, and shows how estimated treatment effects can be higher or lower than 1 depending on the direction and size of spillovers as well as the number of villages treated.

In this case, positive spillovers cause a negative bias and vice-versa. This is because when spillovers are positive, the control group mean is inflated, so the difference-in-means is smaller than it otherwise would have been.[2] The extent of the bias, however, depends on the number of villages treated as well as the magnitude of the spillover effect. In this example, the more villages are treated, the smaller the bias resulting from spillovers. This is because when more villages are treated, both the treatment and control group means are similarly inflated by positive spillovers and deflated by negative spillovers.

Often, evaluators are trying to estimate what would happen if a program were rolled out to everyone. Evidence from an RCT that ignores spillover could greatly over or underestimate the total effects of the intervention.

3. Most experimental analyses implicitly or explicitly assume that there are no spillovers.

The assumption that there are no spillovers is known as the **non-interference assumption**; it is part of a somewhat more elaborate assumption sometimes referred to as the **Stable Unit Treatment Value Assumption** (or SUTVA) that is usually invoked in causal inference.

What does the non-interference assumption mean? Subjects can only reveal one of two "potential outcomes": either their treated outcome or their untreated outcome. Which of these they reveal depends on their own treatment status only. The treatment status of all the other subjects in the experiment doesn't matter at all.

We can state the non-interference assumption more formally using potential outcomes notation: $y_i(z_i,Z)=y_i(z'_i,Z')$, if $z_i=z'_i$, where Z and Z' represent any two possible random assignment vectors. In words, this expression states that subject i is unaffected by other subjects' treatment assignments.

How reasonable is the non-interference assumption? The answer depends on the domain. Every study that finds a statistically significant impact of spillovers is providing evidence that the assumption is incorrect in that particular application. Most papers discussing spillovers tend to focus on examples in which the non-interference assumption is false. But other studies suggest that spillovers are sometimes surprisingly weak. Sinclair, McConnell, and Green (2012) for example find no evidence of within-zip code spillovers of experimental encouragements to vote, bolstering the non-interference claims made by the dozens of previous turnout experiments.

4. You need some kind of non-interference assumption whenever you try to estimate spillover effects

The usual non-interference assumption is very strong: it says that there are no spillover effects. When you try to estimate spillovers, you are replacing this strong assumption with a (slightly) weaker one. Perhaps you think that spillovers take place in geographic space — the treatment status of one location may influence the outcomes of nearby units. Allowing spillovers to take place in geographic space requires the assumption that they do not also occur in, for example, social space. This assumption would be violated if the treatment status of, say, Facebook friends in faraway places affects which potential outcome is revealed. To restate this point more generally: When you relax the non-interference assumption, you replace it with a new assumption: no unmodeled spillovers. The modeling of spillovers itself requires strong, often untestable assumptions about how spillovers can and cannot occur.

Suppose we were to model spillovers in the following way. Every unit has four potential outcomes, which we'll write as $Y(Z_i, Z_j)$, where Z_i refers to a unit's own treatment assignment, and Z_j refers to the treatment assignment of neighboring units (i.e., other units within a specified radius). $Z_j = 1$ when any neighboring units are treated and $Z_j = 0$ otherwise.

