



## PHW250B Week 7 Reader

### Topic 1: Case Control Study Designs

Lecture: Principles of Case Control Studies.....	2
Lecture: Types of Controls in Case Control Studies.....	14
Lecture: Types of Designs in Case Control Studies.....	25
Lecture: Features of Designs of Case Control Studies. ....	44
Wacholder et al. Selection of controls in case-control studies: principles. American Journal of Epidemiology. 1992. 135(9)1019-1028.....	52
Wacholder et al. Selection of controls in case-control studies: types of controls. American Journal of Epidemiology. 1992. 135(9)1029-1041.....	62
Wacholder et al. Selection of controls in case-control studies: design options. American Journal of Epidemiology. 1992. 135(9)1042-1050. ....	75

### Topic 2: Evaluating and Reporting Case Control Studies

Lecture: Evaluating and Reporting Case Control Studies. ....	84
STROBE Checklist for Case-Control Studies.....	93

### Podcast

Epidemiology Case Studies: Risk Factors for Menstrual Toxic Shock Syndrome - Multistate Case-Control Study.....	95
---	----

### Journal Club

Null et al. (2018). ....	106
Reingold et al. (1989). ....	120

## Lecture: Principles of Case Control Studies



# Principles of case-control studies

PHW250 F - Jack Colford

JACK COLFORD: Today we're going to discuss some of the basic principles that underlie the construction of case-control studies. Case-control studies, of course, are a very central type of design used by epidemiologists in many areas of research.

## Seminal papers on case-control methods by Wacholder et al.



Principles



Types of controls



Design options

Berkeley School of Public Health Jadad 2007 1

I want to point out to you that there are three really important papers we expect you to be reviewing by Sholom Wacholder. And these three papers, from the American Journal of Epidemiology, I consider to be just seminal works that are very clear to read and very helpful in coming to a full understanding of case-control design. And basically, the three papers focus on the principles of case-control study design, the types of controls that are used in case-control studies-- because learning how to pick controls is the key feature of building a case-control study-- and some details about specific design options when using case-control study design.

# Principles

<b>1. Study base</b>	Minimizes differences in the way cases and controls are selected from the base (i.e., <b>selection bias</b> )
<b>2. Deconfounding</b>	Minimizes distortion of the true effect by <b>confounders</b>
<b>3. Comparable accuracy</b>	Minimizes differences in accuracy of the information obtained from cases and controls (i.e., <b>information bias</b> )

Let's talk about some of the basic principles of study design that Wacholder outlines in these papers. The three main principles are called the study base principle, the principle of deconfounding, and the principle for comparable accuracy.

First, the study base principle-- this is the idea that out there, there exists some population we wish we could study in full. But we're doing a case-control study because we can't enroll all the people who are in the population entirely. Our cases and controls are selected from that population. And we want to use approaches that will minimize the ways that these people we have selected differ from this larger theoretical big study population that we can't possibly include in full. In other words, we're trying to minimize the selection bias, how these cases and controls differ in any systematic way from this larger theoretical population.

Secondly, we want to observe the deconfounding principle. And this principle is an idea that we want to take steps that will minimize the distortion of any differences or effects we see between the cases and controls that are caused by confounding variables. We have and will talk more about confounders in the course. But we're trying to minimize the impact of confounders, and this is called the deconfounding principle.

And finally, the third principle is comparable accuracy of the cases and controls. The idea here is that, quite simply, we try to measure the cases and controls in the same way, so that anything we measure about them is comparably measured. Say we're measuring smoking cases and controls. We wouldn't want to use a high-technology urine test in the cases and just use a questionnaire in the controls. That would not represent comparable accuracy in our data collection for the cases and controls.

## Study base principle

- Cases and controls should be 'representative of the same base experience' in order to minimize selection bias.
- The base is the set of persons or person-time in which diseased subjects become cases.
  - Immigration and emigration affect whether some is in the study base at a particular time.
  - A subject is in the study base only if she could hypothetically be enrolled as a case if diagnosed with disease at the time.
- The goal is to sample controls in a way that accurately estimates the exposure distribution in the study base.
- The study base might not necessarily be representative of the general population.
  - If this is the case, it is not possible to estimate prevalence or the risk difference

Wacholder et al., 1992

3

Let's elaborate much more on each of these principles. First, I want to talk about the study base principle. And I indicated here the idea with the study base principle is that the cases and controls we pick are representative of some population base experience, or some base experience it's sometimes abbreviated to say, in order to minimize selection bias. What is this population base?

Well, the population base-- again, which is a theoretical construct-- is the set of persons or person-time from which our disease subjects became cases. While immigration in and emigration out affects some people in the study base population at a particular time, we then have to still find a way to represent the whole study base population with this immigration and emigration going on. And a subject is in the study base only if he or she could hypothetically be enrolled as a case if diagnosed with disease at the same time as the case.

If a person is in a study population but would never be diagnosed for whatever reason-- perhaps they're homeless or perhaps they don't have access to care-- they would not necessarily then be comparable to the cases who did get diagnosed. We want people in the population base to be selected into our case-control study who would be diagnosed as a case if they developed disease in the same way that the cases are diagnosed as a case.

The goal here is to sample controls in a way that accurately affects the estimate of exposure distribution in the study base. Think again of this general population we

have, this theoretical construct of person-time or person experience. We want our sample of that, which is our case control study-- our cases and our controls are samples from that population experience. We want that sample to be accurately representing the distribution of exposure in the population also.

The study base might not necessarily be representative of the general population. If there are large elements of the population in our study base that do not represent the general population, then there's a mismatch between the study base population and the general population. It's very possible in that situation-- let's say I used the example of homelessness earlier.

That could cause a difference between who's really in the general population and who's in the study base. We are trying to get our study base cases and controls to be accurate representations of the study base. We hope that our study base is an accurate representation of the general population, but it might not be.

## Primary vs. secondary study base

- **Primary study base:** the base is defined by the population experience that the investigator wishes to target, with the cases being subjects within the base who develop the disease
  - e.g., a population in a geographic area over a certain time frame
  - Major challenge: identifying all cases
  - Easier to sample controls from a well-defined primary base
- **Secondary study base:** the base is defined as the source of cases, and controls are individuals who would have become study cases if they had developed the disease during the study.
  - Major challenge: it may not be obvious whether or not an individual is a member of the base
  - Easier to identify cases
  - It is not always possible to know whether/when a particular person is in the secondary base.

Wacholder et al., 1992

Now, Wacholder makes a distinction between a primary versus a secondary study base. This is two different approaches to coming up with the construct of the study base for a case-control study. For a primary study base, it's defined as the population experience that we want to target with the cases being subjects within the base who develop the disease.

For example, if we have a population in a geographic area over a certain time frame, all the cases in that frame in that population would represent a sample from our study base. Now, the major challenge here might be identifying all of the cases. Because you will find in a primary study base it's easier to sample controls from a well-defined primary base.

What's an example here? Thinking of a big cancer control registry of a population that attempts to capture every case-- well, that's going to give us a good sampling of the-- more than sampling; it might give us complete coverage-- of the entire primary base.

How about a secondary study base? How is that defined? Well, the secondary study base is defined as the source of cases. Controls are individuals who would have become study cases if they had developed the disease during the study. Not capturing everybody like we are in a primary study base, but however we've captured people, that forms our primary study base.

What's an example of this? Imagine I do a case-control study at UCSF where all of

my cases come from UCSF, not from the whole San Francisco Bay Area, just from UCSF, the research center. Then I have defined as a secondary study base that population that would have appeared as a case at UCSF. I want the people in my secondary population who are chosen as controls, then, to be people who would have come to UCSF had they developed disease.

A major challenge here for a secondary study base is it may not be obvious to tell whether or not an individual is a member of this base. If I've picked cases only coming from UCSF, it can be hard to tell whether my controls would have only gone to UCSF had they been selected. But a secondary study base does make it easier to identify cases because we are defining the study base by the way we define the cases, such as, in this example, UCSF patients. And it's not always possible in a secondary study base to know whether or when a particular person is in the secondary base. I hope these comments are going to help you understand the Wacholder articles as you read them.

## Study base example #1: male infertility

- **Cases:** men struggling with fertility who seek medical help
- **Primary base:** not possible to define in practice because infertile men will not become cases unless they are attempting to have children *and* they seek medical attention
- **Secondary base:** Men who, if they were infertile, would seek help



Wacholder et al., 1992

5

Let's do some specific examples. Here's a first example. Let's say we're trying to study risks for male infertility. We may define cases as men who are struggling with fertility who come to seek medical help. We're only studying men who actually come in and ask for help with their infertility.

If you said, how would we do this as a primary base? It's not really possible to do this study using a primary study base because we don't know about all the men who didn't come in to be seen. All we have are the men who actually came in to be seen.

That is not a random sample of the entire population of men with infertility. This is, however, a secondary base because the men we're studying are ones who actually came in and were seen. These men who came in and are seen are defining our secondary study base here. We can't do a primary study base study.

## Study base example #2: breast cancer screening

- **Cases:** women with breast cancer who were screened for breast cancer
- **Primary base:** women who developed early stage breast cancer who were and were not screened (difficult to identify all of them)
- **Secondary base:** women at risk for breast cancer



Wacholder et al., 1992

6

Let's do another example. Let's talk about breast cancer screening. Here our cases are women with breast cancer who were screened for breast cancer. A primary study base, here, would be women who developed early stage breast cancer who were and were not screened.

It might be hard to identify all of them but we're doing a very broad sampling of the population. Whereas, a secondary study base, here, would be women at risk for breast cancer. That's a more targeted, smaller population. I hope you see that distinction, that in the secondary base, here, we're not including all the women; we're just including women who are known to be at risk for breast cancer.

## How to identify controls and satisfy the study base principle

- Choose a **random sample** of individuals from the study base
  - Each eligible individual in the study base has the same probability of selection as a control
  - More complex sampling schemes (e.g., two-stage sampling, cluster sampling) can be used if certain conditions are met (see Wacholder article)
- Choose a **non-random sample** of individuals from the study base
  - Sometimes random selection is not possible or practical
  - A non-random sample can be taken if the distributions of the exposures of interest are the same in the control series as in a random sample of the (secondary) base
- Choose controls from **outside the study base**
  - People who are not in the base can be selected as controls if the distributions of the exposures of interest are the same in the control series as in a random sample of the (secondary) base

Wacholder et al., 1992

How do we identify controls, and how do we satisfy the study base principle? One way is to choose a random sample of individuals from the study base. Once we've defined the study base, we randomly sample individuals from that study base.

Each eligible individual in the study base has the same probability of selection as a control. This is a key feature. Every single person has a probability of selection, and that probability should be the same for every person in that study base or that sample from the study base.

Now, there are more complex sampling schemes, such as two-stage sampling or cluster sampling, that can be used if certain conditions are met. You should see the Wacholder article for more details on those more advanced methods of choosing controls.

A second way to choose controls is to choose a non-random sample of individuals from the study base because sometimes random selection isn't possible or practical. A non-random sample can be taken if the distributions of the exposure of interest are the same in the control series as in a random sample of the secondary base. Take a look at Wacholder for more extended discussion of this.

Or finally, we could choose controls from outside the study base. This is not ideal but sometimes it's the only choice we have. This situation, people who are not in the study base can be selected as controls if the distributions of exposures of interest are the same in the control series as in a random sample of the secondary study base.

## Deconfounding principle

- The measure of association should not be distorted by confounding.
  - Measured confounders should be controlled for in the analysis
  - Unmeasured confounders should be minimized in the design (e.g. stratification, matching)
- **Example:** using siblings as controls matches cases and controls' environmental and genetic risk factors, reducing those sources of confounding.
- The extent of bias from an unmeasured confounder depends on the strengths of the associations between it and the exposure and disease.



Berkeley School of Public Health  
Wacholder et al., 1992

The next principle we want to discuss is the deconfounding principle. This principle basically states that we want the measure of association not to be distorted by confounding. We want to adjust for or deal with, in our design, all these potential confounding factors. If the confounders have been measured by us when we collected data on the subjects, we can control for them in the analysis. Or we can control for confounders that are unmeasured by trying to minimize their importance either by stratification or matching.

For example, if we're using siblings as controls, we're trying to match cases and controls for their environmental and genetic risk factors, thereby reducing genetics and environment as a source of confounding, as a source of difference between the cases and controls. The extent of bias from an unmeasured confounder depends on the strengths of the association between it and the exposure and the disease, as we'll discuss a lot as we discuss confounding.

## Comparable accuracy principle

- The degree of accuracy in measuring the exposure of interest for the cases should be equivalent to the degree of accuracy for the controls
  - Unless the effect of the inaccuracy can be controlled in the analysis
- Cases and controls may be subject to different types of information bias:
  - E.g., Recall of past exposures may be better among cases than controls
- If errors are non-differential between cases and controls, bias is towards the null.
- If errors are differential between cases and controls, bias is in an unknown direction.
- What if cases and controls were recruited from different hospitals?
  - This principle can still be met if a validation study is done to compare the level of accuracy of diagnostic tests at each hospital

Wacholder et al., 1992

And finally, the third and final principle, the comparable accuracy principle-- the degree of accuracy in measuring the exposure of interest for the cases should be equivalent to the degree of accuracy for the controls unless we can control for the inaccuracy somehow in the analysis. The example I gave earlier, here, was if we're measuring smoking as an exposure in cases and controls, we don't want to measure smoking in the cases with some expensive technique that, say, measures cotinine in the urine and then measure smoking in the controls with just a questionnaire. That would not give us comparable accuracy of information.

Also the cases and controls might be subject to different types of information bias. The cases might recall their past exposure in their memory better than the controls do. So that would not give us comparable accuracy. If the errors are non-differential between the cases and controls, this will always be a bias toward the null.

Say, for example, we use questionnaires to ask about smoking in cases and in controls. And with the same rate, they both make errors about recalling with accuracy their prior smoking. If both the cases and controls have similar inaccuracies, that will create a bias toward the null. That's non-differential misclassification. If the errors are differential between the cases and controls, we can't predict which direction the bias will go. It might lead to a stronger effect among the cases or a stronger effect among the controls.

In some situations, you'll see cases and controls recruited from different hospitals. So does this still meet the comparable accuracy principle? The principle could still be met if a validation study is done to compare the level of accuracy of diagnostic tests at each of the two hospitals. If the level of accuracy of tests at the two hospitals is similar and the patients are otherwise similar, you might convince yourself that cases and controls coming from different sites are going to meet the comparable accuracy principle.

## Summary of key points

- Following the three principles for case-control studies outlined by Wacholder et al. helps to minimize selection bias, confounding, and information bias.
- We can use a combination of study design and data analysis techniques to meet these principles, including:
  - Careful selection of controls
  - Matching
  - Stratification
  - Multivariate statistical models
  - Validation studies

To summarize the key points, when designing and analyzing case-control studies that are explicated by Wacholder in the series of articles I mentioned, we want to help to minimize selection bias, confounding and information bias. And we can use a combination of study design and data analysis techniques to meet these principles, including all of these we've discussed-- careful selection of controls, matching of exposures, stratification by certain exposures. Or we can use multivariate statistical models to adjust for differences. Or we can do validation studies to ensure that there's comparable accuracy.



## Types of controls in case-control studies

PHW250 F - Jack Colford

JACK COLFORD: Let's continue our discussion of case control studies by talking about the types of controls that are used in case control studies.

## Types of controls

- Population controls
- Hospital or disease registry controls
- Controls from a medical practice
- Friend controls
- Relative controls
- The case series as a source of controls
- Proxy respondents and deceased controls

Some people say that the most sophisticated or most difficult part of conducting a case control study is in the selection of the controls. So this is a really important topic. And as you'll see on this bullet list here, there are a number of different populations from which we might draw controls.

We might draw controls, for example, from the general population itself, like California, or the US, or however we conceive of the population. We might use tools or what we would call sampling frames like a hospital or a disease registry from which to draw our controls. Or similarly, we might draw controls from a medical practice.

We might ask the cases in our case control study to provide us with the names of friends and use those friends as controls or use the relatives of cases in the study as controls. The series of cases itself, that is the people we actually enroll in the study as cases, in some special situations can themselves serve as controls for the study. And that's a kind of a specialized situation, but it is possible.

And finally, we can involve proxy respondents, that is, people answering on behalf of others as a way to have controls when the controls we wish we had were deceased.

## Population controls

- Appropriate for study with primary base
- Use a random sample of people from the primary base as a control (must know who is in the study base)
- **Advantages:**
  - Cases and controls come from the same study base
  - Can calculate absolute scale measures such as prevalence, risk difference, and population attributable fraction
- **Disadvantages:**
  - It isn't always possible to take a random sample of the study base
  - Not appropriate when cases cannot be fully enumerated from the study base
  - Difficult when there is no population roster



Wacholder et al., 1992

2

OK. So let's talk first about population controls. Now, these types of controls are appropriate for use with a primary study base. And recall that a primary study base is a situation in which we believe we have either the entire population we're interested in, or we have an appropriate random sample of that population from which we can draw people and work.

So when we use population controls, we take a random sample of people from the primary study base. Obviously, we have to know who the primary study base is made up of in order to select these controls.

So some advantages of using a population controls in a case control study is that the cases and controls clearly come from the same study base. And that's one of the principles of case control design that we talked about earlier. And it also allows us to calculate absolute scale measures, such as prevalence, risk difference, and population attributable fraction because of what we know about the population in this design.

There are some disadvantages. One, it isn't always possible to take a random sample of our study base, in which case, we can't use population control. It's not appropriate when cases can't be fully enumerated from the study base.

So if we can't tell who all the cases are in a study base and randomly select from the cases or use all the cases, then we can't use population controls either. So this is difficult in the situation, for example, where there's no population roster identifying the entire underlying population study base.

## Hospital or disease registry controls

- List of people admitted to and/or discharged from a hospital
- Enroll people without the disease of interest as controls
- **Advantages:**
  - Can reasonably assume that people in the list are members of a secondary study base
  - Comparable quality of information between cases and controls
  - Convenient
  - Can calculate absolute scale measures such as prevalence, risk difference, and population attributable fraction
- **Disadvantages:**
  - Must be careful: if controls are subjects with other diseases, they may have a different distribution of exposure than the distribution in the study base
  - Must be careful of Berkson's bias, in which the exposure is related to the risk of hospitalization for the controls



Berkeley School of Public Health  
Wacholder et al., 1992

Another type of popular controls and case control studies are hospital or disease registry controls. So we start with a list of people who are admitted to or are discharged from a certain hospital, let's say Alta Bates in Berkeley. Then we would enroll people without the disease of interest as controls. These would also come from the same hospital or from the catchment area of the hospital. It could be people who aren't in the hospital but who would go to the hospital.

So some advantage of this is we can reasonably assume that people in this list are members of a secondary study base. Our secondary study base here would be all the cases who come to Alta Bates. There's a comparable quality of information between cases and controls when we use this type of control selection.

And hospital or disease registry controls are very convenient to get because we know right where to get them from usually. And we can calculate absolute scale measures, such as prevalence, risk difference, and population attributable fraction.

There are disadvantages. We have to be careful in several situations. For example, if the controls are subjects with other diseases, they might have a different distribution of exposure risk factors than the distribution we would see in the study base we hope we are really studying.

So, we do a case control study. We have our cases, however we got them. We hope that the controls we've picked represent the same study base, whether that's primary

or secondary. If we're getting our controls from a hospital or clinic, with people who are also in the hospital or clinic, they may be there for other exposure reasons that aren't the same as the entire population study base outside the hospital that we hoped we were studying.

So we have to be careful of what's called Berkson's bias. Here, the exposure in Berkson's bias is a situation in which the exposure is related to the risk of hospitalization for the controls. So for example, broad example, say we're studying something related to smoking. Smoking is associated with lots of diseases, so people might be in the hospital because of higher rates of smoking.

If we choose these people who are in the hospital to serve as our controls, and we're studying smoking, our comparison of the cases and controls with respect to smoking isn't really valuable, because it isn't a comparison to the rate of smoking in people outside the hospital, where we really wish we were comparing for our population study base.

## Controls from a medical practice

- Must make the assumption that cases and controls in the same primary care practice with the same case presentation would follow the same pathway through the medical care system (secondary study base)

- **Advantages:**

- Useful strategy when it is otherwise difficult to find controls who are comparable to cases on access to medical care or referral to specialized clinics.

- **Disadvantages:**

- Complexity entailed by a random selection process for controls within several different practices
  - Exposure distribution for controls may differ from that in the study base— e.g., patients choose a physician for reasons relating to particular conditions that are themselves related to exposure.



Wacholder et al., 1992

4

How about controls from a medical practice? This has some of the same sorts of issues as hospital and disease registry controls. Here, we have to make the assumption that the cases and controls in the same primary care practice, with the same case presentation, would follow the same pathway through the medical care system. That is kind of the principle of the secondary study base.

So again, just to restate that, if someone is a lung cancer patient at this VA medical center, we would expect people who are chosen as controls for this study to also, had they developed lung cancer, to have come to the same VA medical center, that their care kind of would have flowed in the same way.

So there are advantages to this type of control. This is a useful strategy for arriving at controls when it's otherwise difficult to find controls who are comparable to cases on access to medical care or referral to specialized clinics.

There are couple of disadvantages. There's some complexity that comes along by the random selection process for controls when you have several different practices, like how do you properly select-- when you have multiple different practices within a facility, how do you select which controls are from which practice?

And the exposure distribution for controls may differ from that in the study base. For example, patients choose a physician for reasons relating to particular conditions that are themselves related to exposure. Obviously, for example, people might be more likely to be in the pulmonary clinic because of smoking issues. So picking controls from the pulmonary clinic for a study that is of some other disease might overrepresent the amount of smoking in the controls because patients in the pulmonary clinic are more likely to smoke.

## Friend controls

- The intuition behind this is that friends are likely to use the medical system in similar ways
- **Advantages:**
  - Convenient and inexpensive
  - Friends are likely to have shared confounders that are difficult to measure (e.g., socioeconomic status)
- **Disadvantages:**
  - Representativeness of exposure is low for factors related to sociability, such as smoking, diet, or alcohol consumption, because sociable people are more likely to be selected as controls than less sociable people
  - Can lead to overmatching (more on this later)



Berkeley School of Public Health  
Wacholder et al., 1992 5

Friend controls are another possibility. And here, the intuition behind this is that friends are likely to use the medical system in the similar way to the cases. The advantages here, it's very convenient and inexpensive, because we just ask the cases to name for us their friends, and then we approach the friends as possible controls.

But the friends are likely to have shared confounders that are difficult to measure. For example, socioeconomic status is itself related to many diseases, but it can be hard to measure. But by picking friend controls, we likely have gotten people in the control set who are similar with respect to socioeconomic status as the cases, which is what we want.

Some disadvantages. One is that the representativeness of the exposure is low for factors related to sociability, such as smoking, diet, alcohol consumption, because sociable people are more likely to be selected as controls than less sociable people. And this use of friend controls can also lead to a difficulty called overmatching, and we'll talk more about that later.

## Relative controls

- Motivated by de-confounding principle (not study base principle)
- Shared genetics, exposures
- **Advantages:**
  - Convenient and inexpensive
- **Disadvantages:**
  - Can lead to overmatching (more on this later)



Berkeley School of Public Health  
Wacholder et al., 1992

Next, we might use relatives as controls for our cases. And this is motivated by the deconfounding principle, not the study base principle. Think about that, because here we're trying to make the cases in control similar on genetics and exposure. This is not the same thing as the study base principle.

So some advantages here. This is very convenient and inexpensive also, but again, can lead to overmatching, to be discussed later.

## The case series as a source of controls

- An individual can serve as their own control for studies of acute events with transient exposures.
- E.g., activities that trigger heart attack
- "Case-crossover design"
- **Advantages:**
  - Clearly satisfies the study base principle
  - Controls for time-invariant confounders
  - Only patients need to be studied
- **Disadvantages:**
  - There may be time-dependent confounders
  - The design may have lower statistical power than alternatives (and require a large sample size for good precision)



Berkeley School of Public Health  
Wacholder et al., 1992

Next, we talked about this unusual situation where we might use the case series itself as a source of a control. And, so for example, an individual might serve as his or her own control for studies of acute events where the exposures are transient. For example, think about activities that trigger a heart attack.

You might be able to observe the person with the heart attack both before and after the occurrence of the heart attack. So the time at which the person is in existence before the heart attack is serving as a control person for the person with the heart attack, only in this case, it's the same person before and after the heart attack. The person serves as his or her own control.

So this clearly satisfies the study base principle, because when a person is him or herself their own control, they clearly are coming from the same population. The control is clearly coming from the same population as the case. This helps us to control for time and variant confounders, that is, things that are fixed like race, and gender, and so forth. And we only need to study the patients-- that is, that people-- who are in our case series.

There are disadvantages to this type of control. There may be time-dependent confounders, for example-- things that change over time that are difficult to adjust for when the person is the same as the case and the control. And this design may have a lower statistical power than alternative designs and require a large sample size to arrive at good precision.

## Proxy respondents and deceased controls

- Proxy respondents are people who answer questions about exposure and disease when the study subject is deceased or too sick to respond. (e.g., spouse, children)
- **Disadvantages:**
  - Tend to be used for cases more than for controls, which can violate the comparable accuracy principle.
  - Need to minimize the amount of time between the exposure and the interview with the proxy.
  - To use dead controls, need to assume controls' exposure distribution is representative of the study base.



Berkeley School of Public Health  
Wacholder et al., 1992

Another type of control is proxy respondents and deceased controls. So proxy respondents are people who answer questions about exposure and disease when the study subject him or herself is either already dead or too sick to respond. So in other words, a spouse or a child might, for example, serve as a proxy control for someone who's a control in a study but is too ill to answer.

Some disadvantages to this are that this tends to be used for cases more than for controls. And so, it violates the comparable accuracy principle. So in some studies, you'll see proxy respondents used for both cases and controls in order to make the response methods similar in both the cases and controls.

Investigator here needs to minimize the amount of time between the exposure and the interview with the proxy-- that is, it's farther away from the actual exposure, say some childhood exposure to a solvent. Asking that of a proxy respondent to answer for that person in a short period of time after the exposure is a much better approach than asking years later.

And to use dead controls, we need to assume the controls exposure distribution is representative of the study base. And that can be hard to double check and confirm.

## Summary of key points

- There are many potential types of controls.
- Each type has different limitations and advantages.
- Typically we are restricted in the set of potential types of control we can use depending on our exposure and disease of interest.
- When selecting controls, we should try to adhere to the three principles for case-control studies as they apply to our research question.

So to summarize, there are many potential types of controls. Each of them has different limitations and advantages. Typically, we're restricted in the set of potential types of control we can use, depending on our exposure in a disease of interest. That obviously drives what we can choose to do. And when selecting controls, we should try to adhere to the three principles for case control studies as they apply to our specific research question.



# Types of case-control study designs

PHW250 B – Andrew Mertens



In this video, I'll review different types of case-control study designs with an emphasis on how controls are sampled in each of these designs, also the measure of association that's estimated within each design.

# Types of designs

- **Study designs**
  - Density (aka, nested case-control)
  - Case-cohort (aka, case-based)
  - Cumulative
- Each samples controls in a different way and estimates a different measure of association



Let's start off with some clarifications on terminology. The Rothman and Szklo textbooks use different terminology for different types of case-control studies. We prefer the Rothman terminology, because we think it's a little more clear and straightforward. So that's the terminology we'll be using in this class. But I did want to briefly outline and compare the two textbooks' different terminology for types of case-control designs, since you may see both terminologies in study publications. So Rothman describes something called a density case-control study, and Szklo calls this a nested case-control study. Rothman also uses the term nested case-control study, but he uses it more broadly so that it's a category that density case-control studies fall within. I think out in the real world, you will find that these two terms are used interchangeably. And I've found that the nested case-control term is a bit more common.

But in this course, we'll try to stick with density case-control study. The next one is a case-cohort study. Thankfully, that's the same term used in both books. And then the final one is what we think of as a traditional case-control study. The Rothman book calls this a cumulative case-control study. And somewhat confusingly, the Szklo textbook calls this a case-based case-control study. There are some situations where the second category, the case-cohort study, is also called a case-based case-control study.

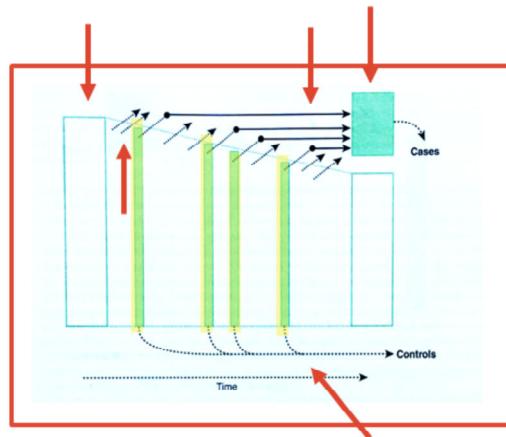
And because that's really confusing, we're just going to try to avoid using the term case-based case-control study in this course. The most important thing for you to

take away is not really to emphasize the terms, but to understand how each of these different designs samples controls in a different way. And as a result, they estimate different measures of association.

So the most important thing, again, try to walk away from this unit understanding how sampling can be done in case-control studies and the implications of that sampling on the measure of association. If you can understand that, then when you're reading papers out there in your real world practice, you'll be able to assess truly what's going on in the study, even if the name of the case-control study design is confusing or is used in a way that's different from what you learned in this course.

## Density sampling

- • Controls are sampled from the person-time of the study base that gave rise to cases.
- • Sampling probability is based on person-time, not on population size.
- • Called a nested case-control study when conducted within a cohort study.
- • It is called density sampling because the odds ratio (OR) estimates the incidence density ratio (IDR)
- • Controls can become a case at a later time (like in a cohort study)



Let's start with density sampling. In a densely sampled case-control study, \*controls are sampled from the person-time of the study base that gave rise to cases. \*This means that the probability is based on the person-time of follow-up, as opposed to the population size or the number of people followed up. \*As I mentioned on the previous slide, this is also referred to as a nested case-control study when it's conducted within a cohort.

\*Let's look at this figure on the right. \*This white rectangle on the left here is the starting population. This is at the beginning of the follow-up period when we define our study base. This is the group of people that will eventually give rise to all of our cases of interest and we'll also be sampling from. And eventually, we'll lose some people from this original population when they become cases or when they're lost to follow-up. \*People who become cases are indicated here with these black circles, \*and they are accumulated in this green box over on the right. \*And then, there were people indicated by these dashed arrows that may die from other causes or be lost to follow-up. And that leaves us with this white rectangle on the right, which is our follow-up population at the end of the study period. Our cases are drawn from the study base throughout the period of follow-up. And then, each time a case is identified, we sample a control from the person-time at that period of time, or close to that period of time.

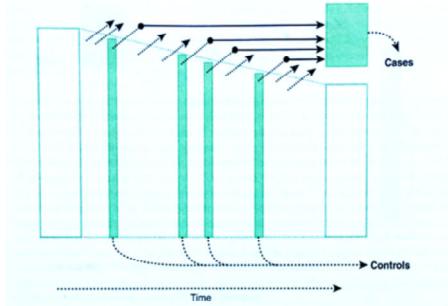
\*This green bar here indicates that a case has been identified \*and a control, or potentially multiple controls, were sampled from people who were being followed up

at that period of time. \*And the second bar here indicates the same thing, just for a different case. \*This figure is showing us four different times when controls were sampled. I'll come back to this in the next few slides.

\*This density sampling method is called density sampling because the odds ratio estimated in this design actually estimates the incidence density ratio. One thing to notice about this design \*is that controls can become cases at a later time. Even though this picture sort of implies that controls are kind of flowing out of the follow-up pool, it's not really the case. We're just counting this first bar here as a control matched to this case here. But in actuality, that control stays within our pool of follow-up time over the course of followup. And that person very well could become a case later on.

## Density sampling

- • Risk-set sampling is a type of density sampling
- Controls are sampled from the person-time at the time the case became a case
- • Cases and controls are matched on time
- • Data must be analyzed using methods to account for the matching



The most common form of density sampling is called \*risk set sampling. It's a type of density sampling. And it means that \*cases and controls are matched very closely on time. \*And as a result, data must be analyzed using special methods that account for that matching. We'll talk more about analysis of matched design later in the course.

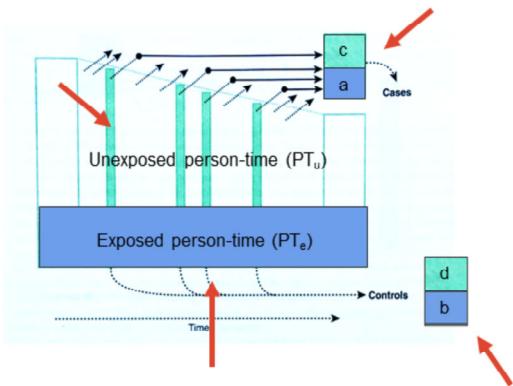
## Density sampling: estimation of the IDR

Data available in case-control study

	Disease	No disease
Exposed	a	b
Unexposed	c	d

Complete data about study base

	Disease	Person-time
Exposed	$\alpha$	$PT_e$
Unexposed	$\gamma$	$PT_u$



Szklo & Nieto, 3rd Ed.  
ME3 Chapter 8



Now, let's talk about how this density sampling estimates the incidence density ratio. First, I want to connect what we just looked at in this figure to two by two tables. \*On the left of the top table here are data available in a case-control study. This is our classic two by two where we have disease indicated in columns, exposure indicated in rows, and then a, b, c, and d. We've seen this many times. \*The table below is the true, complete data about our study base. The third column indicates person-time. The PTe is the total person time exposed, and PTu is the total person time unexposed. In the disease column-- the number of people diseased with exposure is indicated by alpha, and the number of people with disease who were unexposed is indicated by gamma.

Now, let's link this to the figure over here on the right. \*First, if we start up at the top right with the cases, the cases can be divided into exposed and unexposed cases. And we can denote that count by a and c, as in our two by two table, and same for controls down on the bottom right, here. \*The controls are people without disease. And they're either exposed or unexposed, as denoted by b and d in the 2x2 table. \*Let's use blue to indicate those who are exposed. Then we can look at our person-time of follow-up. So the blue people are exposed, \*and the other part of the figure, in green, are unexposed person-time. And if we sum each of those up, that's our total person-time under exposure and person-time without exposure.

# Density sampling: estimation of the IDR

Data available in case-control study

	Disease	No disease
Exposed	a	b
Unexposed	c	d

$$\text{The true IDR} = (\alpha/\text{PT}_e) / (\gamma/\text{PT}_u)$$

How do we know how much the true IDR differs from the IDR estimated by the OR in a case-control study?

If controls are selected so that their exposure distribution matches the exposure distribution of the study base, then the control sampling rates among the exposed and unexposed are equal:

$$b/\text{PT}_e = d/\text{PT}_u \quad \text{and} \quad b/d = \text{PT}_e/\text{PT}_u$$

Complete data about study base

	Disease	Person-time
Exposed	$\alpha$	$\text{PT}_e$
Unexposed	$\gamma$	$\text{PT}_u$

Szklo & Nieto, 3rd Ed.  
ME3 Chapter 8



\*Our true incidence density ratio, if we had complete information about the study base, would be estimated as alpha divided by PT<sub>e</sub>-- person-time of the exposed, and that's the incidence density among the exposed-- divided by gamma over PT<sub>u</sub>, which is the incidence density among the unexposed.

\*How do we know how much the true incidence density ratio differs from the incidence density ratio estimated by the odds ratio in a case control study? We only have the information in the top left two by two table. We don't actually have the information in the bottom left two by two table. So how well can we estimate this incidence density ratio without the complete information?

\*It turns out that if controls are sampled, or selected, so that their exposure distribution matches the exposure distribution of the study base, then the control sampling rates among the exposed and the unexposed are equal. When we say exposure distribution, what we mean is the proportion of people in different categories who are exposed or unexposed. Disease is one category, but confounding variables are additional categories across which exposure distribution need to be balanced.

If this is true, if the controls are selected so that the exposure distributions in the case-control study matches that in the study base, \*then b over PT<sub>e</sub>-- that's the number of people without disease, the controls who are exposed divided by the person-time that's exposed- should be equal to d over PT<sub>u</sub>, which is the same

thing but among the unexposed group. And this is because b over d is the odds of exposure. And that's from the case-control study. And PTe over PTu is the odds of exposure in the complete study base. We can mathematically arrange these as they are here in either of these two formulas.

# Density sampling: estimation of the IDR

Data available in case-control study

	Disease	No disease
Exposed	a	b
Unexposed	c	d

$$\text{The true IDR} = (\alpha/\text{PT}_e) / (\gamma/\text{PT}_u)$$

How do we know how much the true IDR differs from the IDR estimated by the OR in a case-control study?

If controls are selected so that their exposure distribution matches the exposure distribution of the study base, then the control sampling rates among the exposed and unexposed are equal:

$$b/\text{PT}_e = d/\text{PT}_u \quad \text{and} \quad b/d = \text{PT}_e/\text{PT}_u$$

Complete data about study base

	Disease	Person-time
Exposed	$\alpha$	$\text{PT}_e$
Unexposed	$\gamma$	$\text{PT}_u$

If we plug the sampling rates into the OR formula, the OR equals the true IDR if 1) there is complete ascertainment of cases and 2) the sampling rates are equal among the exposed and unexposed

$$\text{OR} = \frac{a/b}{c/d} = \frac{a/(b * \frac{\text{PT}_e}{\text{PT}_e})}{c/(d * \frac{\text{PT}_u}{\text{PT}_u})} = \frac{a/(PT_e * \frac{b}{PT_e})}{c/(PT_u * \frac{d}{PT_u})}$$

Szklo & Nieto, 3rd Ed.  
ME3 Chapter 8

IDR

\*Now, if we plug the sampling rates into the odds ratio formula, the odds ratio equals the true incidence density ratio under two conditions.

\*Let's go through this step-by-step. The odds ratio is equal to a over b divided by c over d. We can multiply the denominator of the numerator by 1 written as PTe over PTe. And we can do the same in the denominator of the denominator using PTu over PTu. And then, if we rearrange things, that gives us a over PTe times b over PTe and c over PTu times d over PTu.

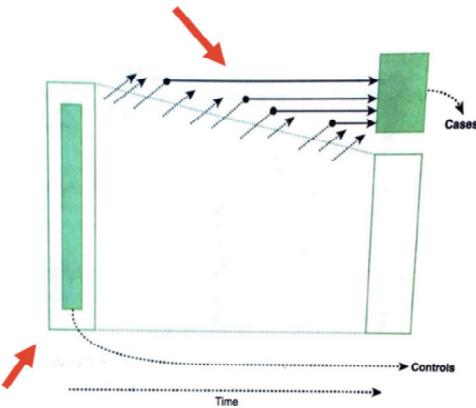
\*Notice that in the left-hand side here, where a over PTe and c over PTu are, that's actually equal to the \*incidence density ratio. For the incidence density ratio to equal the odds ratio, \*we need the second part here-- where it's b over PTe and d over PTu-- we need this part to be equal to 1 so it just goes away. It cancels out. That's another way of saying that the sampling rates need to be equal among the exposed and the unexposed.

If we're looking at a study of smokers and nonsmokers, the percentage of controls sampled from the exposed smoking person-time needs to be equal to the portion of control sampled from the unexposed nonsmoking person-time. And then, this portion of the equation will drop out. For this to equal the true incidence density ratio, we also need a and c to be equal to alpha and gamma. This is another way of saying that we need there to be complete ascertainment of all the cases in the study base. These are our two conditions.

It's pretty tough to get this right, to get these numbers to work out so that the odds ratio exactly equals the incidence density ratio. But if you have information that can give you a good sense of the extent to which these are true, you can make statements about what parameter you're really estimating in this type of case-control study. And so you don't have to just assume that you're stuck with an odds ratio.

## Case-cohort sampling

- • Controls are sampled from the study base at baseline
- • Controls are sampled regardless of how long they stay in the study and whether they develop the disease during the study
- • Sampling probability is based on population size (not person-time)
- • The odds ratio (OR) estimates the cumulative incidence ratio (CIR)
- • Controls can become a case at a later time (like in a cohort study)



Now, let's move on to case-cohort sampling. Here, we have a very similar figure to what we looked at before. But in this figure on the right here,\* what we see is that controls are sampled from the study base at baseline at the beginning of follow-up.

\*And these controls are sampled regardless of how long they contribute to the study's follow-up time and regardless of whether or not they develop the disease during the study. \*And cases are identified in the same way as in the density sampling method. \*The sampling probability is based on population size, not based on person-time, because we're sampling people at baseline. \*And as I'll show you in a moment, the odds ratio for this type of sampling estimates the cumulative incidence ratio. \*And again, controls can become a case later, during follow-up, just like in a cohort study. They just need to not have the disease at the beginning of the follow-up period, and we'll include them as controls regardless of whether they ultimately become a case or not.