- $Y_{00} \equiv Y(Z_i = 0, Z_j = 0)$: Pure Control
- $Y_{10} \equiv Y(Z_i = 1, Z_j = 0)$: Directly treated, no spillover
- $Y_{01} \equiv Y(Z_i = 0, Z_j = 1)$: Untreated, with spillover
- $Y_{11} \equiv Y(Z_i = 1, Z_j = 1)$: Directly treated, with spillover
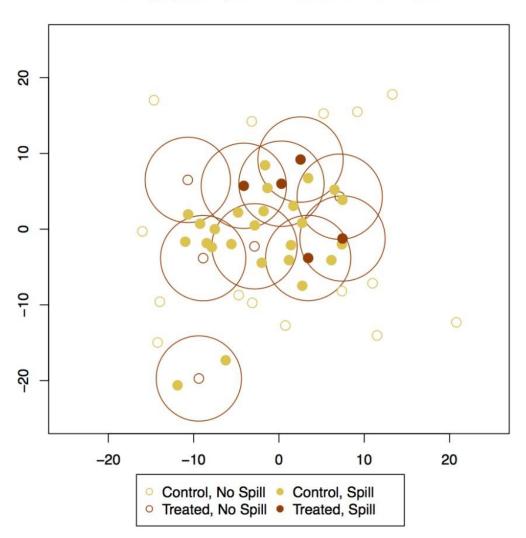
What assumptions are we invoking here? First, we are stipulating that the treatment assignments of non-neighboring units do not alter a unit's potential outcomes. Second, we are modeling spillovers as a binary event: either some neighboring unit is treated, or not — we are ignoring the *number* of neighboring units that are treated, and indeed, their relative proximity.

This potential outcome space is already twice as complex as the one allowed by the conventional non-interference assumption. However, it is important to bear in mind that this potential outcome space can be incorrect in the sense that it does not accurately reflect the underlying social process at work in the experiment.

5. Spillovers are only indirectly "randomly assigned"

The beauty of randomized experiments is that treatment assignments are directly under the control of the researcher. Interestingly in an experiment, spillovers are also randomly determined by the treatment assignment – after all, you're assigning some unit's neighbor to treatment or control on a random basis. The trouble is that the probability that a unit is in a spillover condition is no longer directly under the control of the experimenter. Units that are close to many other units, for example, might be more likely to be in the spillover condition than units that are off on their own.

Take a look at the graph below of 50 units arrayed in geographic space. The 10 red units (both filled and unfilled) were randomly selected for direct treatment and yellow units for control. A filled point represents a unit in a spillover condition, whereas an unfilled point represent a unit that has no treated neighbors within the 5km radius. Notice that the units closer to the center of the graph have a much higher chance of being in a spillover condition than do units towards the edges.

Geographic Spillovers with Radius = 5km

Legend:
- ○ Control, No Spill
- ● Control, Spill
- ○ Treated, No Spill
- ● Treated, Spill

6. To estimate spillovers you need to account for differential probabilities of assignment to the spillover
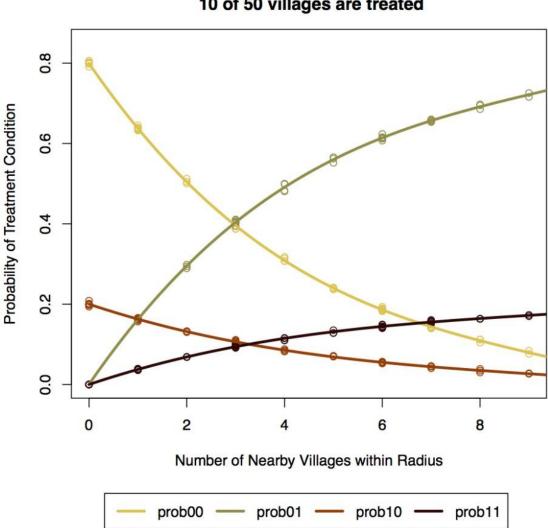
When we estimate causal effects, we have to take account of the probability with which units are assigned to a given treatment condition. Sometimes this is done through matching; sometimes it is done using inverse probability weighting (IPW).