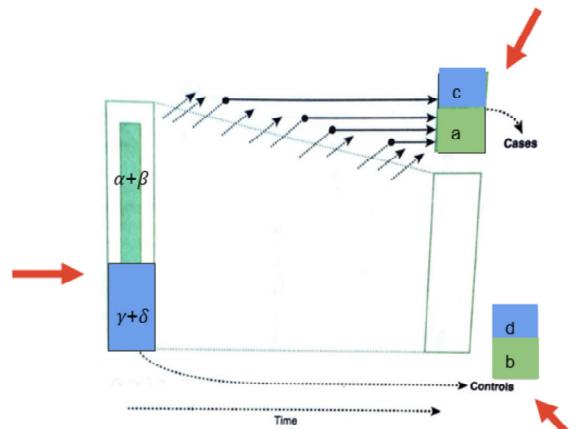
## Case-cohort sampling: estimation of the CIR

Data available in case-control study

	Disease	No disease
Exposed	a	b
Unexposed	c	d

Complete data about study base

	Disease	No disease
Exposed	$\alpha$	$\beta$
Unexposed	$\gamma$	$\delta$



Szklo & Nieto, 3rd Ed.  
ME3 Chapter 8



Now we'll go through some similar logic to illustrate why this type of sampling estimates the cumulative incidence ratio. I've divided up \*the cases and controls on the right side of the figure into a, b, c, and d, just like we did for the incidence density ratio in the density sampling method. \*And this time, our bar on the left side of the figure at the beginning of follow-up is divided into exposed and unexposed. So the exposed is in the white and green, and the unexposed is in blue. The unexposed folks are equal to gamma plus delta, and the exposed folks are equal to alpha plus beta.

# Case-cohort sampling: estimation of the CIR

Data available in case-control study

	Disease	No disease
Exposed	a	b
Unexposed	c	d

$$\text{The true CIR} = (\alpha/N_e) / (\gamma/N_u)$$

How do we know how much the true CIR differs from the CIR estimated by the OR in a case-control study?

If controls are selected so that their exposure distribution matches the exposure distribution of the study base, then the control sampling rates among the exposed and unexposed are equal:

$$b/N_e = d/N_u \quad \text{and} \quad b/d = N_e/N_u$$

Complete data about study base

	Disease	No disease	Total
Exposed	$\alpha$	$\beta$	$N_e$
Unexposed	$\gamma$	$\delta$	$N_u$

Szklo & Nieto, 3rd Ed.  
ME3 Chapter 8



We have the information in the upper left two by two table from our case-control study. The bottom left two by two table is complete data about our study base that we don't have. I'll show you the same exact logic that we used for the density sampling method to show you how this sampling structure estimates the cumulative incidence ratio. \*In our case, the cumulative incidence ratio is equal to alpha over the total N that's exposed divided by gamma over the total N that's unexposed. Alpha over the N that's exposed is the cumulative incidence among the exposed. And gamma over Nu is the cumulative incidence among the unexposed. \*We have the same requirement here for the OR to estimate the CIR. Controls need to be selected so that their exposure distribution matches that of the study base. And that would give us that b over Ne equals d over Nu.

# Case-cohort sampling: estimation of the CIR

Data available in case-control study

	Disease	No disease
Exposed	a	b
Unexposed	c	d

Complete data about study base

	Disease	No disease	Total
Exposed	$\alpha$	$\beta$	$N_e$
Unexposed	$\gamma$	$\delta$	$N_u$

$$\text{The true CIR} = (\alpha/N_e) / (\gamma/N_u)$$

How do we know how much the true CIR differs from the CIR estimated by the OR in a case-control study?

If controls are selected so that their exposure distribution matches the exposure distribution of the study base, then the control sampling rates among the exposed and unexposed are equal:

$$b/N_e = d/N_u \quad \text{and} \quad b/d = N_e/N_u$$

If we plug the sampling rates into the OR formula, the OR equals the true CIR if 1) there is complete ascertainment of cases and 2) the sampling rates are equal among the exposed and unexposed

$$\text{OR} = \frac{a/b}{c/d} = \frac{a/(b * \frac{N_e}{N_e})}{c/(d * \frac{N_u}{N_u})} = \frac{a/(N_e * \frac{b}{N_e})}{c/(N_u * \frac{d}{N_u})}$$

Szklo & Nieto, 3rd Ed.  
ME3 Chapter 8

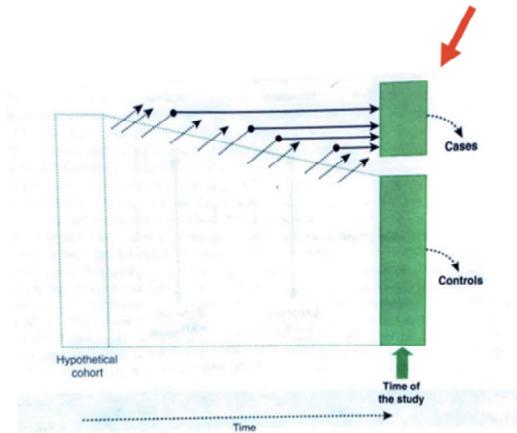


10

And then, we can plug this into \*a formula, as we did before. Again, we start with a over b and c over d. We can add in  $N_e$  over  $N_e$ , which is equal to 1, and  $N_u$  over  $N_u$ , which is equal to 1, in our second step. And then, we can rearrange things, \*the third part of the formula, over here. And what we see is we have a over  $N_e$  and c over  $N_u$ . And as long as we have complete case ascertainment, a equals alpha, c equals gamma, and that right there is our cumulative incidence ratio. b over  $N_e$  and d over  $N_u$  will drop out if they're equal to each other.

## Cumulative sampling

- • Controls are sampled from the study base without disease at the end of follow-up
- Sampling probability is based on population size (not person-time)
- • Appropriate for acute diseases (e.g. diarrhea)
- This is a “traditional” case-control study.
- • The OR does not approximate the IDR or CIR unless the disease is rare.



Now, let's talk about cumulative sampling. This is the type of sampling that gives us the traditional case-control study. \*In this design, we sample controls from the study base without disease at the end of the follow-up period. We can still think of this in terms of a hypothetical cohort, though. On the right, we have the same kind of figure, and it shows us how the cases are being identified over time from this hypothetical cohort. But in practice, really, everything is done at one time.

\*At the end, here, where it says time of the study, we identify the cases, and we identify the controls from the non-disease population. But they did arise from a cohort. We just might not have the information about that cohort.

\*This type of cumulative sampling is more appropriate for acute diseases, such as diarrhea-- things that last for short periods of time.

\*And in this case, the odds ratio does not approximate the incidence density ratio or the cumulative incidence ratio unless the disease is rare

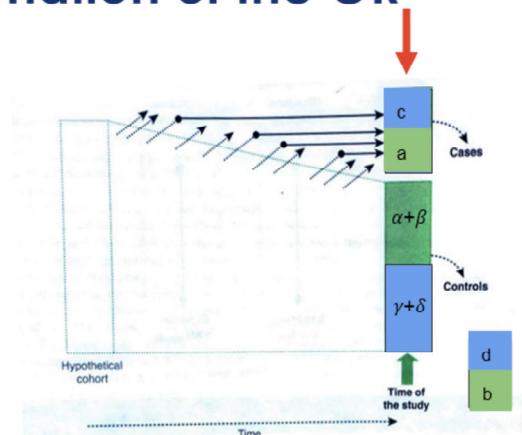
## Cumulative sampling: estimation of the OR

Data available in case-control study

	Disease	No disease
Exposed	a	b
Unexposed	c	d

Complete data about study base  
(That would be captured in a cohort study)

	Disease	No disease
Exposed	$\alpha$	$\beta$
Unexposed	$\gamma$	$\delta$



Szklo & Nieto, 3rd Ed.  
ME3 Chapter 8



12

Now, I'll go through the same steps to show you how, with the cumulative sampling design, we can estimate the true odds ratio. So we see the same figure as before, except this time, \*on the right side of our figure, we're looking at the end of follow-up. Our true odds ratio is equal to alpha over beta divided by gamma over delta.

And now, we'll go through these same steps to understand when the true odds ratio is approximated by the odds ratio from our cumulative case-control study.

# Cumulative sampling: estimation of the OR

Data available in case-control study

	Disease	No disease
Exposed	a	b
Unexposed	c	d

Complete data about study base  
(That would be captured in a cohort study)

	Disease	No disease	Total
Exposed	$\alpha$	$\beta$	$N_e$
Unexposed	$\gamma$	$\delta$	$N_u$

$$\text{The true OR} = (\alpha/N_e) / (\gamma/N_u)$$

How do we know how much the true OR differs from the OR estimated by the OR in a case-control study?

If controls are selected so that their exposure distribution matches the exposure distribution of the study base, then the control sampling rates among the exposed and unexposed are equal:

$$b/\beta = d/\delta \quad \text{and} \quad b/d = \beta/\delta$$

If we plug the sampling rates into the OR formula, the OR equals the true OR if 1) there is complete ascertainment of cases and 2) the sampling rates are equal among the exposed and unexposed

$$\text{OR} = \frac{a/b}{c/d} = \frac{a/(b * \frac{\beta}{\delta})}{c/(d * \frac{\delta}{\delta})} = \frac{a/(\beta * \frac{b}{\delta})}{c/(\delta * \frac{d}{\delta})}$$

Szklo & Nieto, 3rd Ed.  
ME3 Chapter 8

So this time, \*we're doing the same thing, \*but we're swapping in beta over beta and delta over delta in the second step. And when we rearrange things, \*we find that, as long as b over beta is equal to d over delta, this right-hand portion of the formula is equal to 1 and will disappear, which leaves us with a over beta over c over delta. And if a and c are equal to alpha and gamma, then this is equal to our true odds ratio.

## Summary of key points

- Different case-control designs estimate different measures of association.

Design	Control sampling source	Exposure odds ratio estimates:
Density sampling / Nested case-control	Person-time at approximate time when cases occur during follow-up	Incidence density ratio
Case-cohort	Total cohort at baseline	Cumulative incidence ratio
Cumulative	Total non-diseased cohort at time of study	Odds ratio



To summarize, different case-control study designs use different control sampling techniques that estimate different measures of association.

\*In density sampling and nested case-control sampling, the control sampling source is person-time at approximate time when cases occur during follow-up, and it estimates the incidence density ratio.

\*In the case-cohort design, the control sampling source is the total cohort at baseline, and it estimates the cumulative incidence ratio.

\*And in the cumulative design, the control sampling source is the total non-disease cohort at the time of study, and it estimates the odds ratio. This just goes to show that the traditional view that the odds ratio can only be estimated in a case-control study is not 100% true. When we're able to think about a case-control study as part of a cohort study, under certain conditions, we can estimate other kinds of parameters.



## Features of case-control designs

PHW250 F - Jack Colford

PRESENTER: Let's talk now about some additional features of case-control designs. These are slightly more advanced, complex issues that arise in the design of case-control studies.

## Types of designs

- Matching
- Ratio of controls to cases
- One control group for several diseases
- Cluster sampling
- Two-stage sampling

Berkeley School of Public Health  
Wacholder et al., 1992

And I'll talk about several particular features related to design issues, which include matching, how do you choose the ratio of number of controls to the number of cases, can you use one control group for several diseases, what is cluster sampling, and what is two-stage sampling.

# Matching

- Cases and controls can be matched on characteristics (e.g., age, sex)
- Reasons to match:
  - **Matching can increase statistical efficiency**
    - If the number of cases available is small, "efficiency" may be a concern. The precision of the measure of association will be low due to the small sample size.
  - **Improve control of unmeasured confounders**
    - Matching on a variable that is identifiable but not quantifiable (e.g., neighborhood, which can be a proxy for socioeconomic status) can help control for unmeasured confounders
  - **Increase time comparability between cases and controls** (by matching on time-related variables)
- Matching has implications for the statistical analysis (more on this later in the course)

Wacholder et al., 1992

2

First, let's talk about matching. We can match cases and controls on characteristics, any characteristics we might choose or think are important to match on, such as age or sex. But we have to be very careful about A, deciding to match, and B, which factors to match on, because there can be problems with matching on the wrong factors or too many factors.

Matching can increase statistical efficiency. If the number of cases that you have is small, quote unquote "efficiency" might be a concern. And the precision of the measure of the association that you see will be low, if there's a small sample size.

If you're forced to have a small sample size because you have a limited number of cases, then sometimes matching will help you to maximize the efficiency with which you conducted the study. Matching can also improve control of unmeasured confounders. So for example, if you match on a variable that's identifiable but not quantifiable-- for example, neighborhood, which can serve as a proxy for socioeconomic status-- matching can help control for the unmeasured confounders that make up the kind of construct neighborhood.

And additionally, matching can increase the time comparability between cases and controls, if you match on time related variables. Matching does have implications for the statistical analysis that you do with your data, and we'll talk more about this later in the course. And I'll again refer you back to the Wacholder articles that we've discussed earlier.

## Ratio of controls to cases

- Typically the number of cases is fixed, but the number of controls to enroll per case can be chosen by the investigator.
- Widening the study base geographically is a common method of identifying more potential controls.
- Having a ratio  $> 1$  control : 1 case can help increase statistical power, but there is typically no benefit above 4:1.

What about the ratio of the number of controls to the number of cases? Oftentimes, investigator has the ability to have multiple controls per case. Is that a useful feature? Well, typically the number of cases is fixed. We use as many cases as we can find often, but the number of controls to enroll per case can be chosen by the investigator.

Widening the study base geographically is a common method of identifying more potential controls to allow the investigator to have more than one control per case. And having a ratio greater than one control to one case can help increase statistical power. But with empiric study, it can be seen that there's really not a lot of benefit above four controls for every case. When you get to four controls per case, it's usually not of much more value to go to five or six, and so forth.

## One control group for several diseases

- Using a single control for multiple case studies (essentially in different case-control studies) can reduce cost and resources.
  - E.g., cases are patients admitted to Children's Hospital Oakland for different diseases. Controls are children who live in the catchment area of the hospital.
- This approach induces statistical dependence between measures of association for different diseases. Special statistical methods are needed when comparing exposures for different diseases using a single control series.
- An advantage to this approach is that it can identify exposures that have associations with several diseases.



How about using one control group to serve for several different diseases that are being studied? Using a single control group for multiple cases, which is essentially conducting different case-control studies, can reduce cost and resources. For example, cases are admitted to the Children's Hospital Oakland for different diseases. We might study those different diseases against various different control groups, where the controls are our children who live in the catchment area of a hospital but we, in a sense, reuse the control groups to do different case-control studies with the different groups of cases.

This approach induces statistical dependence between the measures of association for different diseases. And one would need to use special statistical methods when comparing exposures for different diseases, if one has used a single control series. And that's beyond the scope of the course right at the moment, but it is the possibility.

And one advantage to this approach is that it can identify exposures that have associations with several disease. By doing several different case-control studies, you might repeatedly see, for example, that smoking comes up as a risk factor in the different case-control studies, obviously confirming or helping to suggest more strongly that smoking is a risk factor for multiple diseases.

## Cluster sampling

- Controls can be selected in groups instead of as individuals.
  - E.g., controls in the same household or family can be enrolled together.
- Reduces the cost of a case-control study by increasing the speed of recruitment.
  - Especially useful when controls must provide biological specimens (e.g., blood samples), which are costly to collect.
- Clusters must be selected so that each member of the study base has an equal chance of being selected.
- The statistical analysis must account for clustering.



Another technique that is employed is called cluster sampling. And in cluster sampling, we select controls in groups instead of as individuals. For example, we might choose as controls the entire household of a case, or the entire family of a case. And they are all enrolled together.

This reduces the cost of a case control study because it increases the speed of recruitment. We get multiple controls quickly for each case. And it's really useful when controls have to provide biological specimens. For example, blood samples or stool samples or salivary samples and so forth, because these can be costly to collect if we have to go out and find multiple controls.

Clusters have to be selected so that each member of the study base has an equal chance of being selected. Back to our principle study base in the design of a case-control study, we want to make sure that everybody in the study base has an equal probability of selection. So there are techniques used to ensure that each selected control had the same probability of selection. And the statistical analysis has to take into account this clustering that was done.

## Two-stage sampling

- This design is intended to reduce the cost of case-control studies.
- In the first stage, exposure and/or confounder information is collected among all subjects.
- In the second stage, the more expensive-to-collect variables remaining are collected among a subset of subjects.
- Example: study of residential radon exposure and lung cancer
  - First stage collects age and smoking status
  - Second stage prioritizes nonsmoking cases and smoking controls for fieldwork to measure radon exposure
  - The purpose is to enforce the proportion of cases and controls in smoking and non-smoking groups to be close to 0.5 to enhance statistical power.



Yet another technique for sampling is called two-stage sampling, and it's intended to reduce the cost of case-control studies. In the first stage of two-stage sampling, exposure and/or confounder information is collected among all the subjects. In the second stage, the more expensive to collect variables that remain are collected among a subset of subjects only.

Example-- say we're doing a study of residential radon exposure in lung cancer. Lung cancer is the outcome. Radon exposure is the exposure variable we're trying to study. In the first stage, we collect age and smoking status. Then in the second stage, we might prioritize nonsmoking cases and smoking controls for the field work in which we measure radon exposure.

And here, the purpose is to enforce the proportion of cases and controls in smoking and nonsmoking groups to be close to 0.5, if we're trying to have a one to one ratio to enhance statistical power. But this is a situation where we might not be using multiple controls per case. But we want one control for every case, so at 0.5 ratio.

## Summary of key points

- Various case-control design features exist to:
  - Reduce the cost of conducting the study
  - Increase the statistical efficiency (e.g., reduce the width of 95% confidence intervals)
  - Control for unmeasured confounders



In summary, various case-control design features are available for us to reduce the cost of conducting studies, increase the statistical efficiency when we conduct these studies-- that is, reduce the width of the 95% confidence interval. And finally, to control for unmeasured confounders.

## Selection of Controls in Case-Control Studies

### I. Principles

Sholom Wacholder,<sup>1</sup> Joseph K. McLaughlin,<sup>1</sup> Debra T. Silverman,<sup>1</sup> and Jack S. Mandel<sup>2</sup>

A synthesis of classical and recent thinking on the issues involved in selecting controls for case-control studies is presented in this and two companion papers (S. Wacholder et al. *Am J Epidemiol* 1992;135:1029–50). In this paper, a theoretical framework for selecting controls in case-control studies is developed. Three principles of comparability are described: 1) *study base*, that all comparisons be made within the study base; 2) *deconfounding*, that comparisons of the effects of the levels of exposure on disease risk not be distorted by the effects of other factors; and 3) *comparable accuracy*, that any errors in measurement of exposure be nondifferential between cases and controls. These principles, if adhered to in a study, can reduce selection, confounding, and information bias, respectively. The principles, however, are constrained by an additional efficiency principle regarding resources and time. Most problems and controversies in control selection reflect trade-offs among these four principles. *Am J Epidemiol* 1992;135:1019–28.

bias (epidemiology); epidemiologic methods; prospective studies; retrospective studies

The purpose of this series of papers is to present a theoretical framework for control selection in case-control studies and show how practical issues can be addressed within this framework. We discuss controversial areas of control selection using the framework and attempt to offer advice when there is relevant empiric information or experience to guide us. For the most part, issues of analysis will not be addressed in the review.

Received for publication May 8, 1991, and in final form February 11, 1992.

Abbreviations: AIDS, acquired immunodeficiency syndrome; HIV, human immunodeficiency virus.

<sup>1</sup> Biostatistics Branch, National Cancer Institute, Bethesda, MD.

<sup>2</sup> Department of Environmental and Occupational Health, School of Public Health, University of Minnesota, Minneapolis, MN.

Reprint requests to Dr. Sholom Wacholder, Biostatistics Branch, National Cancer Institute, 6130 Executive Blvd., EPN 403, Rockville, MD 20892.

The authors thank Dr. Robert Hoover, Dr. Peter Inskip, Dr. Mitchell Gail, Dr. William Blot, Dr. Patricia Hartge, Dr. Jack Siemiatycki, and Dr. Olli Miettinen for their comments on earlier versions of the manuscripts in this set.

In this paper, the first of three, the principles underlying control selection are developed. These principles also apply to the design of cohort studies, as would be expected since the case-control design is simply an efficient sampling technique to measure exposure-disease associations in a cohort or study base. In theory, every case-control study takes place within a cohort, although in practice it can be difficult to characterize the cohort or study base. The identification of the appropriate study base from which to select controls is the primary challenge in the design of case-control studies.

In our second paper (1), we apply the principles presented in this paper to the selection of control groups used in case-control studies, including population controls, hospital controls, medical practice controls, friend controls, and relative controls. We also discuss the use of proxy respondents and deceased controls.