Sometimes, the only practical way to calculate assignment probabilities is through computer simulation (though analytic probabilities can be calculated for some designs). For example you could conduct 10,000 simulated random assignments and count up how often each unit is in each of the four conditions described in the previous section. In R:

```
# Define two helper functions
```

```r
complete_ra <- function(N,m){
  assign <- ifelse(1:N %in% sample(1:N,m),1,0)
  return(assign)
}


get_condition <- function(assign, adjmat){
  exposure <-  adjmat %*% assign
  condition <- rep("00", length(assign))
  condition[assign==1 & exposure==0] <- "10"
  condition[assign==0 & exposure>0] <- "01"
  condition[assign==1 & exposure>0] <- "11"
  return(condition)
}


N <- 50  # total units
m <- 20  # Number to be treated


# Generate adjacency matrix
set.seed(343)
coords <- matrix(rnorm(N*2)*10, ncol = 2)
distmat <- as.matrix(dist(coords))
true_adjmat <- 1 * (distmat<=5) # true radius = 5
diag(true_adjmat) <-0


# Run simulation 10000 times
Z_mat <- replicate(10000, complete_ra(N = N, m = m))
cond_mat <- apply(Z_mat, 2, get_condition, adjmat=true_adjmat)


# Calculate assignment probabilities
prob00 <- rowMeans(cond_mat=="00")
prob01 <- rowMeans(cond_mat=="01")
prob10 <- rowMeans(cond_mat=="10")
prob11 <- rowMeans(cond_mat=="11")
```

We can display the resulting probabilities plotted below against the number of units within the 5km radius. The further from the center a unit is, the higher the probability of not being in the spillover condition.

Probabilities of Assignment Vary with Geographic Location
10 of 50 villages are treated

We must account for these differential probabilities of assignment using IPW. Below is a block of R code that shows how to include IPWs in a regression context.

```
# Define helper functions
get_prob <- function(cond,prob00,prob01,prob10, prob11){
  prob <- prob00
  prob[cond=="10"] <- prob10[cond=="10"]
  prob[cond=="01"] <- prob01[cond=="01"]
  prob[cond=="11"] <- prob11[cond=="11"]
  return(prob)
}
```

```r
get_Y <- function(cond, Y00, Y01, Y10, Y11){
  Y <- Y00
  Y[cond=="10"] <- Y10[cond=="10"]
  Y[cond=="01"] <- Y01[cond=="01"]
  Y[cond=="11"] <- Y11[cond=="11"]
  return(Y)
}


# Generate potential outcomes as a function of position
Y00 <- rnorm(N)


# Treatment Effects
t10 <- 10   # direct effect
t01 <- -3   # indirect effect
t11 <- 5    # direct + indirect


Y01 <- Y00 + t01
Y10 <- Y00 + t10
Y11 <- Y00 + t11


# Randomly generate treatment assignment
assign <- complete_ra(N, m)


# Reveal true conditions
cond <- get_condition(assign = assign, adjmat = true_adjmat)


# Reveal potential outcomes
Y <- get_Y(cond = cond, Y00 = Y00, Y01=Y01, Y10=Y10, Y11=Y11)


# calculate weights
weights <- 1/get_prob(cond=cond, prob00=prob00,prob01=prob01,prob10=prob10,pr
ob11=prob11)


# combine data into a dataframe
df <- data.frame(Y, cond, weights, prob00, prob01, prob10, prob11)


# conduct estimation comparing the spillover condition to the pure control
```

```
fit <- lm(Y ~ cond=="01", weights=weights,

          data = subset(df, prob00 >0 & prob00 <1 & prob01 >0 & prob01 < 1 &
cond %in% c("00", "01")))
```

There are two very important things to remember when using IPW:

- Only include units that have a non-zero and non-one probability of being in all conditions being compared. The code above only compares the pure control condition to the untreated spillover condition (see the subsetting in the lm call).
- Remember the IPW mantra: units are weighted by the inverse of the probability of being in the condition that they are in.

7. Choosing the wrong interference assumption will yield incorrect estimates

You might be tempted to simply construct a model for a particular type of spillover and estimate it. But unfortunately, just as spillovers can produce biased estimates of treatment effects, incorrectly modeled spillovers can create biased estimates of spillover effects (as well as treatment effects).