In the third paper of the series (2), we focus on issues encountered after a particular control group has been selected. Some of

the areas discussed are matching, ratio of controls to cases, number of control groups, nested case-control studies, two-stage sampling designs, and issues relating to information bias such as contemporaneity of cases and controls.

We do not intend the principles described and illustrated in these papers to be used for determining whether a study is up to standard. Perfect adherence to a principle can be as difficult to achieve as perfect experimental conditions in a laboratory. Sometimes, one principle can conflict with another. Indeed, tolerating a minor violation of a principle is often the only way to study a particular exposure-disease association. Such a study can still provide valuable information, particularly when the impact of the violation can be evaluated or bounded.

## COMPARABILITY PRINCIPLES

Three basic tenets of comparability underlie attempts to minimize bias in control selection. These are the principles of *study base*, *deconfounding*, and *comparable accuracy*.

**Study base principle.** Cases and controls should be "representative of the same base experience" (3, p. 545). The base is the set of persons or person-time, depending on the context, in which diseased subjects become cases. The base can also be thought of as the members of the underlying cohort or source population for the cases during the time periods when they are eligible to become cases (4). Typically in chronic disease epidemiology, membership in the base is dynamic in the sense that a subject may be in the base at certain times and out of it at other times. The simplest way to satisfy this principle is to choose a random sample of individuals from the same source as the cases; if comparability of time, e.g., age or calendar time, is essential, the sampling should be from the members of the base at risk at the same time as the case's diagnosis. Immigration and emigration from the catchment area affect whether someone is in the study base at a particular time; a subject is in the base only when he or she would be

enrolled as a case if diagnosed with disease at the time. A useful paradigm with an explicitly defined study base is the "nested case-control study" (2, 5-7) where controls are selected randomly from the "risk set," the subjects in the cohort who are at risk at the time of diagnosis of each case.

**Deconfounding principle.** Confounding should not be allowed to distort the estimation of effect. Confounders that are measured can be controlled in the analysis. Unknown or unmeasured confounders should have as little variability as possible. Since this variability is measured conditionally on the levels of other variables being studied, the use of stratification or matching can, in effect, reduce or eliminate the variability of the confounder. For example, using siblings as matched controls in a study of environmental risk factors may result in less variability for genetic risk factors within the matched set and, hence, less confounding than using controls who are not siblings. The extent of bias from an unmeasured or uncontrolled confounder depends on the strengths of the associations between it and the study exposure and disease risk.

**Comparable accuracy principle.** The degree of accuracy in measuring the exposure of interest for the cases should be equivalent to the degree of accuracy for the controls, unless the effect of the inaccuracy can be controlled in the analysis.

We believe that the results of a case-control study become more credible to the extent that these three principles are met. Strict adherence to the principles of comparability outlined here ensures that an apparent effect is not due to 1) differences in the way cases and controls are selected from the base; 2) distortion of the effect by other, unmeasured, risk factors related to exposure; or 3) differences in the accuracy of the information obtained from cases and controls. The aim of the principles is to reduce or eliminate, respectively, selection bias, confounding bias, and information bias.

However, there is an additional practical principle that constrains attempts at comparability.

**Efficiency principle.** The study should be

implemented so as to learn as much as possible about the questions being investigated for a fixed expenditure of time and resources.

### **Study base principle**

The importance of defining the study base in epidemiologic investigations has been recognized for a long time (8). Miettinen (3, 9, 10) distinguishes between a primary base and a secondary base. In a study with a primary base, the base is defined by the population experience that the investigator wishes to target, with the cases being subjects within the base who develop the disease. A population-based case-control study is an example of a study that uses a primary base, where population experience is defined geographically and temporally. However, particularly when ascertainment of all cases in a primary base is difficult or impractical, it may be preferable to use a secondary base, where the cases are defined before the base is identified. In this approach, the base is defined as the source of the cases, and controls are individuals who would have become study cases if they had developed disease during the time of the investigation (9). For example, in a hospital-based study, the cases might be all patients diagnosed with the study disease at one hospital; the individuals contributing to the (secondary) base would be all subjects who would be diagnosed at that hospital had they developed the study disease.

Thus, while the major challenge with a primary base is complete case identification in the base, the major challenge with a secondary base is definition of the study base. Sometimes it may not be possible to resolve definitively whether and when a particular person is in the secondary base. Whether the base is primary or secondary, the critical point is that the *base* and the *cases* need to be defined so that the cases consist, exclusively, of all (or a random sample of) subjects experiencing the study outcome in the base, and that the controls are derived from the base and can be used to estimate the exposure distribution in that base.

The fundamental trade-off between a primary base and a secondary base is that it is easier to sample for controls from a well-defined primary base than from a secondary base, where it may not be obvious whether or not an individual is a member of the base; on the other hand, case ascertainment is complete by definition in a secondary base but can be problematic with a primary base. Selection factors affecting which cases are ascertained and included in the study or the accuracy of identification of the base can cause bias in either a primary or secondary base setting. Identification of a setting where no selection factor operates on the cases or on the sample of the base is often a major challenge in case-control studies, as in the three following examples.

**Referral hospital.** In a study where the cases are subjects who were treated at a referral hospital, the (secondary) base consists of those individuals who would have been treated at that hospital had they been diagnosed with the study disease. The difficulty, of course, is in identifying exactly who would have been referred to that hospital had they developed the study disease.

**Underascertainment of cases.** Incomplete case identification can be substantial for diseases with mild symptoms and for those that do not require medical attention; hence, there could be a spurious association with variables related to utilization of medical services in a study using self-identified cases. A primary base would be unworkable in a study of male infertility, since infertile men will not become cases unless they are attempting to have children *and* seek medical help (11). A secondary base approach would restrict controls to men who, if they were infertile, would seek help, just as all the cases have. Failure to restrict the secondary base accordingly, and thereby failure to exclude controls who would not seek medical advice, could result in a misleading association with correlates of seeking medical attention.

**Temporal differences.** When cases are diagnosed long before controls are selected, it can be difficult to reconstruct the base that was contemporaneous with disease incidence.

Problems in identifying the base sometimes make it very difficult to choose the study base that would be the most scientifically informative. This is particularly true when becoming a case is contingent on a previous condition, as in the following examples.

**Screening.** A simple and powerful approach to evaluate the efficacy of screening for breast cancer would compare mortality in screened and unscreened women in a base of women who had developed early stage breast cancer (12, 13). However, it would be difficult for a case-control study to use this approach because of the problem in identifying members of the base for the denominator of the mortality rate, particularly in unscreened women. Thus, the standard but less efficient approach for case-control studies is to choose controls from a broader base consisting of women at risk for breast cancer (14, 15).

**Prenatal survival.** Exposure to human teratogens may affect prenatal survival and, thus, the opportunity to observe a congenital malformation. This can lead to misleading estimates of effects in case-control studies using livebirths as controls (16).

**Spontaneous abortion.** The ideal base for a case-control study of previous therapeutic abortion on risk of ectopic (tubal) pregnancy would be women who conceive. However, identification of women with intrauterine pregnancies who spontaneously abort would be incomplete, since the women themselves may never become aware of the conception (17). If women who had a previous therapeutic abortion are at extra risk for unnoticed spontaneous abortions, the proportion of missed intrauterine conceptions will differ by exposure, and use of this base could be prone to bias. If the base is women who are trying to conceive, it would be difficult to separate the effects of factors related to conception itself, such as contraceptive use, from those leading to ectopic pregnancy in women who do conceive.

**Acquired immunodeficiency syndrome (AIDS) after human immunodeficiency virus (HIV) infection.** In studies of progression to AIDS after HIV infection, the time of

seroconversion is typically unknown. Thus, defining a base of HIV-positive subjects is difficult (18).

**Sampling from the study base.** In simple random sampling, controls are selected randomly from the base. Therefore, each eligible individual has the same probability of selection as a control, and the sampling is independent; i.e., the presence of a specific subject in the sample does not make the presence of any other more or less likely. In stratified sampling and frequency matching, the base is subdivided into strata determined by factors such as age and sex, and the sampling fraction is allowed to vary across strata. More complex sampling schemes, such as two-stage (19–21) and cluster sampling plans (22), can be used as long as the joint distribution of the exposures of interest in the base can be estimated without bias; generally these require knowledge of the relative sampling fractions and a nonstandard analysis.

Selection bias can be introduced when the sampling fractions for individuals in the base depend on an exposure variable in an unknown way. This dependence is typically indirect and inadvertent, such as when control selection by telephone tends to exclude poor people without phones. However, an analysis of the effects of *other* variables will be unbiased when the source of the dependence can be identified and handled in the analysis as if it were a confounder (23, 24). Unfortunately, recognizing the presence of selection bias can be quite difficult, and this solution requires identification of the selection factor. As with confounding, there is no bias when the selection probability depends on a factor that is unrelated to the exposure.

The study base principle entails the requirement of representativeness of the base but not necessarily of the general population. Representativeness of the general population is crucial in estimating the prevalence of disease, the attributable risk, or the distribution of a variable in a population based on a sample (25). But representativeness, *per se*, is not needed in analytical studies of the relation between an exposure and disease (9, 25). An association found in any

subpopulation may be of interest in itself; in a representative population, an association that is limited to one group may be obscured because the effect is weaker in other groups or because of differences in the distribution of the exposure. On the other hand, detection of variability of the strength of association (effect modification) can be missed if the study base is narrowly defined. If there is reason to believe that an effect is strongest in one particular subgroup, exclusion of other subgroups might be the best strategy for demonstrating that effect; thus, a study of the effect of a possible risk factor for myocardial infarction might restrict the base to subjects who had a previous one. The power of a study targeted at a subgroup can even be greater than the power of a study of the entire population, despite the reduced number of subjects, when the effect is larger in the subgroup (26). Other grounds for exclusions that may increase statistical or economic efficiency include 1) inconvenience (e.g., subjects likely to be too hard to reach); 2) anticipated low or inaccurate responses (e.g., exclusion of subjects who do not speak the language of the interview); 3) lack of variability in the exposure (27, 28) (e.g., a study of the effects of oral contraceptives on subsequent risk of breast cancer should probably exclude women who were past reproductive age when oral contraceptives were introduced into common use); or 4) subjects at increased risk of disease due to other causes (e.g., subjects at high risk for leukemia as a result of chemotherapy for Hodgkin's disease), because cases from the treated group are likely to be attributable to the treatment and therefore may not contribute much to the understanding of other risk factors.

An exclusion rule that applies equally to cases and controls is valid (29) because it simply refines the scope of the study base. One that applies to one but not the other violates the study base principle. For example, a study design that excludes potential controls who had changed their residence between the time of diagnosis of the matched case and the time of selection but places no analogous restriction on the residential mo-

bility of the cases (30) violates the study base principle, and the estimate of effect for an exposure associated with such residential mobility could be biased (30).

**Nonrandom selection from the study base.** In theory, choosing the controls to be a random sample from the base ensures that the controls are representative of the base. When random selection is not practical, as when identification of the base is difficult, a nonrandom subset can be selected if a representativeness assumption regarding the study exposure is met: that the distributions of the exposures of interest are the same in the control series as in a random sample of the (secondary) base (3, 9).

For example, hospital controls are a non-random subset of the study base rather than a random sample from the study base; the validity of a hospital-based study rests on the (perhaps tenuous) assumption that the distribution of exposure among the chosen hospital controls is the same as in the base itself or differs because of measurable factors (1, 9). This assumption is reasonable when the following two conditions apply.

**Identical catchment populations.** Subjects who are admitted to the hospital for the case disease would have been admitted to the same hospital for the control disease, and, conversely, subjects who are admitted for the control disease would have been admitted for the case disease. Thus, determinants of hospitalization and the choice of hospital must be considered carefully in studies with hospital controls.

**Exposure independent of admission.** The exposure is unrelated to the reason for admission of the control.

In the male infertility example considered above, a control series consisting of men whose wives have been identified as infertile at an infertility clinic (11) would be a non-random sample of the appropriate secondary base that would have the same determinants of seeking medical attention as the cases. However, it could introduce selection bias for male correlates of causes of female infertility, such as sexually transmitted disease in the husbands of women with pelvic inflammatory disease (11).

Use of deterministic (nonrandom) schemes for control selection, such as choosing the case's best friend or neighbors, can avoid the need for a representativeness assumption for exposure if 1) the base is divided into nonoverlapping strata and 2) all members of the base in the stratum that includes the case are selected as controls (31). Thus, instead of random selection, a 100 percent sample (31) from a (typically very small) stratum of the study base is chosen. (Strictly speaking, this would not be a case-control study, since no sampling is involved; it is a cohort study where all the strata with no cases can be ignored.) Together, these two requirements imply reciprocity (31). If *A* is included as a control for *B*, then *B* would have to have been included as a control for *A*, if *A* had become the case; this is exactly what is done in a cohort study. In practice, selection of a subset of the stratum deterministically would not produce bias, unless the selection were related to exposure (31). But the possibility of bias does exist with any scheme that allows control selection to be determined by the case or the case's physician.

**Controls from outside the study base.** A proxy control series from outside the base can be used as an "indirect way to probe the base" (9, p. 82), if the representativeness of exposure assumption is met. For example, in a study where blood group is the exposure of interest, use of females as controls when the actual base consists only of males would be theoretically acceptable, under the assumption that blood group distribution does not vary by sex (32). (Of course, published rates on the distribution of blood group might obviate the need for any controls.) In more common situations, it may not be known whether the representativeness assumption actually holds for a given exposure. The validity of the assumption for each exposure studied needs to be assessed individually.

Controls currently living in a neighborhood who are chosen to match cases diagnosed several years earlier should be excluded since they are outside the study base. Excluding controls who have moved into

the neighborhood since diagnosis of the case reduces the problem but does not solve it, since people who moved out of the neighborhood will still be missed.

### Deconfounding principle

While the study base principle clarifies who can be entered into the study, the deconfounding principle addresses the problem created when the study exposure is associated with other risk factors. The principle applies to control selection with respect to unmeasured confounders, since measured confounders can be handled in the analysis.

Confounding can bias the results of any epidemiologic study. Complete assurance of control of confounding is achieved (in theory) by eliminating the variability in the confounding factor. Thus, if the study base consists entirely of males, there can be no confounding by sex. Some control for confounding by genotype might be achieved by the use of relatives of the cases as matched controls. Similarly, controls are sometimes selected to match the neighborhood of the case in order to control for unknown risk factors relating to socioeconomic and ethnic variables or, particularly, access to medical care, which is difficult to control for otherwise. However, controlling for the confounding effects of a risk or selection factor by matching on its correlate or proxy does not eliminate confounding bias (33).

This principle, however, can conflict with the efficiency principle. Selecting controls to have the same values of confounders as cases results in controls who are likely to be more similar to cases with respect to exposure (34); i.e., restricting the variability of the confounding variable will also reduce the conditional variability of the exposure of interest when the exposure and confounder are highly correlated. Studying a population that is almost uniform with respect to unmeasured confounders but also nearly uniform on the exposures of interest is not an effective strategy (35); it is a form of overmatching (in the sense that subjects are effectively, if not deliberately, "matched" on the exposure) that can reduce the precision of esti-

mates of effect without affecting validity (2, 35, 36). Generally, matching on variables that are not risk factors is also overmatching, since the matching may reduce the variability in the exposure of interest without controlling for any confounding (2, 36, 37). On the other hand, reduced precision might be inevitable in the presence of confounding, since it can be a consequence of control for confounding in the design and analysis.

### **Comparable accuracy principle**

Error in the measurement of variables is unavoidable in epidemiologic studies, particularly when information is obtained retrospectively. When the bias due to measurement error can be removed in the analysis, as when the relations between the observed and true exposure measurements are known for cases and controls or an appropriate validation study can be used (38, 39), this principle need not influence control selection. For example, measurements made using both "gold standard" and error-prone methods on some study subjects can allow unbiased estimation of the effects of a poorly measured exposure (38–40). Even when cases' information was obtained from one clinic and that of controls from another, subjects for whom information from both clinics was available can be used as a validation study and can yield unbiased estimates under the assumption that being interviewed in both clinics is unrelated to the responses given (41).

When no correction is possible in the analysis, the comparable accuracy principle calls for all measurement errors that result in distortion of the estimates of effect to be nondifferential; i.e., the error distributions should be the same for cases and controls, as seems reasonable when the mechanisms generating the errors for both groups are the same and are not influenced by disease status. In control selection, one needs to consider the accuracy of information that can be obtained from the controls, e.g., whether recollection of past exposures is better if hospital controls are used rather than healthy population controls.

With nondifferential errors, the bias is typically (but not always) in a predictable direction (toward lack of association) and, unless the measurement is so bad as to be *negatively* correlated with the truth, seldom reverses the direction of the association (42, 43). On the other hand, the effect of differential measurement error on estimates of association is usually unpredictable.

Thus, adherence to the comparable accuracy principle does not eliminate its corresponding bias—information bias. Only elimination of errors (or correction for bias in the analysis using additional information or assumptions (39)) can remove bias entirely. Adherence to this principle may not even reduce bias, as in the hypothetical example presented in table 1. The true odds ratio is 6. When the exposure of the cases is misclassified with specificity and sensitivity both equal to 80 percent, the observed odds ratio from controls with 100 percent specificity and sensitivity will be 3.2 (table 2), which is *less* biased than the 2.7 that would be observed from controls with 80 percent sensitivity and specificity (table 3). So why make this a principle if adhering to it can increase bias? The rationale is to ensure that a positive finding cannot be induced simply by differences in the accuracy of information about cases and controls. While recent work (42, 44) indicates that equal accuracy does not guarantee bias toward the null, a reversal of the *direction* of the association seems unlikely.

Differential errors can be hard to avoid in case-control studies in which exposure information is obtained from interviews with the subjects. Even when interviewers can be blinded to the disease status of a subject, the case generally knows the diagnosis at the time of interview. The disease itself and hospitalization and treatment of the disease may change actual habits as well as perception of current and past habits.

The comparable accuracy principle should not be taken to mean that creating strata within which the errors are equal will be helpful. In fact, stratification designed to achieve nondifferential error within strata can increase bias (45). Thus, creating a stra-

**TABLE 1. Hypothetical example: exposure classified correctly**

Measured exposure	No. of cases	No. of controls	Observed odds ratio
Present	800	400	6.00
Absent	200	600	1.00
Total	1,000	1,000	

**TABLE 2. Hypothetical example: exposure misclassified for cases only\***

Measured exposure	No. of cases	No. of controls	Observed odds ratio
Present	680	400	3.19
Absent	320	600	1.00
Total	1,000	1,000	

\* Specificity and sensitivity are 80% for cases and 100% for controls.

**TABLE 3. Hypothetical example: exposure misclassified for cases and controls\***

Measured exposure	No of cases	No. of controls	Observed odds ratio
Present	680	440	2.70
Absent	320	560	1.00
Total	1,000	1,000	

\* Specificity and sensitivity are 80% for both cases and controls.

tum of direct-interview cases and controls and another for proxy-interview cases and controls does not necessarily reduce bias. Examining the interaction, however, may be helpful since the bias will be greatest in the strata with poorest classification (46).

#### Comparable opportunity for exposure?

Since the focus of a study should be on whether the risk of disease is related to the level of exposure actually received, cases and controls do not need to have equal *opportunity* to be exposed (3, 25, 47). Thus, in a study of cancer treatment on subsequent risk of leukemia, a case who received a treatment could be matched to a control whose physician never prescribed that treatment. Of course, when it is easy to identify subsets of subjects without exposure opportunity, they

can be excluded on efficiency grounds. For example, in a study of oral contraceptive use and risk of myocardial infarction, it would be foolish to include males since sex is a confounder and since there is no variability in exposure in the male stratum.

#### Comment

The use of the term "comparability" in the principles delineated above does not necessarily entail equality. Instead, it means that the study results should be as valid as those that would be obtained under equality. Therefore, our framework of comparability principles, under certain assumptions, allows controls to be selected from outside the study base (1, 9); allows external information to be used to correct for an unmeasured confounder (48); and allows for the use of separate validation studies of the exposure for cases and controls to correct for unequal accuracy (49, 50). Thus, violations of "equality" do not always violate the comparability principles.

#### EFFICIENCY PRINCIPLE

Savings in money and time are two motivations for choosing a case-control design. These factors also affect decisions about other aspects of design, such as the ratio of controls to cases, whether and on which variables to match (3), the source of controls (2), and how they will be recruited. The efficiency principle calls for consideration of costs as well as validity in selection of controls. Statistical efficiency refers to the amount of information obtained per subject; more broadly, efficiency encompasses the time and energy needed to complete the study. For example, even when matching can improve statistical efficiency, the payoff may not be worth the extra effort needed to recruit subjects (51).

We have already seen how the efficiency principle can conflict with the deconfounding principle. When control of confounding is essential for bias reduction, the efficiency principle must be subordinated. However, the principle is important in choosing

among control selection strategies, for example, whether to match or to control in the analysis for each potential confounder (3, 52). Precision of the estimates of effect of a given exposure depends on the variance of the exposure, conditional on the matching factors and the other variables that are adjusted for in the model, regardless of whether or not they are confounders. When a second risk factor is strongly related to exposure and there is a need to control for its confounding effect, any strategy for controlling the effects of the confounder will reduce the conditional variance of the exposure and can reduce efficiency substantially. In a matched-pairs study, this phenomenon is manifested as a reduction in the number of discordant pairs.

## SUMMARY

In this paper, we have presented and described what we believe are the major principles underlying control selection in case-control studies. The principles of study base, deconfounding, and comparable accuracy all address the issue of comparability between cases and controls. Perhaps the key concept is that of the study base. If the study base is identified correctly and if controls are chosen from it properly, the exposure experience of the controls should be representative of the individuals who compose the base. At times, however, the pragmatic principle of efficiency limits the investigator's ability to achieve comparability, reflecting the tension between efficiency and comparability inherent in epidemiologic research.

## REFERENCES

1. Wacholder S, Silverman DT, McLaughlin JK, et al. Selection of controls in case-control studies. II. Types of controls. *Am J Epidemiol* 1992;135: 1029-41.
2. Wacholder S, Silverman DT, McLaughlin JK, et al. Selection of controls in case-control studies. III. Design options. *Am J Epidemiol* 1992;135: 1042-50.
3. Miettinen OS. The "case-control" study: valid selection of subjects. *J Chronic Dis* 1985;38:543-8.
4. Miettinen OS. Response. (Letter). *J Clin Epidemiol* 1986;39:567.
5. Breslow NE, Lubin JH, Marek P, et al. Multiplicative models and cohort analysis. *J Am Stat Assoc* 1983;78:1-12.
6. Mantel N. Synthetic retrospective studies and related topics. *Biometrics* 1973;29:479-86.
7. Liddell FDK, McDonald JC, Thomas DC. Methods of cohort analysis: appraisal by application to asbestos mining (with discussion). *J R Stat Soc A* 1977;140:469-91.
8. Dorn HF. Some problems arising in prospective and retrospective studies of the etiology of disease. *N Engl J Med* 1959;261:571-9.
9. Miettinen OS. Theoretical epidemiology: principles of occurrence research in medicine. New York: John Wiley & Sons, Inc, 1985.
10. Miettinen OS. The concept of secondary base. *J Clin Epidemiol* 1990;43:1016-17.
11. Savitz DA, Pearce N. Control selection with incomplete case ascertainment. *Am J Epidemiol* 1988; 127:1109-17.
12. Tarone RE, Gart JJ. Significance tests for cancer screening trials. *Biometrics* 1989;45:883-90.
13. Chu KC, Smart CR, Tarone RE. Analysis of breast cancer mortality and stage distribution by age for the health insurance plan clinical trial. *J Natl Cancer Inst* 1988;80:1125-32.
14. Morrison AS. Case definition in case-control studies of the efficacy of screening. *Am J Epidemiol* 1982;115:6-8.
15. Weiss NS. Control definition in case-control studies of the efficacy of screening and diagnostic testing. *Am J Epidemiol* 1983;118:457-60.
16. Khoury MJ, Flanders WD, James LM, et al. Human teratogens, prenatal mortality, and selection bias. *Am J Epidemiol* 1989;130:361-70.
17. Weiss NS, Daling JR, Chow WH. Control definition in case-control studies of ectopic pregnancy. *Am J Public Health* 1985;75:67-8.
18. Brookmeyer R, Gail MH. Biases in prevalent cohorts. *Biometrics* 1987;41:739-49.
19. White JE. A two stage design for the study of the relationship between a rare exposure and a rare disease. *Am J Epidemiol* 1982;115:119-28.
20. Weinberg CR, Wacholder S. The design and analysis of case-control studies with biased sampling. *Biometrics* 1990;46:963-75.
21. Breslow NE, Cain KC. Logistic regression for two-stage case-control data. *Biometrika* 1988;75: 11-20.
22. Graubard B, Fears TR, Gail MH. Effects of cluster sampling on epidemiologic analysis in population-based case-control studies. *Biometrics* 1989;45: 1053-71.
23. Breslow N. Design and analysis of case-control studies. *Annu Rev Public Health* 1982;3:29-54.
24. Breslow NE, Day NE, eds. Statistical methods in cancer research. Vol I. The analysis of case-control studies. Lyon: International Agency for Research on Cancer, 1980. (IARC scientific publication no. 32).
25. Rothman KJ. Modern epidemiology. Boston: Little, Brown & Company, 1986.
26. Rosenbaum PR. Case definition and power in case-control studies. *Stat Med* 1984;3:27-34.
27. McKeown-Eyssen GE, Thomas DC. Sample size determination in case-control studies: the influence of the distribution of exposure. *J Chronic Dis* 1985;

- 38:559–68.
28. Lubin JH, Samet JM, Weinberg C. Design issues in epidemiologic studies of indoor exposure to Rn and risk of lung cancer. *Health Phys* 1990;59: 807–17.
  29. Lubin JH, Hartge P. Excluding controls: misapplications in case-control studies. *Am J Epidemiol* 1984;120:791–3.
  30. Savitz DA, Wachtel H, Barnes FA, et al. Case-control study of childhood cancer and exposure to 60-Hz magnetic fields. *Am J Epidemiol* 1988;128: 21–38.
  31. Robins J, Pike M. The validity of case-control studies with nonrandom selection of controls. *Epidemiology* 1990;1:273–84.
  32. Miettinen OS, Cook EF. Confounding: essence and detection. *Am J Epidemiol* 1981;114:593–603.
  33. Greenland S. The effect of misclassification in the presence of covariates. *Am J Epidemiol* 1980;112: 564–9.
  34. Cole P. The evolving case control study. *J Chronic Dis* 1979;32:15–27.
  35. Cole P. Introduction. In: Breslow NE, Day NE, eds. *Statistical methods in cancer research. Vol 1. The analysis of case-control studies*. Lyon: International Agency for Research on Cancer, 1980:14–40. (IARC scientific publication no. 32).
  36. Miettinen OS. Matching and design efficiency in retrospective studies. *Am J Epidemiol* 1970;91: 111–18.
  37. Day NE, Byar DP, Green SB. Overadjustment in case-control studies. *Am J Epidemiol* 1980;112: 696–706.
  38. Greenland S. Variance estimation for epidemiologic effect estimates under misclassification. *Stat Med* 1988;7:745–57.
  39. Armstrong BG. The effects of measurement errors on relative risk regressions. *Am J Epidemiol* 1990; 132:1176–84.
  40. Armstrong BG, Whittemore AS, Howe GR. Analysis of case-control data with covariate measurement error: application to diet and colon cancer. *Stat Med* 1989;8:1151–65.
  41. Elton RA, Duffy SW. Correcting for the effect of misclassification bias in a case-control study using data from two different questionnaires. *Biometrics* 1983;39:659–65.
  42. Dosemeci M, Wacholder S, Lubin JH. Does non-differential misclassification of exposure always bias a true effect toward the null value? *Am J Epidemiol* 1990;132:746–8.
  43. Dosemeci M, Wacholder S, Lubin JH. The authors clarify and reply. (Letter). *Am J Epidemiol* 1991; 134:441–2.
  44. Wacholder S, Dosemeci M, Lubin JH. Blind assignment of exposure does not always prevent differential misclassification. *Am J Epidemiol* 1991; 134:433–7.
  45. Greenland S, Robins JM. Confounding and misclassification. *Am J Epidemiol* 1985;122:495–506.
  46. Walker AM, Velema JP, Robins JM. Analysis of case-control data derived in part from proxy respondents. *Am J Epidemiol* 1988;127:905–14.
  47. Poole C. Exposure opportunity in case-control studies. *Am J Epidemiol* 1986;123:352–8.
  48. Axelson O. Aspects of confounding in occupational health epidemiology. *Scand J Work Environ Health* 1978;4:98–102.
  49. Greenland S, Kleinbaum DG. Correcting for misclassification in two-way tables and matched pair studies. *Int J Epidemiol* 1983;12:93–7.
  50. Espeland MA, Hui SL. A general approach to analyzing epidemiologic data that contain misclassification error. *Biometrics* 1987;43:1001–12.
  51. Thompson WD, Kelsey JL, Walter SD. Cost and efficiency in the choice of matched and unmatched case-control study designs. *Am J Epidemiol* 1982; 116:840–51.
  52. Thomas DC, Greenland S. The relative efficiencies of matched and independent sample designs for case-control studies. *J Chronic Dis* 1983;36: 685–97.

## Selection of Controls in Case-Control Studies

### II. Types of Controls

Sholom Wacholder,<sup>1</sup> Debra T. Silverman,<sup>1</sup> Joseph K. McLaughlin,<sup>1</sup> and Jack S. Mandel<sup>2</sup>

Types of control groups are evaluated using the principles described in paper 1 of the series, "Selection of Controls in Case-Control Studies" (S. Wacholder et al. *Am J Epidemiol* 1992;135:1019-28). Advantages and disadvantages of population controls, neighborhood controls, hospital or registry controls, medical practice controls, friend controls, and relative controls are considered. Problems with the use of deceased controls and proxy respondents are discussed. *Am J Epidemiol* 1992;135:1029-41.

bias (epidemiology); epidemiologic methods; retrospective studies

In this paper, we apply the comparability principles of study base, deconfounding, and comparable accuracy presented in our previous paper (1) to the practical problem of choosing a control group. A number of choices for sources of controls are discussed and evaluated within the framework of the principles. We also offer specific suggestions that we believe are useful in choosing controls.

#### POPULATION CONTROLS

In a study with a primary base, where the focus is the disease experience of a population during a specified time interval in a defined geographic area, randomly sampled controls from that population satisfy the study base principle. More complex sampling schemes, such as frequency matching or cluster sampling (2, 3), are also appropriate, as long as the analysis properly accounts

for the sampling plan. When a roster identifying all members of the base is available, controls can be selected simply as a random sample from that roster as in a nested case-control study (2) or a case-cohort study (2).

When the probability of case identification among members of a primary base depends on a variable, the study base principle is violated and there can be selection bias, unless control selection depends proportionally on values of that variable. For example, when the probability of disease diagnosis depends on access to medical care, a hospital control series with similar dependence on access might be more appropriate than population controls (4). Or, in a case-control study of occupational risk factors using population controls, investigators might consider excluding cases diagnosed at smaller hospitals in the catchment region for logistic reasons. If these hospitals tend to serve rural communities, however, urban occupations may be overrepresented among the cases. An alternative type of control, such as hospital controls, or stratification by geographic factors in the design or analysis may alleviate this problem.

Received for publication May 8, 1991, and in final form February 11, 1992.

Abbreviation: RDD, random digit dialing.

<sup>1</sup> Biostatistics Branch, National Cancer Institute, Bethesda, MD.

<sup>2</sup> Department of Environmental and Occupational Health, School of Public Health, University of Minnesota, Minneapolis, MN.

Reprint requests to Dr. Sholom Wacholder, Biostatistics Branch, National Cancer Institute, 6130 Executive Blvd., EPN 403, Rockville, MD 20892.

#### Advantages of population controls

There are a number of advantages of population controls.

**Same study base.** Selection of population controls from a primary base ensures that the controls are drawn from the same source population as the case series (the study base principle (1)). The mechanisms for sampling from a population are similar to those commonly used in survey research.

**Exclusions.** The definition of the base can encompass the exclusions, e.g., not being in the catchment area at the appropriate time, being a previous case of the disease under study, or not being at risk of the disease under study, such as women who have had a hysterectomy in a uterine cancer study.

**Extrapolation to base.** The distribution of exposures in the controls can be readily extrapolated to the base for purposes such as calculations of absolute or attributable risk (5) or to learn about the distribution of exposures in the population; for example, a detailed diet questionnaire could be used to study differences in food consumption between blacks and whites. By contrast, a hospital control series that is appropriate for a study using cases drawn from the same hospital cannot be used for estimating attributable risk in a population without making assumptions about the representativeness of the case series in that population.

### Disadvantages of population controls

Population controls can be inappropriate when there is incomplete case ascertainment or when even approximate random sampling of the study base is impossible because of nonresponse or inadequacies of the sampling frame (6). When case ascertainment is incomplete and probability of ascertainment depends on the factors being studied, hospital-based or other controls may be preferable (7).

Population controls have some other disadvantages.

**Inconvenience.** Sampling from the population instead of using a more readily available series, such as other hospitalized patients, can be less convenient and more expensive.

**Recall bias.** Differences between cases and healthy controls can lead to violation of the

comparable accuracy principle. Despite an interviewer's best efforts to have the subject's response refer to the period before disease, the responses by a previously hospitalized case may reflect modifications in exposure due to the disease itself, such as drinking less coffee or alcohol after an ulcer, or due to changes in perception of past habits after becoming ill.

**Less motivation.** Population controls may be less motivated to cooperate than hospital controls.

### Selection of population controls when a roster exists

Selection of population controls is simplest when there is a complete listing of the study base. It is useful if the roster has a telephone number or at least an address with which to make contact with the subject. Rosters that have been used include the following: annual residence lists, compiled by law in some areas such as Massachusetts (8) and several European and Asian countries (9); birth certificate records for studies of disease in children (10); Health Care Financing Administration files for Medicare recipients, with coverage of about 98 percent of the United States population aged 65 years and over (11); and electoral lists prepared for each election by a door-to-door survey in such countries as Great Britain and Canada. It is important to remember that, when the cases are identified by a method other than follow-up of subjects on the roster, such as through a disease registry, cases who are not included on the roster, such as those who are not citizens when using electoral lists, should be excluded from the study.

### Selection of population controls when no roster exists

When no roster exists, it can be difficult to ensure that every eligible subject in the study base has the same chance of selection. Bias can be induced when methods rely on contacts with a household, either by telephone (12) or in person, and only one eligible control per household is selected. Con-

sider a study of a disease in children where controls must be within 2 years of age of the case to which they are matched and only one control per household will be selected. A child with a sibling of similar age is less likely to be selected as a control than one with no siblings. This violates the study base principle and, since cases are selected regardless of the ages of other family members, can lead to bias in estimating the effects of variables related to family size (12). Independent determination of whether to select each subject would eliminate this problem, or more than one control can be selected from the same home, and the possible dependence in their responses can be accounted for in the analysis (3).

*Random digit dialing.* Random digit dialing (RDD) can be used to select population controls when no roster exists. RDD and its variants generate sets of telephone numbers without relying on a directory that would not have new or unpublished numbers (13). The aim of RDD is to ensure that each residential number has an equal chance of selection, while minimizing the number of phone calls to nonresidential or inappropriate numbers (13).

Investigators have flexibility in the details of RDD (13, 14). In the standard method (13), a random sample is drawn from working sets of telephone exchanges provided by the telephone company, i.e., the first several numbers of the complete telephone number, typically eight of ten (including area code) in the United States. The number is then completed with two random numbers. The complete number is dialed to determine whether it is a working residential phone. If it is, a predetermined number of calls are made to that exchange; if not, the exchange is discarded. The extra steps are included in order to reduce the numbers of calls to exchanges that have relatively few residential numbers.

Does RDD satisfy the study base principle or can it generate a selection bias? It is easy to show that each phone number in the area has the same chance of being reached. The probability that an exchange is selected is proportional to the number of working

numbers in the exchange, but the probability that a particular number is chosen from that exchange is inversely proportional to the number of working numbers in the exchange. However, the goal in control selection is a random sample of eligible subjects, not of telephone numbers. Incomplete phone coverage, residences that can be reached by more than one phone number, more than one person in the household who is eligible to be a control, and nonresponse can all lead to possible selection bias, unless accommodation is made in the design or analysis. Stratification on numbers of telephone lines and eligible residents in the household can alleviate some of the problems. Advances in technology, such as answering machines and call forwarding, have added complications to this method.

The first contact with a household is most often used for screening and to obtain a census of the household. Information on the address of the house and on the name, age, sex, and race of each household member is obtained. Based on the responses, a sampling frame is generated, and a random sample of identified eligible subjects is selected to be controls. These individuals can then be contacted by letter, by telephone, or in person for interview. When the sampling scheme is simple, such as when it is based simply on age and sex, the census and the interview itself can be done in a one-step process (14, 15), thereby reducing the overall percentage of refusal.

When telephone coverage is low, RDD will miss a substantial proportion of subjects and can result in biased estimates of the effects of exposures related to telephone coverage, such as socioeconomic status. While telephone coverage in the United States is 93 percent, it is lower for residents of the South, for blacks, and for the poor; in 1986, only 56 percent of Southern blacks living in households with yearly income below \$5,000 had telephones (16). In the United States, the difference in the proportion of smokers in households with phones and in those without phones is 1 percent, though this difference is 4 percent among those with income below \$5,000 (16). Incomplete tele-

phone coverage is often, therefore, a smaller problem than nonresponse and refusal.

The clustering of exchanges in RDD results in a sample that is not random, since not every possible subset of the population can be chosen. This clustering does not bias estimates of effects but can lead to overly optimistic estimates of the precision of point estimates unless addressed in the analysis (3).

Sometimes a variation of RDD is used with controls selected using the exchange of the case in order to generate subjects matched on factors that are difficult to measure but are believed to be related to geographic area (17, 18). It is not clear, however, whether "neighborhood matching" would satisfy the study base criterion. Would controls selected to have similar phone numbers actually become cases (e.g., be admitted to the study hospital, if the cases were a hospital series) had they been diagnosed with disease? Moreover, the extent to which such a procedure matches on area or region has not been empirically demonstrated. Matching on primary care practice, discussed below, may be a better approach in some situations.

RDD can be expensive and time-consuming when targeting subgroups of the population. For example, an average of almost 35 households must be screened to identify one black male, 64 households to identify one Hispanic male between 20 and 29 years of age (19), and almost 70 to identify one male aged 75 or older (13). Hence, some studies in the United States use Health Care Financing Administration rosters as a substitute to identify controls over age 65.

RDD is an option for identifying controls from a particular ethnic group who tend to be clustered in certain neighborhoods because less effort is expended in nonproductive exchanges (13). The Donnelley data base (Donnelley Marketing Information Services, Stamford, Connecticut), which classifies exchanges by race and income, can also be used for stratification to identify blacks and Hispanics more cost effectively (19); it is not so helpful, however, for identifying groups such as Asian Americans whose residential patterns are less concen-

trated geographically (19). However, this technique may violate the study base principle and lead to bias when the exposures are related to cultural factors associated with the diversity of the subject's neighborhood, since eligible subjects living in heterogeneous areas may be underrepresented in the control group. For example, Asian Americans living in so-called Chinatowns are likely to have life-styles and diets different from those living in an ethnically heterogeneous neighborhood.

**Neighborhood controls.** Population controls can also be selected using residences, rather than telephone numbers, as the sampling unit. This strategy can be particularly useful when telephone coverage is low. In area probability sampling, controls are selected randomly from a roster of residences, perhaps obtained from a recent census. However, creating a roster when one is not available can be extremely expensive.

Instead of using a random sample from a roster of residences, "neighborhood controls" are typically selected from residences in the same city block or other geographic area as the case, in an attempt to reduce the variability of factors such as access to medical care and socioeconomic status. For neighborhood controls to satisfy the study base principle, one must consider the base as divided into geographically defined strata. Use of neighborhood controls in a study with a secondary base may not satisfy the principle; for example, a neighbor who would not be admitted to the same hospital under the circumstances that led to admission of the case would be outside the base. Thus, bias could result if the source of cases is a religiously affiliated hospital and the neighborhood is religiously heterogeneous.

Neighborhood controls are usually chosen deterministically (nonrandomly) within a stratum defined geographically. If the selection is not random, one must rely on an assumption that the selection process is independent of the exposure, which is equivalent to the exposure distribution being the same as that in the study base (1, 20, 21). To protect that independence, the interviewer should not be given the flexibility to

choose which house to select. Instead, a particular algorithm, such as the one depicted in reference 22 or perhaps based on a reverse directory (sorted by address), should be used in order to avoid bias arising from interviewer selection of residences that appear more likely to cooperate.