To get some intuition for the problem, the simulator below lets you pick an interference assumption: the radius beyond which spillovers cannot occur. As in section 4, we assume there are only 4 potential outcomes. The three causal effects that interest us are the average differences between Y_{00} and the other three potential outcomes. The tension in the simulator is between the true (in principle, unknown) spillover network that generates outcomes and the assumed spillover network used for estimation.
The causal effect estimates are only correct when the spillover assumption is correct. The potential outcomes were generated under a true radius of 5km. When any radius other than 5km is selected, some if not all of the estimates are biased. This simulator underlines a discouraging point about spillover analysis: it is generally not possible to know if you've got the "correct" model of spillovers. Short of doing so, the answers yielded by the model will be incorrect.
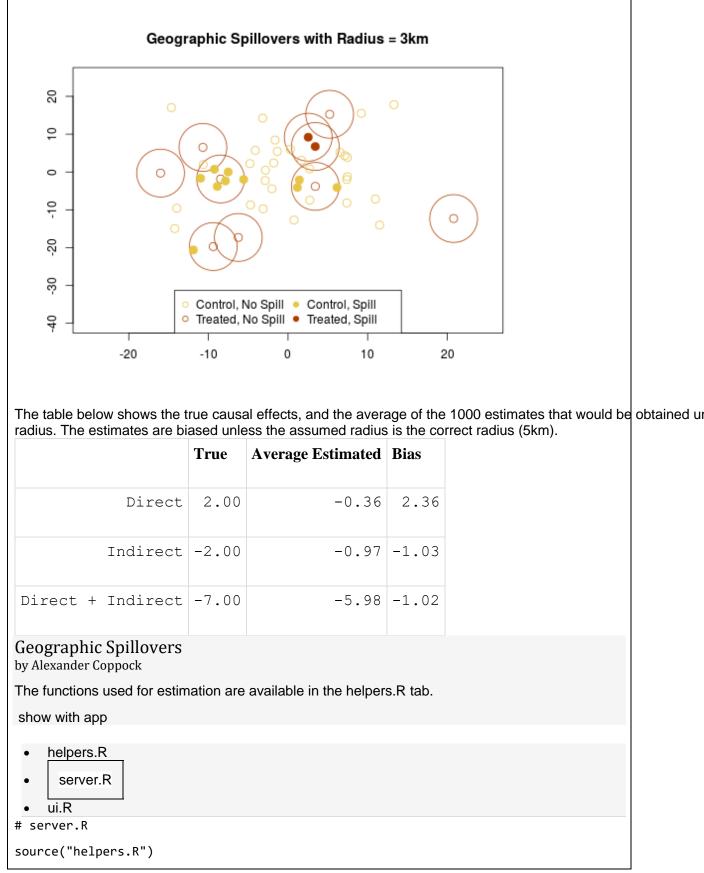
Geographic Spillovers

Estimates of causal effects depend on the assumed spillover structure. In this example, you can choose the rad
spillovers are assumed to be zero. Units are in a spillover condition if there is a treated unit within the specified r
were generated under a true radius of 5km.
Spillover Radius

0 2 0 3 0 2 4 6 8 10 12 14 16 18 20

Choose the magnitude of the true causal effects:
Direct Effect

Indirect Effect

Direct + Indirect Effect

Geographic Spillovers with Radius = 3km

The table below shows the true causal effects, and the average of the 1000 estimates that would be obtained un[der each]
radius. The estimates are biased unless the assumed radius is the correct radius (5km).

	True	**Average Estimated**	**Bias**
Direct	2.00	-0.36	2.36
Indirect	-2.00	-0.97	-1.03
Direct + Indirect	-7.00	-5.98	-1.02

Geographic Spillovers
by Alexander Coppock

The functions used for estimation are available in the helpers.R tab.

show with app

- helpers.R
- server.R
- ui.R

```
# server.R

source("helpers.R")
```

```r
N <- 50  # total units
m <- 10  # Number to be treated


# Generate coordinates
set.seed(343)
coords <- matrix(rnorm(N*2)*10, ncol = 2)


# Generate distance matrix
distmat <- as.matrix(dist(coords))


# Potential outcomes are a function of position
# We write treatment effects Yij
Y00 <- rep(0, N)


# Randomly generate treatment assignment
assign <- complete_ra(N, m)


# True basis for spillovers
true_radius <- 5
true_adjmat <- 1 * (distmat<=true_radius)
diag(true_adjmat) <-0


# Reveal true conditions
true_cond <- get_condition(assign = assign, adjmat = true_adjmat)


# Generate outcomes under all random assignments (using true spillover radius)
Z_mat <- replicate(1000, complete_ra(N,m))
true_cond_mat <- apply(Z_mat, 2, get_condition, adjmat=true_adjmat)


cols <- c("#E7C545","#948E45","#AE3D00",   "#310708")


shinyServer(
  function(input, output) {
```

```
    estimates_fun <- reactive({


    # Treatment Effects
    t10 <- input$direct   # direct effect
    t01 <- input$indirect   # indirect effect
    t11 <- input$directplusindirect    # direct + indirect


    Y01 <- Y00 + t01
    Y10 <- Y00 + t10
    Y11 <- Y00 + t11


    # Generate observed outcomes
    Y <- get_Y(cond = true_cond, Y00 = Y00, Y01=Y01, Y10=Y10, Y11=Y11)
    Y_mat <- apply(true_cond_mat, 2, FUN = get_Y, Y00 = Y00, Y01=Y01, Y10=Y10, Y1
1=Y11)


    # Radius for spillover
    radius <- input$radius


    # Generate numeric adjacency matrix
    adjmat <-  1 * (distmat<=radius)
    diag(adjmat) <- 0


    cond <- get_condition(assign=assign, adjmat=adjmat)


    # Generate probabilies of assignment
    cond_mat <- apply(Z_mat, 2, get_condition, adjmat=adjmat)
    prob00 <- rowMeans(cond_mat=="00")
    prob01 <- rowMeans(cond_mat=="01")
    prob10 <- rowMeans(cond_mat=="10")
    prob11 <- rowMeans(cond_mat=="11")


    # Generate probabilities that units are in the condition that they are in
    prob <- get_prob(cond = cond, prob00=prob00,prob01=prob01,prob10=prob10, prob
11=prob11)
```

```
    prob_mat <- apply(X = cond_mat, 2, FUN = get_prob, prob00=prob00,prob01=prob0
1,prob10=prob10, prob11=prob11)


    # Estimate Treatment effects
    estimates <- mapply(FUN = estimator_ipw, split(Y_mat, col(Y_mat)), split(cond
_mat, col(cond_mat)), split(prob_mat, col(prob_mat)), MoreArgs = list(prob00=prob00
,prob01=prob01,prob10=prob10, prob11=prob11))

    estimates[is.nan(estimates)] <-NA

    means <- apply(estimates, 1, mean, na.rm=TRUE)

    means <- means[c(2,1,3)]

    list(means=means)

  })


  output$radiusmap <- renderPlot({
    # Radius for spillover
    radius <- input$radius

    adjmat <-  1 * (distmat<=radius)

    diag(adjmat) <- 0


    t10 <- input$direct    # direct effect

    t01 <- input$indirect    # indirect effect

    t11 <- input$directplusindirect     # direct + indirect

    cond <- get_condition(assign=assign, adjmat=adjmat)


    # plot exposure
    exposure_cols <-ifelse(assign==1, cols[3], NA)

    treat_cols <-ifelse(assign==1, cols[3], cols[1])

    exposure_pchs <- ifelse(cond %in% c("01", "11"), 19, 1)


    plot(coords,type="n", xlim=c(-25,25), ylim= c(-40,25),xlab="",ylab="",main=pa
ste0("Geographic Spillovers with Radius = ", radius, "km"))

    symbols(coords,circles=rep(radius,N),inches=FALSE,fg=exposure_cols, add=TRUE)

    points(coords,cex=1.5, pch=exposure_pchs, col=treat_cols)

    legend("bottom", legend = c("Control, No Spill","Treated, No Spill", "Control
, Spill", "Treated, Spill"),ncol = 2,
            col=c(cols[1], cols[3], cols[1], cols[3]), pch=c(1,1,19,19))

  })
```

```
    output$esttable <- renderTable({

      t10 <- input$direct   # direct effect

      t01 <- input$indirect  # indirect effect

      t11 <- input$directplusindirect    # direct + indirect

      results <- estimates_fun()

      resultstable <- cbind(c(t10, t01, t11), results$means, c(t10, t01,t11) - resu
lts$means)

      rownames(resultstable) <- c( "Direct", "Indirect", "Direct + Indirect")

      colnames(resultstable) <- c("True", "Average Estimated", "Bias")

      resultstable

    })
})
```