Ideally, the neighborhood control should have been a resident of the house when the index case was diagnosed. Controls who recently moved into a neighborhood and are chosen to match cases from the neighborhood diagnosed several years earlier should be excluded since they are outside the study base. Excluding controls who have moved into the neighborhood since diagnosis of the case reduces the study base problem but does not solve it, since people who moved out of the neighborhood will still be missed (1). Whenever cases are diagnosed several years before control selection, use of current residents (or any other current roster) raises the possibility of distortion of the distributions of factors associated with mortality and migration, particularly if socioeconomic or ethnic characteristics of the neighborhood have changed. Old reverse directories, visits to long-term residents, property tax records, old plat maps, and registers of deeds can be used to historically reconstruct neighborhoods (23) in order to find neighborhood controls contemporaneous with the case.

Neighborhood controls have two main advantages. 1) Control selection does not require the existence of a roster or use of a telephone, and 2) confounding factors associated with neighborhood may be balanced between cases and controls.

Thus, neighborhood controls can be an attractive alternative for studies with a primary base when no roster of the population is available or, possibly, for studies where the cases are obtained from hospital lists. The disadvantages of neighborhood controls include the potential for not satisfying the study base principle, particularly in studies with a secondary base; the high cost associated with contacting each potential control (24); the use of the household as the sampling unit, as in selection of controls by telephone; and the difficulty in documenting

nonresponse, since one does not know the number of eligible subjects in homes for which there is no response. In addition, there may be overmatching on the study exposure because of similarities between cases and controls from the same neighborhood on exposures related to residence, particularly in buildings with more than one household unit. These multiunit dwellings, especially apartment buildings, present additional problems because of the difficulty of enumerating all household units within the building for sampling. Moreover, access to the buildings themselves can be a problem in many large cities.

#### **HOSPITAL OR DISEASE REGISTRY CONTROLS**

When a list of admissions or discharges from a hospital or clinic is the source of the cases, the same list can be used as the source of controls, too. The points below regarding hospital controls also apply to disease registry controls, drawn, for example, from a tumor or malformation registry. One attraction of hospital controls is that one can reasonably assume that patients admitted to the same hospital as the cases are members of the same (secondary) base (1, 20). The most serious danger with hospital controls is that choosing subjects with other diseases may jeopardize the assumption of representativeness of exposure (1, 20, 21), namely, that the distribution of the exposures under study in the controls is the same as that in a random sample from the base that produced the cases. This is equivalent to assuming that there is no relation between exposure and the diagnoses used to determine inclusion of controls. For example, use of other women who undergo dilatation and curettage as controls for a study of endometrial cancer (25) probably meets the assumption regarding membership in the secondary base but fails the representativeness of exposure assumption when estrogen use is related to conditions indicating dilatation and curettage (26-28).

The representativeness of exposure assumption is not quite as difficult to meet as

it may seem because it must hold only within strata used in the analysis or conditionally on factors for which adjustment will be made in the analysis (1, 20). Thus, for example, stratification by sex eliminates bias for an exposure even if the exposure is associated with a control disease unconditionally, as long as it is unassociated with the control disease separately for males and females.

Hospital or registry controls are usually more appropriate than are population controls if a sizable fraction of diseased subjects in the base will not become cases in the study and if the ones who do have different exposures from those who do not. For example, multiple sclerosis patients referred to an academic center about 60 miles (about 100 km) away were found to have demographic and severity characteristics different from those of other multiple sclerosis patients at the center from the same area who were not referred (29). It would be difficult to reflect this heterogeneous referral pattern using population controls. Use of hospital controls from another disease with a similar referral pattern might provide more assurance that all subjects share the same study base; alternatively, stratifying by geography or by referral status might be effective.

Use of a hospital control series consisting of subjects with a disease the outward manifestations of which are identical to those of the disease of interest can eliminate one source of selection bias. When differential diagnosis is made on these subjects, the ones with the index disease become cases and the ones with the "imitation" disease become controls. If the imitation disease is unrelated to the exposure of interest, these controls would be appropriate; Miettinen (20, p. 79) describes them as "ideal."

Hospital controls have other advantages.

**Comparable quality of information.** A major advantage is that generally hospital controls are more comparable to cases with respect to quality of information, since they too have been ill and hospitalized. However, careful consideration of the environment where information is gathered, the content

and phrasing of the questions, and the diseases to be included in the control series is needed, since simply being sick does not necessarily entail comparable accuracy and avoidance of recall bias. Selecting hospital controls with conditions that are believed to lead to similar errors in recall may alleviate some of the problems that cause this form of information bias. For example, in a study of birth-related risk factors for testicular cancer in men treated at a military or tertiary care hospital, controls were age-matched men with other cancers, presumed to be unrelated to the study exposures, at the same hospitals (30). Since subjects' mothers provided information on the key exposures, which occurred during early childhood, the use of hospital controls was particularly appropriate because it ensured that the sons of all the mothers interviewed for the study had had malignancies (30). Further, the study hospitals drew patients from across the United States, so these controls were likely to have referral patterns similar to those of the cases (30).

**Convenience.** Hospital controls may be the most convenient choice when controls will be asked to provide bodily fluids or to undergo a physical examination, as when looking for dysplastic nevi in a study of skin melanoma.

In addition to the need to satisfy the representativeness of exposure assumptions noted above, hospital controls have other difficulties.

**Different catchments.** Even when controls are identified from the same registry or hospital as the cases, the catchments for different diseases within the same hospital may be different (31), violating the study base principle. For example, an urban teaching hospital associated with a medical center may provide primary medical care to poor people in the neighborhood and also serve as a tertiary referral center providing sophisticated services for certain medical conditions. Restricting the study base to people living in the vicinity of the hospital can alleviate the problem (20, 21) but may reduce the number of cases substantially.

Stratification by distance between hospital and residence or by referral status might be an effective alternative.

**Berkson's bias.** If the study exposure is related to the risk of being hospitalized for the control disease, the exposure distribution in the series may not reflect the base. For example, diabetics are more likely to be admitted to the hospital with heart disease than are nondiabetics, which could bias studies focusing on diet. This is an example of Berkson's bias, which is caused by selection of subjects into a study differentially on factors related to exposure (32).

### Composition of a hospital control series

We believe that the best strategy regarding the selection of diseases to form a hospital- or disease registry-based control group is to exclude from the control series all conditions likely to be related to exposure (20, 33). The payoff for the extra effort in the study design will be more confidence in the validity of the results. If there is an association, subjects admitted to the hospital for the disease need to be excluded from the control series (34); however, a previous history of the disease should not be grounds for exclusion, unless the exclusion is also applied to cases (35). These exclusion rules apply regardless of whether the association is positive or negative, causal or not.

In theory, a possible association of exposure with a control disease should be assessed after controlling for confounders included in the analysis. If adjustment for a confounder eliminates a crude association between exposure and a potential control disease, adjustment for that same confounder in the analysis of the study will eliminate the bias caused by using that disease as a source of controls.

Similarly, an association between the control disease and a confounder is acceptable, if the effect of the confounder is controlled in the analysis. Also, patients with any disease that cannot be clearly distinguished from the study disease should be excluded

from the control series to reduce bias due to misclassification of disease.

If there is complete confidence that a single disease is unrelated to the exposure of interest, the entire control series may be selected from among patients with that disease. However, only rarely is there convincing evidence that the assumption of independence of the study exposure and a control disease is satisfied. Therefore, inclusion of patients with several diseases minimizes potential bias if any one disease turns out to be related to exposure (36, 37). When related diseases, such as other cancers for a study of a particular cancer or other perinatal outcomes for a study of birth defects, are used, the possibility of information bias may be reduced (38, 39). Again, however, any of the diseases that are related to the exposure (based on *a priori* knowledge) should be excluded (33). Overall, we recommend using more diseases rather than fewer; this protects the investigators if later evidence links one or more of the control diseases positively or negatively to an exposure.

### CONTROLS FROM A MEDICAL PRACTICE

Choosing controls from the primary medical practice of the cases can be a useful strategy when it is otherwise difficult to find controls who are comparable to cases on access to medical care or referral to specialized clinics. For example, medical practice controls may be more appropriate than hospital controls when cases are drawn from an urban teaching hospital, because potential subjects admitted to this hospital may be mixtures of poor clinic patients and high-socioeconomic-status private patients in far different ratios from those of the case series.

Controls selected from the same medical practices as the cases are drawn from the appropriate secondary base (1, 20), if one can make the assumption that two patients in the same primary care practice with the same presentation would follow the same pathway through the medical care system. A disadvantage of medical practice controls is

the extra complexity entailed by a random selection process for controls within several different practices.

The study base principle can be jeopardized with medical practice controls, since the exposure distribution for controls may not be the same as that in the study base, as when patients choose a physician for reasons relating to particular conditions that are themselves related to exposure. Similarly, if those conditions lead to modification of exposure subsequent to symptoms or treatment, the study base principle can be violated, unless the timing of the changes can be ascertained. For example, medical practice controls were used in a study that reported a positive association between coffee drinking and pancreatic cancer risk (40). Because the controls were patients with gastrointestinal disorders, some of them had conditions that were either caused by coffee drinking or treated by removing coffee from the diet, and thus the level of coffee drinking was not representative of the study base. The magnitude of the bias introduced by inclusion of such controls is dependent on the proportion of controls with diet-altering conditions and the relations of these diseases to coffee consumption (41). Control conditions associated with a confounder do not need to be excluded, if there will be adjustment for the confounder in the analysis. Thus, smoking-related conditions not related to coffee drinking might be included in the control series (34).

## FRIEND CONTROLS

Friends of cases may be a more convenient and inexpensive source of controls than are other alternatives. Controls can be selected from a list of friends or associates obtained from the case at little extra effort while the case is being interviewed. Friends may be likely to use the medical system in similar ways. Moreover, biases due to social class are reduced since usually the case and friend control will be of a similar socioeconomic background.

Nonetheless, we have strong reservations

about the use of friend controls. A possible theoretical justification of friend controls is that the base is divided into mutually exclusive "friendship strata" and that the exposure of a friend control is representative of that of the friendship stratum. Alternative justifications for friend controls are as a nonrandom sample from the base, if friends will all be in the same study base, or, if not, as an indirect way to probe the base (1, 20). None of these rationales is very persuasive. It is unrealistic to believe that the study base is divided into mutually exclusive friendship strata and that the controls are selected from only within the case's stratum (42). Even if this were true, the control selection within the stratum is deterministic and possibly related to exposure (1, 42). The credibility of representativeness of exposure is low for factors related to sociability, such as gregariousness or, possibly, smoking, diet, or alcohol consumption, because sociable people are more likely to be selected as controls than are loners (42).

"Friendly control" bias was suspected in a case-control study of patients with insulin-dependent diabetes mellitus, where friend controls were designated by the parents of cases (43). Insulin-dependent diabetes mellitus cases were found to be more likely to have learning problems, to have few friends, to dislike school, and to have recent illness in the family. While these findings could be due to true risk factors or to recall bias, they are more likely to be due to selection bias, since children perceived to have problems may be less likely to be identified as friends and, therefore, as controls (44). Further, there was some evidence suggesting that parents gave names of children from families with "mainstream" social characteristics (44).

A less serious problem is that the use of friend controls can lead to overmatching, since friends tend to be similar with regard to life-style and occupational exposures of interest, as in a study of head trauma and seizures (45); the loss of efficiency due to overmatching depends on how strongly head trauma is correlated among friends (e.g.,

motorcycle racers and boxers) and how closely it is related to gregariousness.

These problems can be alleviated to some extent by asking cases for the names of several friends and choosing controls randomly from the list or by asking for names of associates rather than friends (31, 46) so that the control will not be the case's closest, and perhaps most sociable, friend. However, those on the list will still tend to be more sociable than is a loner who is not on anyone's list (but can become a case), and there is no reason to believe that the extra friends named will have different characteristics from those who would be named on a shorter list (47). In addition, some cases may not be willing to provide names of friends (48), increasing nonresponse.

Despite serious shortcomings, friend controls may be useful in some exceptional circumstances, such as in a study of exposures unrelated to friendship characteristics, as is likely in a study of a genetically determined metabolic disorder (48, 49).

## RELATIVE CONTROLS

The choice of relative controls is motivated by the deconfounding principle, not the study-base principle (1). When genetic factors confound the effect of exposure, blood relatives of the case have been used as a source of controls (50) in an attempt to match on genetic background. Spousal and sibship relationships form strata and meet the reciprocity requirement (1, 42), but the theoretical justification for other relatives is more tenuous. Spouses might be a suitable control group if matching on adult environmental risk factors is sought. When sibling controls are used in studies of the association between genetic markers and the risk of cancer, confounding by factors related to ethnicity is minimized (51); however, cases and controls may be overmatched on a variety of genetic and environmental factors that are not risk factors but are related to the exposure under study. For example, effects of risk factors associated with family size cannot be assessed in a study using

sibling controls because of overmatching (31).

Trade-offs in using relative controls are illustrated in a study of the association between tonsillectomy and Hodgkin's disease (52) that used two control groups, siblings and spouses, to control for socioeconomic status in childhood and adulthood, respectively. A higher risk for tonsillectomy was found with spousal controls than with sibling controls, suggesting either positive confounding by childhood socioeconomic status or negative confounding by adult socioeconomic status.

## THE CASE SERIES AS THE SOURCE OF CONTROLS

An individual can serve as his own control for a study of an acute event when the effect of an exposure is transient (53), such as the effect of a possible triggering activity on myocardial infarction. The impact of the exposure is evaluated by comparing the proportions of events occurring during the putative period of elevated risk and the proportions of time each individual has been at elevated risk. This "case-crossover" design (53) can be thought of as a case-control design where each stratum consists of a single individual (or as a cohort study with many noninformative strata). The study base principle is clearly satisfied. Although there is no possibility of between-subject confounding, a second exposure that tends to occur at the same time or at different times from the study exposure can cause confounding (53). This design has the advantages that only patients need to be studied (53) and that recurrences can be handled easily.

However, for studies of chronic diseases where the main focus is on more stable time-dependent covariates, the use of a study series of cases only, as might be found in a disease registry, requires a complete and accurate exposure history and the strong assumption that the exposure of interest is unrelated to overall mortality (54). This

study design may also have lower power than more conventional studies (54).

### **PROXY RESPONDENTS AND DECEASED CONTROLS**

Interviews with proxy respondents are often used when subjects are deceased or too sick to answer questions or for persons with perceptual or cognitive disorders (55). Because proxy respondents will tend to be used more often for cases than for healthy controls, violation of the comparable accuracy principle is likely. Surrogates, particularly spouses and children, generally provide accurate responses for broad categories of exposure information, although more detailed information is usually less reliable (56–60). For some variables, such as cigarette smoking, and consumption of coffee and alcohol, spouses and children are remarkably accurate, even when compared with reinterviewed living subjects (61, 62). Proxies may even provide better information than the index subjects, such as in nutritional studies among older subjects, where a subject's wife may have prepared much of her husband's food (63).

When feasible, reducing the time interval between diagnosis and interview of cases can reduce the number of proxy interviews required. When information is obtained from a surrogate because the case is dead, using a living control sampled properly from the base can violate the comparable accuracy principle. However, insisting on a dead control (64) violates the study base principle, since the base consists of living subjects and subjects who die represent a special sample from that base. In order to use dead controls, one needs to assume representativeness of exposure (1), that the dead controls have the same distribution of exposure variables as does the base. This assumption has been demonstrated to be incorrect for a number of personal behavior variables, including use of tobacco and alcohol (65), even after deaths from causes believed to be associated with the study exposure are excluded (66).

Interviews with surrogates of appropri-

ately selected living controls do not make accuracy fully comparable since the controls are still alive while the cases are deceased, and responses by their surrogates may be influenced by factors associated with the subject's death (67). Nonetheless, validation studies (68, 69), in which the responses of a proportion of the living subjects and their proxies are obtained, can be used to reduce the bias due to errors from proxy responses.

In studies with deceased cases, the use of proxy interviews for appropriately selected live controls is usually preferable to the use of dead controls, particularly if the study exposure is likely to be associated with overall mortality. The advisability of insisting on a proxy interview for a live control depends on what information will be obtained from the interview. When exposure is assessed directly, comparable accuracy for cases and controls in an interview designed primarily to elicit information on confounders does not necessarily reduce the bias in the estimate of the effect of exposure (1, 67, 70); therefore, a proxy interview for the control may not help. Using proxy interviews for live controls should be considered only when 1) information about a key study exposure is to be obtained by interview *and* 2) a proxy report for the case is likely to be substantially less accurate than the control's self-report about the key study exposure.

### **Controls in proportionate mortality studies**

A proportionate mortality study can be viewed as a case-control study with controls obtained from a registry consisting of deaths (71). The underlying assumption is that the distribution of the exposure under study among subjects who died from other causes is the same as that in the base, which consists of living persons only. Just as in other registry-based studies, deaths from causes related to the study exposure must be excluded from the control series (21). Thus, this kind of study may not be suitable for investigating exposures such as smoking that are risk factors for causes of death repre-

senting a high proportion of mortality. One advantage of the approach is that a roster for the eligible controls can be established conveniently; any absences from the base typically will not lead to selection bias, since the efficiency of the system for registering deaths from most causes is unlikely to vary substantially with cause of death. However, errors in attribution of cause of death do occur, for example for AIDS, suicide, or cirrhosis of the liver, resulting in misclassification bias and over- or underexclusion of subjects.

## NUMBER OF CONTROL GROUPS

Some researchers have suggested choosing more than one control group (72, 73). It certainly is reassuring when the results are concordant across control series. However, when the results are discordant (25, 27, 52), the investigators must decide which result is "correct" and essentially discard the other. We therefore believe that doubt is not a good basis for choosing an additional control group. Rather, the best strategy usually is to decide which series is preferable at the design stage.

Multiple control groups might be helpful when each serves a different purpose, as when each control group provides the ability to control for a particular confounder. In this situation, the second control group can act as a form of replication.