Applied researchers often favor two responses to the "unknowability" of the spillover process. First, they specify "theoretically-driven" models of spillover. Usually, this involves the careful application of qualitative information from the experimental context. Second, researchers conduct robustness checks: they present estimates under a series of spillover assumptions, for example the estimates under increasing radii.

8. Sometimes you can avoid spillovers with "buffer rows"

One approach to addressing the problem of spillovers is to ensure that other units' treatment assignments cannot interfere with potential outcomes, by including "buffer rows" between experimental units. The buffer row analogy comes from agricultural studies in which experimental crop rows were physically separated by non-experimental rows that presumably prevented interference due to local changes in soil nitrogen content, insect behavior, or water usage.

The analogous design choice in our villages experiment would be to sample a set of 50 experimental villages from a larger set of villages, such that all 50 experimental villages were a healthy distance away from each other – say, separated by a minimum of 75km. Of course, we still must make a non-interference assumption along the lines of: "No spillovers between villages that are 75km or more apart." This assumption also rules out spillovers that might take place over non-geographic networks, such as an information network via radio, telephone, or internet.

The main advantage of buffer-row-inspired design is the massive reduction in complexity. You can get a clean estimate of a direct treatment effect using standard analytic techniques, without needing to posit complicated assumptions about the possible avenues for spillover.

The main disadvantage of this design, however, is that by design you cannot estimate natural spillover patterns — which could be critical in understanding normal social processes. (Note: if you do do a buffer design, don't ignore the buffers themselves — data on

these can give you a better handle on spillover effects even though they are never going to receive treatment directly.)

9. There are other design-based approaches for detecting spillover effects.

Some researchers employ a "multilevel" design for exploring spillover effects. The "levels" of the experiment correspond to the spillover network. For example, Sinclair, McConnell, and Green (2012) employ a multilevel design to investigate the possible spillover effects of an encouragement to vote. The levels in their experiment are the neighborhood (nine-digit ZIP code), the household, and the individual. The authors' non-interference assumption is that the treatment assignments of units in other neighborhoods do not matter. What determines which potential outcome is revealed is a combination of three things:

- An individual's own treatment assignment
- The treatment assignment of his or her housemate
- The treatment assignment of others in the neighborhood

Following a relatively complex randomization scheme, the authors assigned treatments so as to create variation in all three levels.

What are the advantages of this design? First, it requires the researcher to stipulate a non-interference assumption ex ante, so there can be no question of fiddling around with interference assumptions until a statistically significant result pops up. Second, it assigns individuals to treatment (including spillover) conditions with known probabilities, so IPW can proceed without having to resort to the simulation method discussed above.

What are the disadvantages? As ever, the difficulty is that the non-interference assumption used in the design stage could be wrong. Perhaps there are significant spillovers across neighborhoods – after all, neighborhood boundaries as described by nine-digit ZIP codes are arbitrary; it could be that the best of friends happen to straddle these boundaries. Or it could be that the spillover network is only indirectly governed by geography. Workplace social ties may be the true means by which the treatment assignment of one unit influences the outcome expressed by others. Of course, nothing about a multilevel randomization scheme prevents the exploration of such alternative spillover structures.