## REFERENCES

- Wacholder S, McLaughlin JK, Silverman DT, et al. Selection of controls in case-control studies. I. Principles. *Am J Epidemiol* 1992;135:1019-28.
- Wacholder S, Silverman DT, McLaughlin JK, et al. Selection of controls in case-control studies. III. Design options. *Am J Epidemiol* 1992;135:1042-50.
- Graubard B, Fears TR, Gail MH. Effects of cluster sampling on epidemiologic analysis in population-based case-control studies. *Biometrics* 1989;45:1053-71.
- Cole P. Introduction. In: Breslow NE, Day NE, eds. *Statistical methods in cancer research. Vol 1. The analysis of case-control studies*. Lyon: International Agency for Research on Cancer, 1980:14-40. (IARC scientific publication no. 32).
- Whittemore AS. Estimating attributable risk from case-control studies. *Am J Epidemiol* 1983;117:76-85.
- Gail M, Lubin JH, Silverman DT. Elements of design in epidemiologic studies. In: DeLisi C, Eisenfeld J, eds. *Statistical methods in cancer epidemiology*. Amsterdam: North Holland Publishing Co, 1985:313-23.
- Savitz DA, Pearce N. Control selection with incomplete case ascertainment. *Am J Epidemiol* 1988;127:1109-17.
- Cole P, Monson RR, Haning H, et al. Smoking and cancer of the lower urinary tract. *N Engl J Med* 1971;284:129-34.
- Adami HO, Rimsten A, Stenkvist B, et al. Reproductive history and risk of breast cancer. *Cancer* 1978;41:747-57.
- Gold EB, Diener MD, Szklo M. Parental occupations and cancer in children. *J Occup Med* 1982;24:578-84.
- Hatten J. Medicare's common denominator: the covered population. *Health Care Finan Rev* 1980;2:53-64.
- Greenberg ER. Random digit dialing for control selection. A review and a caution on its use in studies of childhood cancer. *Am J Epidemiol* 1990;131:1-5.
- Waksberg J. Sampling methods for random digit dialing. *J Am Stat Assoc* 1978;73:40-6.
- Harlow BL, Davis S. Two one-step methods for household screening and interviewing using random digit dialing. *Am J Epidemiol* 1988;127:857-63.
- Harlow B, Harge P. Telephone household screening and interviewing. *Am J Epidemiol* 1983;117:632-3.
- Thornberry TO, Massey JT. Trends in United States telephone coverage across time and subgroups. In: Groves RM, Biemer PB, Lyberg LE, et al., eds. *Telephone survey methodology*. New York: John Wiley & Sons, Inc, 1988:25-49.
- Ward EM, Kramer S, Meadows AT. The efficacy of random digit dialing in selecting matched controls for a case-control study of pediatric cancer. *Am J Epidemiol* 1984;120:582-91.
- Robison LL, Daigle A. Control selection using random digit dialing for cases of childhood cancer. *Am J Epidemiol* 1984;120:164-6.
- Mohadjer L. Stratification of prefix areas for sampling rare populations. In: Groves RM, Biemer PB, Lyberg LE, et al., eds. *Telephone survey methodology*. New York: John Wiley & Sons, Inc, 1988:161-73.
- Miettinen OS. *Theoretical epidemiology: principles of occurrence research in medicine*. New York: John Wiley & Sons, Inc, 1985.
- Miettinen OS. The "case-control" study: valid selection of subjects. *J Chronic Dis* 1985;38:543-8.
- Cohen BH. Family patterns of longevity and mortality. In: Neel JV, Shaw MW, Schull WJ, eds. *Genetics and the epidemiology of chronic diseases*. Washington, DC: US Department of Health, Education, and Welfare, 1963:237-63. (DHEW publication no. 1163).
- Massey FJ, Bernstein FS, O'Fallon WM, et al. Vasectomy and health. *JAMA* 1984;252:1023-9.
- Vernick LJ, Vernick SL, Kuller KH. Selection of

- neighborhood controls: logistics and fieldwork. *J Chronic Dis* 1984;37:177-82.
25. Horwitz RI, Feinstein AR. Alternative analytic methods for case-control studies of estrogens and endometrial cancer. *N Engl J Med* 1978;299:1089-94.
  26. Hutchison GB, Rothman KJ. Correcting a bias. *N Engl J Med* 1978;299:1129-30.
  27. Hulka BS, Grimson RC, Greenberg BG, et al. "Alternative" controls in a case-control study of endometrial cancer and exogenous estrogen. *Am J Epidemiol* 1980;112:376-87.
  28. Rothman KJ. Modern epidemiology. Boston: Little, Brown & Company, 1986.
  29. Nelson LM, Franklin GM, Hamman RF, et al. Referral bias in multiple sclerosis research. *J Clin Epidemiol* 1988;41:187-92.
  30. Brown LM, Pottier LM, Hoover RN. Prenatal and perinatal risk factors for testicular cancer. *Cancer Res* 1986;46:4812-16.
  31. MacMahon B, Pugh TF. Epidemiology: principles and methods. Boston: Little, Brown & Company, 1970.
  32. Flanders WD, Boyle CA, Boring JR. Bias associated with differential hospitalization rates in incident case-control studies. *J Clin Epidemiol* 1989;42:395-401.
  33. Wacholder S, Silverman DT. Re: "Case-control studies using other diseases as controls: problems of excluding exposure-related diseases." (Letter). *Am J Epidemiol* 1990;132:1017-18.
  34. MacMahon B, Yen S, Trichopoulos D, et al. Coffee and cancer of the pancreas. (Letter). *N Engl J Med* 1981;304:1605-6.
  35. Lubin JH, Hertge P. Excluding controls: misapplications in case-control studies. *Am J Epidemiol* 1984;120:791-3.
  36. Jick H, Vessey MP. Case-control studies in the evaluation of drug-induced illness. *Am J Epidemiol* 1978;107:1-7.
  37. Axelson O. The case-referent study—some comments on its structure, merits, and limitations. *Scand J Work Environ Health* 1985;11:207-13.
  38. Linet MS, Brookmeyer R. Use of cancer controls in case-control cancer studies. *Am J Epidemiol* 1987;125:1-11.
  39. Smith AH, Pearce NE, Callas PW. Cancer case-control studies with other cancers as controls. *Int J Epidemiol* 1988;17:298-306.
  40. MacMahon B, Yen S, Trichopoulos D, et al. Coffee and cancer of the pancreas. *N Engl J Med* 1981;304:630-3.
  41. Silverman DT, Hoover RN, Swanson GM, et al. The prevalence of coffee drinking among hospitalized and population-based control groups. *JAMA* 1983;249:1877-80.
  42. Robins J, Pike M. The validity of case-control studies with nonrandom selection of controls. *Epidemiology* 1990;1:273-84.
  43. Siemiatycki J, Colle S, Campbell S, et al. Case-control study of insulin-dependent (type I) diabetes mellitus. *Diabetes Care* 1989;12:209-16.
  44. Siemiatycki J. Friendly control bias. *J Clin Epidemiol* 1989;42:687-8.
  45. Hochberg F, Toniolo P, Cole P. Head trauma and seizures as risk factors of glioblastoma. *Neurology* 1984;34:1511-14.
  46. Kelsey JL, Thompson WD, Evans AS. Methods in observational epidemiology. New York: Oxford University Press, 1986.
  47. Thompson WD. Nonrandom yet unbiased. *Epidemiology* 1990;1:262-5.
  48. Shaw GL, Tucker MA, Kase RG, et al. Problems ascertaining friend controls in a case-control study of lung cancer. *Am J Epidemiol* 1991;133:63-6.
  49. Flanders WD, Austin H. Possibility of selection bias in matched case-control studies using friend controls. *Am J Epidemiol* 1986;124:150-3.
  50. Goldstein AM, Hodge SE, Haile RW. Selection bias in case-control studies using relatives as the controls. *Int J Epidemiol* 1989;18:985-9.
  51. Petrakis NL, King MC. Genetic markers and cancer epidemiology. *Cancer* 1977;39:1861-6.
  52. Gutensohn N, Li F, Johnson R, et al. Hodgkin's disease, tonsillectomy, and family size. *N Engl J Med* 1975;292:22-5.
  53. Maclure M. The case-crossover design: a method for studying transient effects on the risk of acute events. *Am J Epidemiol* 1991;133:144-53.
  54. Prentice RL, Vollmer WM, Kalbfleisch JD. On the use of the case series to identify disease risk factors. *Biometrics* 1984;40:445-58.
  55. Nelson LM, Longstreth WT Jr, Koepsell TD, et al. Proxy respondents in epidemiologic research. *Epidemiol Rev* 1990;12:71-86.
  56. Rogot E, Reid DD. The validity of data from next of kin in studies of mortality among migrants. *Int J Epidemiol* 1951;4:51-4.
  57. Kolonel LN, Hirohata T, Nomura AMY. Adequacy of survey data collected from substitute respondents. *Am J Epidemiol* 1977;106:476-84.
  58. Marshall J, Priore R, Haughey B, et al. Spouse-subject interviews and the reliability of diet studies. *Am J Epidemiol* 1980;112:675-83.
  59. Lerchen ML, Samet JM. An assessment of the validity of questionnaire responses provided by a surviving spouse. *Am J Epidemiol* 1986;123:481-9.
  60. Blot WJ, McLaughlin JK. Practical issues in the design and conduct of case-control studies: use of next-of-kin interviews. In: Blot WJ, Hirayama T, Hoel DG, eds. Statistical methods in cancer epidemiology. Hiroshima: Radiation Effects Research Foundation, 1985:49-62.
  61. McLaughlin JK, Dietz MS, Mehl ES, et al. Reliability of surrogate information on cigarette smoking by type of informant. *Am J Epidemiol* 1987;126:144-6.
  62. McLaughlin JK, Mandel JS, Mehl ES, et al. Reliability of next-of-kin and self-respondents for cigarette, coffee, and alcohol consumption. *Epidemiology* 1990;1:408-12.
  63. Samet JM. Surrogate sources of dietary information. In: Willett W, ed. Nutritional epidemiology. New York: Oxford University Press, 1990:133-42.
  64. Gordis L. Should dead cases be matched to dead controls? *Am J Epidemiol* 1982;115:1-5.
  65. McLaughlin JK, Blot WJ, Mehl ES, et al. Problems in the use of dead controls in case-control studies. I. General results. *Am J Epidemiol* 1985;121:131-9.
  66. McLaughlin JK, Blot WJ, Mehl ES, et al. Problems in the use of dead controls in case-control studies. II. Effect of excluding certain causes of death. *Am*

- J Epidemiol 1985;122:485-94.
- 67. Walker AM, Velema JP, Robins JM. Analysis of case-control data derived in part from proxy respondents. Am J Epidemiol 1988;127:905-14.
  - 68. Armstrong BG, Whittemore AS, Howe GR. Analysis of case-control data with covariate measurement error: application to diet and colon cancer. Stat Med 1989;8:1151-65.
  - 69. Armstrong BG. The effects of measurement errors on relative risk regressions. Am J Epidemiol 1990; 132:1176-84.
  - 70. Greenland S. The effect of misclassification in the presence of covariates. Am J Epidemiol 1980;112: 564-9.
  - 71. Miettinen OS, Wang J-D. An alternative to the proportionate mortality ratio. Am J Epidemiol 1981;114:144-8.
  - 72. Ibrahim MA, Spitzer WO. The case-control study: the problem and the prospect. J Chronic Dis 1979; 32:139-44.
  - 73. The case-control study. (Editorial). Br Med J 1979; 2:884-5.

## Selection of Controls in Case-Control Studies

### III. Design Options

Sholom Wacholder,<sup>1</sup> Debra T. Silverman,<sup>1</sup> Joseph K. McLaughlin,<sup>1</sup> and Jack S. Mandel<sup>2</sup>

Several design options available in the planning stage of case-control studies are examined. Topics covered include matching, control/case ratio, choice of nested case-control or case-cohort design, two-stage sampling, and other methods that can be used for control selection. The effect of potential problems in obtaining comparable accuracy of exposure is also examined. A discussion of the difficulty in meeting the principles of study base, deconfounding, and comparable accuracy (S. Wacholder et al. *Am J Epidemiol* 1992;135:1019-28) in a single study completes this series of papers. *Am J Epidemiol* 1992;135:1042-50.

bias (epidemiology); epidemiologic methods; retrospective studies

In our previous papers, we presented basic principles of control selection (1) and discussed different kinds of control groups (2). This paper addresses some of the other decisions involved in control selection, of which the major themes are issues of stratification and efficiency and the effects of time. The principles of deconfounding and efficiency are the main concerns in considering some special sampling techniques, including matching, cluster sampling, and two-stage sampling. Efficiency is paramount in our discussion of the control/case ratio, replacement of controls, and using a single control group for multiple case series. Consideration of the time of membership in the study base is crucial in the discussion of nested case-control, case-cohort, and case-base designs.

### MATCHING

Random sampling from the study base, where controls are chosen independently of characteristics of the cases, is the simplest strategy for control selection. Matching is an option that sometimes can improve efficiency in the estimation of the effect of exposure by protecting against the situation where the distributions of a confounder are substantially different in cases and controls (3). However, the improvement is typically small (3), except for strong confounders. There are several other reasons to match.

*Control of unmeasured confounders.* Identifiable but not quantifiable variables with many categories, such as neighborhood or telephone exchange, can serve as proxies for environmental or socioeconomic confounding factors that are difficult to measure (4). Matching on such a variable may balance cases and controls with respect to unknown confounders. Use of the cases' identical twins as controls is an extreme form of this kind of matching.

*Power.* Matching can ensure that there are sufficient controls to estimate an effect in a particular subgroup or to identify an interaction (5). For example, matching on smoking instead of choosing controls indepen-

Received for publication May 8, 1991, and in final form February 11, 1992.

<sup>1</sup> Biostatistics Branch, National Cancer Institute, Bethesda, MD.

<sup>2</sup> Department of Environmental and Occupational Health, School of Public Health, University of Minnesota, Minneapolis, MN.

Reprint requests to Dr. Sholom Wacholder, Biostatistics Branch, National Cancer Institute, 6130 Executive Blvd., EPN 403, Rockville, MD 20892.

dently of smoking could make it easier to find an interaction between the exposure and smoking if smoking had a large effect on risk and was rare in the population, by ensuring a sufficient number of smoking controls. The two-stage design, a generalization of matching discussed below, can also be used to achieve this goal.

**Time comparability.** In unmatched studies, it can be difficult to achieve time comparability between cases and controls for exposures that vary over time. Matching on time-related variables provides a simple reference point for variables based on these exposures.

**Feasibility.** Matching may be the most feasible method of obtaining controls. For example, when a case is defined as a perinatal death, the next live birth can be chosen as a control (6). Use of a random sample of live births from a computer tape of births would delay the interview of controls, leading to possible violation of the comparable accuracy principle as well as possibly increasing nonresponse.

**Completeness of control for confounding.** Perfect matching, followed by a matched analysis, results in complete control for a continuous confounder under a multiplicative model of the joint effects. Alternative strategies, such as regression adjustment for the confounder, can result in bias if its effect is misspecified, e.g., if linearity is wrongly assumed. Categorization may leave some residual confounding, but this is of little importance unless there is a substantial gradient in risk within strata (7).

On the other hand, matching has several disadvantages.

**Cost.** Matching can add cost and complexity to a sampling scheme by requiring extra effort to recruit controls (5).

**Exclusion of cases.** Matching can result in the exclusion of cases when no matched control can be found (8, 9), particularly when matching on several variables.

**Longer study duration.** Matching will slow down a study when control selection must wait for cases to be identified and for complex matching variables to be obtained for cases and potential controls.

**Reduced flexibility in analysis.** Stratified or matched analyses can be considered, even when there was no matching or stratification in the design. But matching at the design stage reduces the investigator's flexibility during the analysis. For example, over-matching that occurs in the design cannot be corrected in the analysis. Furthermore, matching usually precludes the ability to directly estimate or test the effect of the matching variable as a risk factor and the fitting of a nonmultiplicative risk model involving the matching variable and the exposure (5). However, in the multiplicative model, it is possible to fit interactions with the matching variables. Below, we discuss a method that can be used to estimate the effect of matching variables and to fit nonmultiplicative models, as long as the values of the matching variables are retained for subjects who are identified for the study but excluded because of the matching criteria.

### Variables on which to match

Matching reduces the possibility of severe loss of efficiency due to a major discrepancy in the empiric distributions of a strong risk factor between cases and controls. Matching should be considered only for risk factors whose confounding effects need to be controlled for but that are not of scientific interest as independent risk factors in the study. Matching on variables that are unrelated to risk of disease is pointless; it can only reduce a study's efficiency (4). Age, sex, and race are often used as matching variables because they are usually strong confounders and because their effects are usually well-known from descriptive epidemiology (10).

### Forms of matching and stratification

One form of matching is individual matching where a selected control must have exactly or approximately the same value of the matching factor as the corresponding case. Frequency matching or quota matching results in equal distributions of the matching factors in the cases and the selected controls. For these forms of matching, the control cannot be recruited until the case

is identified. Approximate frequency matching can begin immediately; it uses the anticipated, rather than the actual, case distribution and thereby allows the control selection process to operate independently of the case selection process. However, if some of the matching strata are extremely small, approximate frequency matching can be wasteful (11), since the control/case ratio will vary. Probability matching (12) defines strata based on the matching variables. A random mechanism is used to select eligible subjects, with the probabilities of inclusion for each subject determined by the investigators based on the odds ratios for disease associated with the subject's stratum. This approach does not require knowledge of the exposure distribution in the cases and allows for a more informative analysis (13), as discussed below.

### **Overmatching**

We use the term "overmatching" to refer to matching that is counterproductive, by either causing bias or reducing efficiency (14). Matching on an intermediate variable in a causal pathway between exposure and disease can bias a point estimate downward (7), since the exposure's effect on disease, adjusting for (conditional on) intermediate variable, is less than the unadjusted effect. For example, matching on presence of endometrial hyperplasia in a study of the relation between estrogen and endometrial cancer is overmatching leading to bias (14, 15). Why? The parameter we seek to estimate is a measure of the impact on disease risk of a change in the level of exposure. Matching on endometrial hyperplasia effectively restricts the comparisons of exposure to subjects concordant on presence of hyperplasia; this does not allow the full impact of estrogen on cancer risk to be assessed, since presence of hyperplasia itself is strongly influenced by estrogen use. Matching on a factor that is a surrogate for or a consequence of disease or matching on a correlate of an imperfectly measured exposure (15) can also lead to overmatching and bias.

The other main form of overmatching can reduce the efficiency of a study by restricting the variability of an exposure that is correlated with the exposure under study (16). This form of "overmatching" can occur even when matching per se was not used in the selection of controls, as when an overly homogeneous base is used for the study. Miettinen and Cook (17) note that the use of a variable indicating the presence of yellow fingers, presumed to be related to smoking but unrelated to risk of lung cancer (after controlling for smoking), would be an example of overmatching. There is much less variability in smoking, conditional on presence of yellow finger, than unconditionally; since having yellow fingers is not a risk factor, it does not affect the point estimate but does reduce efficiency.

### **RATIO OF CONTROLS TO CASES**

Determination of the number of controls to be selected is another important design decision. It is useful to consider the ratio of controls to cases. There is usually little marginal increase in precision from increasing the ratio of controls to cases beyond four (18), except when the effect of exposure is large (19). In general, the best way to increase precision in a case-control study is to increase the number of cases by widening the base geographically or temporally rather than by increasing the number of controls, because the marginal increase in precision from an additional case is greater than from an additional control (assuming there are already more controls than cases in the study). In matched and stratified studies, the most efficient allocation of a fixed number of controls into strata is usually one that sets the ratio of controls to cases to be approximately equal (4).

### **REPLACEMENT OF CONTROLS**

Controls who refuse to participate in a study should sometimes be replaced on efficiency grounds, as when replacement can prevent wasting a case who otherwise would have no matched control. However, subjects who refuse to participate in case-control

studies may have a different exposure distribution from those who do participate. Replacing refusals will not increase the validity of a study, since refusals will still be excluded.

The situation is different when information on the primary exposure, perhaps obtained from medical records, is available, but some information on a confounder (perhaps obtained by interview) is not. Then, the impact of excluding the control is probably more serious than that of the missing information regarding the confounder, and the control should not be replaced. When controls are replaced, reported response rates should reflect the actual percentage of eligible subjects who refuse to participate.

### **ONE CONTROL GROUP FOR SEVERAL DISEASES**

Use of a single control group for more than one case series can lead to savings of money and effort (20–22). Systematic errors in assembling the control series would presumably affect each individual series equally, but the availability of a larger number of controls would increase the precision of point estimates. While the use of the same controls for different diseases induces some dependence in the estimates of effect for different diseases (23), no special analysis is required, except when *comparing* the risk factors for the various diseases (24). In fact, this strategy can have another advantage; i.e., it can help to calibrate the control series by identifying exposures having stronger (or weaker) than expected associations with several diseases, resulting from special characteristics of the control group. The fact that the same control series was used for several diseases should be discussed in the reports from the studies, so that readers can judge whether findings resulted from the unique characteristics of the control series.

### **NESTED CASE-CONTROL AND CASE-COHORT STUDIES**

#### **Controls in nested case-control studies**

A particular form of case-control study that, in fact, does have a roster of subjects available for control selection is the nested

case-control study or case-control within a cohort study (19, 25, 26). This design paradigmatically satisfies the study base principle since the base is the cohort as it moves through time. Typically, for each case, a set of controls is selected from subjects at risk at the time of disease occurrence of the case. The matching in the design allows for tight control of the confounding effects of time in the analysis. Thus, this design is useful when close matching on time is required, as in studies of the incidence of a rare disease, such as cancer.

Just as in the calculation of person-years in a cohort study, membership in the base depends on time (1, 4). A subject is in the base while under follow-up, i.e., when the subject would be enrolled as a case in the study upon development of disease. So a subject cannot be selected before entry to the cohort, after loss to follow-up, after death, or after becoming a case (unless subsequent occurrences of disease would make the subject a case again, as, perhaps, in a case-control study of ear infection in young children). In nested case-control studies with age matching, a subject chosen to be a control for a case at a given age should not be excluded from the set of controls because of subsequent development of disease. Thus, a control who subsequently develops disease can also serve as a case (27).

If the times when a subject is actually in the base are available with the roster, selection of controls from the base is simply a matter of sampling. Sampling can be stratified according to factors available for all members of the roster at the time of entry to the cohort. Control selections at the various times of diagnosis of the cases should be mutually independent and should not be influenced by future disease status of the subject or by use as a control for another case (27–29); thus, the same individual can serve as a control for more than one case (27, 28). These rules mirror the approach in the analysis of the proportional hazards model with time-to-event data, where virtually all cases served previously as “controls” and virtually all controls are used more than once (30).

## Controls in case-cohort studies

The case-cohort design is an alternative to the nested case-control design with a simpler sampling scheme but a more complex analysis (24, 31, 32). In its simplest form, a sub-cohort or random sample from all members of the cohort is selected to be the source of all controls. Adjustment for the confounding effects of time is achieved in the analysis, by comparing the exposures of each case to those of a set of controls consisting of all members of the subcohort who were at risk at the time of diagnosis of the case. We think of a case-cohort study as a variant of a nested case-control study where controls are selected without matching on time (24, 32). The following several advantages of the case-cohort study are due to the use of "unmatched" controls.