10. Even if a treatment is binary, spillovers might not be. The right model might require dealing with "dosage"

We've explored a non-parametric approach to estimating spillover effects. (See Aronow and Samii 2015 for a fuller treatment of this method as well as a conservative variance estimator in the presence of spillovers.) Units were randomly assigned to one of four conditions with a

complex (but knowable) probability. Our estimates of causal effects were calculated as differences in weighted average outcomes between the treatment conditions. This approach has the advantage of making IPW estimation easy – simply weight each observation by the inverse of the probability of it being in the condition that it's in.

But what about "dosage"? Perhaps in fact spillovers work as a decreasing function of the distance to every other treated unit or in some other more complex way. The spillover is then a continuous variable that describes the "dosage" of exposure to spillovers. The non-parametric IPW approach would require us to chop up the continuous variable in to bins and then calculate average outcomes according to the bin. The IPW estimator quickly becomes quite noisy, as fewer and fewer units occupy each bin.

Bowers, Fredrickson, and Panagopoulos (2013) propose a framework that can accommodate any causal model that maps treatment assignments into potential outcomes. The potential outcomes can be in discrete categories (as we've been assuming for most of this guide) or a continuous function of the dosage of spillovers.

A schematic sketch of their method is as follows. Suppose the causal model has two parameters: β_1, the direct treatment effect and β_2, the indirect effect of a single unit of spillover dosage. A joint test of the hypothesis that $\beta_1=\beta_2=0$ is equivalent to a test of the sharp null hypothesis of no effect. Such a test yields a p-value — the probability that the observed data were generated according to the causal model in which $\beta_1=\beta_2=0$. But we aren't restricted to only obtaining p-values for the hypothesis that $\beta_1=\beta_2=0$. Those parameters could take on any values, and we could associate a p-value with any hypothesized pair of values. The essence of their proposed estimation method is to pick the pair that generates the highest p-value by searching through the set of plausible pairs.

For further reading

Aronow, Peter M., and Cyrus Samii (2015). "Estimating Average Causal Effects Under Interference Between Units." arXiv

Athey, Susan, and Guido W. Imbens (2017a). "The Econometrics of Randomized Experiments." In *Handbook of Economic Field Experiments*, vol. 1 (E. Duflo and A. Banerjee, eds.). arXiv DOI

Athey, Susan, and Guido W. Imbens (2017b). "The State of Applied Econometrics: Causality and Policy Evaluation." *Journal of Economic Perspectives* 31(2): 3–32.

Bowers, Jake, Mark M. Fredrickson, and Costas Panagopoulos (2013). "Reasoning about Interference Between Units: A General Framework." *Political Analysis* 21: 97–124.

Gerber, Alan S., and Donald P. Green (2012). *Field Experiments: Design, Analysis, and Interpretation*, chapter 8.

Glennerster, Rachel, and Kudzai Takavarasha (2013). *Running Randomized Evaluations: A Practical Guide*, modules 4.2, 7.3, and 8.2.

Paluck, Elizabeth Levy, Hana Shepherd, and Peter M. Aronow (2016). "Changing Climates of Conflict: A Social Network Experiment in 56 Schools." *Proceedings of the National Academy of Sciences* 113: 566–571.

Sinclair, Betsy, Margaret McConnell, and Donald P. Green (2012). "Detecting Spillover Effects: Design and Analysis of Multilevel Experiments." *American Journal of Political Science* 56: 1055–1069.

1. Originating author: Alex Coppock, 31 Jul 2014. Minor revisions: Don Green and Winston Lin, 20 July 2016. The guide is a live document and subject to updating by EGAP members at any time; contributors listed are not responsible for subsequent edits.↩

2. Note though, it is generally very difficult to guess the direction of the bias that would be induced by spillover. Claims like, "spillover would only make our treatment effects appear stronger" usually depend on assumptions of treatment (and spillover) effect homogeneity.↩