**External comparisons.** It is easy to obtain estimates comparing the risk in the cohort with that of an external population (32). For example, the case-cohort design has been used to compare the risk of breast cancer in a cohort of women who received noncontraceptive estrogen treatment with the risk for other women living in the same region (33).

**Ease of selection.** Sampling of controls can begin before the roster of subjects and the list of cases have been completely identified. As soon as each member of the roster is identified, randomization can be used immediately to determine inclusion in the subcohort (24).

**Multiple diseases.** Since there is no time matching of cases to controls, a single subcohort can serve as a source of controls for multiple disease types (20, 24, 31).

**Primary time scale.** Unlike the nested case-control design, the case-cohort design does not require a decision about the primary time scale until the analysis stage (24). Thus, for example, all analyses of a nested case-control design with age matching control for age in a study of a treatment-related second cancer, while controlling for age alone or time since first cancer alone is possible in a case-cohort design (24).

It is worth noting that for either design, there is a possibility of differential misclas-

sification if information about cases is obtained before that about controls (24, 34). Other practical considerations have recently been discussed (24). The statistical efficiency of the nested case-control design is often slightly higher than that for the case-cohort design when a single disease is being studied (34), except when there are small amounts of censoring and late entry (31). More refined approximations of the efficiency of nested case-control, case-cohort, and related designs can be obtained when the cohort has been assembled, i.e., when the cases and their event times, as well as the interval at risk for all subjects, are known, but exposure information is not yet available (29).

## CASE-BASE STUDIES

Controls are sampled for the case-base design (20, 35) in the same way as the subcohort is selected in the case-cohort design; it differs from the standard case-control design only in that control subjects are sampled from the base, regardless of their disease status. The case-base design can be thought of as a variant of the case-control design that allows estimation of the risk ratio, because the exposure odds in the base (not just in the cases and noncases as in the case-control study) can be estimated.

## ARE CASES ELIGIBLE TO BE CONTROLS?

We have noted several situations, including the nested case-control, case-cohort, and case-base designs, where subjects who qualify as cases can be included as controls. In general, a future case is in the base until his or her disease is diagnosed and, therefore, should not be excluded from the sampling of controls for a case diagnosed at an earlier time.

## TWO-STAGE SAMPLING

Recently, some two-stage sampling designs have been proposed as economical alternatives to standard case-control studies (13, 36-38). The savings accrue from not

requiring all exposure and confounder measurements for all subjects; sometimes a substantial reduction in expense can be achieved with little loss in statistical efficiency compared with the study with information on all subjects. In these designs, first-stage variables, i.e., exposure or confounder measurements that are relatively easy to obtain, are gathered for all subjects. The remaining, second-stage variables are obtained on only a subset of subjects, with the sampling fractions depending on disease status and the first-stage variables. The first-stage variable might be an exposure that would be obtained from a record, while the second-stage variable might be a confounder, such as smoking, which required a personal interview. Alternatively, the first-stage variable might be a confounder and the second-stage variable might be the exposure (13). This approach could be helpful, for example, in a study of the effect of residential radon exposure on lung cancer risk (12, 13). First-stage variables might include age and smoking. Nonsmoking cases and smoking controls will have higher probabilities of selection for the part of the study requiring expensive fieldwork for residential radon measurements. The power for assessing a radon-smoking interaction will be enhanced, compared with a matched or an unmatched design, by forcing the proportions of cases and controls in the smoking and nonsmoking strata to be near 0.5 (13, 37). A two-stage design can also be considered where the second-stage variable is a more refined version of the first-stage variable. For example, the probability of obtaining a detailed occupational history can be allowed to vary, depending on the subject's current job title.

The two-stage design proposed originally uses random samples of the subjects in each cell of the cross-classification of the first-stage variable and disease to determine which subjects to include for second-stage measurements (36–38). An alternative randomized recruitment approach uses randomization, with the probabilities, which are dependent on the approximate odds ratios (determined *a priori*) associated with the

subject's level of the first-stage variable (12, 13).

A two-stage approach can be more efficient than matching for estimating the main effects of exposures and interactions (37) and allows for estimation of the effects of first-stage variables, in contrast to standard matched studies (13). However, any matched study can be viewed and analyzed as a special case of the two-stage study, if information on the matching factor is retained for all eligible subjects, including those excluded because they did not satisfy the matching criteria. This allows estimation of the main effect of the matching variable as well as the fitting of nonmultiplicative models.

### **CLUSTER SAMPLING**

In cluster sampling, controls are selected in groups, to reduce expense, rather than independently (39). Choosing several controls who live in the same household or near one another can be economical when less effort is needed to include an additional member of the cluster than to include an independent control. Thus, cluster sampling might be appropriate for population control groups when blood samples are needed. The clusters themselves must be selected so that each member of the base population has an equal chance of being selected (1), and an analysis taking clustering into account must be used (39).

### **AVOIDING INFORMATION BIAS**

A widespread concern about interview-based case-control studies is that cases recall previous exposures differently than do controls. Cases may spend time thinking about possible reasons for their illness, may search their memories for past exposure or even exaggerate or fabricate exposure, or may try to deny any responsibility for the disease. Therefore, some suggest using control groups of diseased subjects in the name of equal accuracy (40). While accuracy of information and how that accuracy differs between cases and controls are considerations

in the choice of control group, one must also be concerned about choosing controls with conditions possibly related to exposure (2, 41, 42).

Unfortunately, the literature on the question of differential recall for cases and controls is sparse, and the interpretation of the published studies is difficult (43). Most recent empiric research suggests that differential accuracy does not cause serious distortion (40, 44–48). Empiric work on the accuracy of recall for a broader range of variables would help in the decision of what is the appropriate source of controls in situations when there is a suspicion of differential accuracy.

In many studies, information about time-dependent exposures (variables whose values can vary over time), such as consumption of food items or cigarette smoking, should be obtained in such a way that the entire history will be available. This is typically not practical, and, instead, questions usually refer to a particular period of time. Unless the periods of time correspond for cases and controls, the comparable accuracy principle may be violated. If controls are matched to cases on age, questions about exposures should refer to exposures at the same age for cases and controls. In a diet and cancer study, if a case is asked about usual diet 5 years prior to the diagnosis of cancer at age 60, the questions to a perfectly age-matched control should refer to usual diet at age 55, regardless of the control's age at interview. For frequency-matched or unmatched studies, some sort of average might be attempted, such as starting exposure questions for all subjects with, "Before 1985, . . ." This should be the practice even if a control, who was selected from among those free of disease at age 60, is not interviewed until age 63. (Of course, one has to assume that the respondents are answering the questions the way they are asked.) The time intervals between interviews of cases and of controls should be similar (and as short as possible), so the elapsed times from the period to which the questions refer in the interview will be similar in cases and controls; also, any secular trends in exposure

prevalence would be less likely to cause bias (49, 50). Similarly, matching on calendar time should be considered, if it can ensure that exposure measurements are comparable with respect to time, as in case-control studies performed in an occupational setting, where industrial hygiene data from different years may be affected by changes in the quality of the measuring instruments.

## DISCUSSION

In the first paper of this series (1), we presented four basic principles—study base, deconfounding, comparable accuracy, and efficiency—that we believe provide a theoretical framework for the evaluation of issues in control selection. Various practical problems have been addressed, and possible solutions have been examined using these principles.

It may be difficult or impossible to satisfy all principles in a study. Sometimes an attempt that is feasible turns out to be harmful. Just as unnecessary matching can reduce efficiency and even cause bias, avoiding violations of principles that are purely theoretical and have no effect on inference is not advisable. It is important to remember that the validity of a study can be undermined more by an equivocal violation of principle than by a clear violation of principle that results in only minor bias. Since all biases are not created equal, quantification of the extent of bias is important (51, 52); otherwise, an attempt to avoid a violation of one principle may induce a more serious violation of another. Therefore, the implications of alternative approaches need to be considered carefully. For example, allowing concerns about the theoretical possibility of recall bias to determine the type of controls to choose, when in fact little or no recall bias may exist, could lead to a more biased study, if controls were drawn from a diseased group related to the study exposure.

It is important to recognize that development of a protocol that deals with the theoretical considerations discussed here is not enough; careful fieldwork is needed to make sure the study is properly executed. Thus, a

low response rate, particularly when nonresponse might depend on exposure level, may violate the study-base principle and threaten a study's validity.

The "ideal" (53) control group rarely exists in epidemiologic studies. Besides additional theoretical work, empiric studies are needed to measure the impact of violations of the principles so intelligent trade-offs can be made when planning a study. We believe, however, that although proper control selection will continue to be problematic, the most serious mistakes in control selection can be avoided by keeping a few basic principles in mind.

## REFERENCES

1. Wacholder S, McLaughlin JK, Silverman DT, et al. Selection of controls in case-control studies. I. Principles. *Am J Epidemiol* 1991;135:1019-28.
2. Wacholder S, Silverman DT, McLaughlin JK, et al. Selection of controls in case-control studies. II. Types of controls. *Am J Epidemiol* 1991;135:1029-41.
3. Thomas DC, Greenland S. The relative efficiencies of matched and independent sample designs for case-control studies. *J Chronic Dis* 1983;36:685-97.
4. Miettinen OS. Theoretical epidemiology: principles of occurrence research in medicine. New York: John Wiley & Sons, Inc, 1985.
5. Thomas DC, Greenland S. The efficiency of matching in case-control studies of risk factor interactions. *J Chronic Dis* 1985;38:569-74.
6. Clarke M, Clayton D. The design and interpretation of case-control studies of perinatal mortality. *Am J Epidemiol* 1981;113:636-45.
7. Breslow NE, Day NE, eds. Statistical methods in cancer research. Vol 1. The analysis of case-control studies. Lyon: International Agency for Research on Cancer, 1980. (IARC scientific publication no. 32).
8. McKinlay SM. Pair matching—a reappraisal of a popular technique. *Biometrics* 1977;33:725-35.
9. Thompson WD, Kelsey JL, Walter SD. Cost and efficiency in the choice of matched and unmatched case-control study designs. *Am J Epidemiol* 1982;116:840-51.
10. MacMahon B, Pugh TF. Epidemiology: principles and methods. Boston: Little, Brown & Company, 1970.
11. McTiernan A, Thomas DB, Whitehead A, et al. Efficient selection of controls for multi-centered collaborative studies of rare diseases. *Am J Epidemiol* 1986;123:901-4.
12. Weinberg CR, Sandler DP. Randomized recruitment in case-control studies. *Am J Epidemiol* 1991;134:421-32.
13. Weinberg CR, Wacholder S. The design and analysis of case-control studies with biased sampling. *Biometrics* 1990;46:963-75.
14. Cole P. Introduction. In: Breslow NE, Day NE, eds. Statistical methods in cancer research. Vol 1. The analysis of case-control studies. Lyon: International Agency for Research on Cancer, 1980:14-40. (IARC scientific publication no. 32).
15. Breslow N. Design and analysis of case-control studies. *Annu Rev Public Health* 1982;3:29-54.
16. Day NE, Byar DP, Green SB. Overadjustment in case-control studies. *Am J Epidemiol* 1980;112:696-706.
17. Miettinen OS, Cook EF. Confounding: essence and detection. *Am J Epidemiol* 1981;114:593-603.
18. Ury HK. Efficiency of case-control studies with multiple controls per case: continuous or dichotomous data. *Biometrics* 1975;31:643-9.
19. Breslow NE, Lubin JH, Marek P, et al. Multiplicative models and cohort analysis. *J Am Stat Assoc* 1983;78:1-12.
20. Kupper LL, McMichael AJ, Spiertas R. A hybrid epidemiologic study design useful in estimating relative risk. *J Am Stat Assoc* 1975;70:524-8.
21. Siemiatycki J, Wacholder S, Richardson L, et al. Discovering carcinogens in the occupational environment. *Scand J Work Environ Health* 1987;13:486-92.
22. Greenland S. Control-initiated case-control studies. *Int J Epidemiol* 1985;14:130-4.
23. Thomas DC, Goldberg M, Dewar R, et al. Statistical methods for relating several exposure factors to several diseases in case-heterogeneity studies. *Stat Med* 1986;5:49-60.
24. Wacholder S. Practical considerations in choosing between the case-cohort and nested case-control design. *Epidemiology* 1991;2:155-8.
25. Mantel N. Synthetic retrospective studies and related topics. *Biometrics* 1973;29:479-86.
26. Liddell FDK, McDonald JC, Thomas DC. Methods of cohort analysis: appraisal by application to asbestos mining (with discussion). *J R Stat Soc A* 1977;140:469-91.
27. Lubin JH, Gail MH. Biased selection of controls for case-control analysis of cohort studies. *Biometrics* 1984;40:63-75.
28. Robins J, Gail MH, Lubin JH. More on biased selection of controls. *Biometrics* 1986;42:293-9.
29. Wacholder S, Gail MH, Pee D. Selecting an efficient design for assessing exposure-disease relationships in an assembled cohort. *Biometrics* 1991;47:63-76.
30. Cox DR. Regression models and life tables (with discussion). *J R Stat Soc B* 1972;34:187-220.
31. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986;73:1-11.
32. Wacholder S, Boivin J-F. External comparisons with the case-cohort design. *Am J Epidemiol* 1987;126:1198-1209.
33. Bergkvist L, Adami HO, Persson I, et al. The risk of breast cancer after estrogen and estrogen-progestin replacement. *N Engl J Med* 1989;321:293-7.
34. Langholz B, Thomas DC. Nested case-control and case-cohort methods of sampling from a cohort: a critical comparison. *Am J Epidemiol* 1990;131:169-76.

35. Miettinen OS. Design options in epidemiologic research: an update. *Scand J Work Environ Health* 1982;8(suppl 1):7-14.
36. White JE. A two stage design for the study of the relationship between a rare exposure and a rare disease. *Am J Epidemiol* 1982;115:119-28.
37. Breslow NE, Cain KC. Logistic regression for two-stage case-control data. *Biometrika* 1988;75: 11-20.
38. Cain KC, Breslow NE. Logistic regression analysis and efficient design for two-stage studies. *Am J Epidemiol* 1988;128:1198-1206.
39. Graubard B, Fears TR, Gail MH. Effects of cluster sampling on epidemiologic analysis in population-based case-control studies. *Biometrics* 1989;45: 1053-71.
40. Werler MM, Pober BR, Nelson K, et al. Reporting accuracy among mothers of malformed and nonmalformed infants. *Am J Epidemiol* 1989;129: 415-21.
41. Swan SH, Shaw GM. Re: "Reporting accuracy among mothers of malformed and nonmalformed infants." (Letter). *Am J Epidemiol* 1990;131: 935-6.
42. Berg AT. Re: "Reporting accuracy among mothers of malformed and nonmalformed infants." (Letter). *Am J Epidemiol* 1990;131:936-7.
43. Harlow SD, Linet MS. Agreement between questionnaire data and medical records: the evidence for accuracy of recall. *Am J Epidemiol* 1989;129: 233-48.
44. Preston-Martin S, Bernstein L, Maldonado AA, et al. A dental x-ray validation study: comparison of information from patient interviews and dental charts. *Am J Epidemiol* 1985;121:430-9.
45. Mackenzie SG, Lippman A. An investigation of report bias in a case-control study of pregnancy outcome. *Am J Epidemiol* 1989;129:65-75.
46. Linet MS, Van Natta ML, Brookmeyer R, et al. Familial cancer history and chronic lymphocytic leukemia: a case-control study. *Am J Epidemiol* 1989;130:655-64.
47. Drews CD, Kraus JF, Greenland S. Recall bias in a case-control study of sudden infant death syndrome. *Int J Epidemiol* 1990;19:405-11.
48. Friedenreich CM, Howe GR, Miller AB. An investigation of recall bias in the reporting of past food intake among breast cancer cases and controls. *Ann Epidemiol* 1991;1:439-53.
49. Greenland S, Thomas DC. On the need for the rare disease assumption in case-control studies. *Am J Epidemiol* 1982;116:547-53.
50. Gail M, Lubin JH, Silverman DT. Elements of design in epidemiologic studies. In: DeLisi C, Eisenfeld J, eds. *Statistical methods in cancer epidemiology*. Amsterdam: North Holland Publishing Co, 1985:313-23.
51. Ibrahim MA, Spitzer WO. The case-control study: the problem and the prospect. *J Chronic Dis* 1979; 32:139-44.
52. Rothman KJ. *Modern epidemiology*. Boston: Little, Brown & Company, 1986.
53. Greenberg BG. The future of epidemiology. *J Chronic Dis* 1983;36:353-9.

Lecture: Evaluating and Reporting Case Control Studies



## Evaluating and reporting case-control studies

PHW250 F - Jack Colford

JACK COLFORD: Let's talk now how to evaluate and report case control studies.

## Assessment of case-control studies

### How much selection bias was present?

- Was the study base principle met with either a primary or secondary study base?
- Is the distribution of the exposure among controls representative of the distribution of exposure in the study base?
- Was the outcome clear, specific, and measurable?

### Were potential confounding factors sought and controlled for in the analysis?

- Was the deconfounding principle met?
- Did the investigators anticipate and gather information on potential confounding factors?
- What method(s) were used to assess and control for confounding?

### What steps were taken to minimise information bias?

- Was the comparable accuracy principle met?
- Was the exposure clear, specific, and measurable?
- Was the outcome identified in the same way for cases and controls?
- Was the accuracy of outcome and exposure ascertainment similar for cases and controls?
- Was determination of exposure made by an observer blinded as to case/control status?

When we talk about how to assess case control studies, we think of three broad categories of questions. First, how much selection bias was present? Secondly, were there potential confounding factors that were sought and controlled for by the investigators in the analysis? And finally, what steps did the investigators take to minimize information bias?

With respect to selection bias, the first idea to consider is what we discussed when we talked about whether the case control study was drawn from a primary or secondary study base-- that is, the study base principle. Which study base principle was used? Then we want to ask, was the distribution of the exposure among the controls who were selected representative of the distribution of exposure in the study base?

Are our controls a good representation of the broader population in the study base from which these controls are drawn? And finally, was the outcome clear, specific, and measurable? In a very detailed kind of way, we want to know what the authors were specifically measuring, was the outcome clear, and how was it measured.

Next, our big category of questions for evaluation of a case control study is, were potential confounding factors sought and controlled for in the analysis? Because this is an observational study, it's particularly worrisome that there might be confounding factors that distort our relationships. First, we want to ask, was the deconfounding principle met?

What did the authors do to remove confounding variables? Did they anticipate and gather information about potential confounding factors and then how did they assess and control for confounding? Did they do design steps such as matching or did they do analysis steps such as stratification or multivariate analysis?

And finally, what steps did the authors or investigators take to minimize information bias. This is back to that principle in study design for case control studies of, was there comparable accuracy between the measurement of exposures and outcomes in the cases and the controls? Were the exposures clear, specific, and measurable?

Was the outcome identified in the same way for both cases and controls? If coronary heart disease is the outcome, were the same sorts of tests used to measure it that were applied to cases as were applied to controls? Was the accuracy of the outcome and exposure ascertainment similar for cases and controls? And finally, was the determination of exposure made by an observer blinded as to case control status? We often think of case control studies as not being able to use blinding, but in fact, we can have case control studies in which the people who are measuring the exposure in the cases and the controls were blinded as to whether the person they're measuring at a given point is a case or a control.

## STROBE Checklist for reporting case-control studies



- The checklist arose out of concerns that observational studies were poorly and inconsistently reported.
- Poor reporting makes it difficult to assess strengths and weaknesses of a study.
- A group of methodologists, researchers, and editors developed the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) recommendations to improve the quality of reporting of observational studies.
- 22 item checklist
- 18 items are common to cohort studies, case-control studies and cross-sectional studies.
- 4 items are specific to each of the three study designs.

Berkeley School of Public Health

We've used the STROBE checklist when we discussed cohort studies and here's the STROBE checklist again for observational studies in general, and it has sections for case control studies in particular. And again, this checklist arose out of concern that observational studies were being poorly and inconsistently reported. And when reporting is poor, it's difficult for readers to assess the strength and weaknesses of the study.

The STROBE checklist, which stands for the Strengthening and Reporting of Observational Studies and Epidemiology, is a set of recommendations to improve the quality of reporting. It has 22 items. 18 of the items are common to cohort studies, case control studies, and cross-sectional studies. We've used it before. And then there are four extra items to use for cohort studies, four extra items to use for case control studies, and four extra items to use for cross-sectional studies, depending on what sort of study you're investigating or reviewing.

**STROBE Statement—Checklist of items that should be included in reports of *case-control studies***

Item No	Recommendation
<b>Title and abstract</b>	1 (a) Indicate the study's design with a commonly used term in the title or the abstract (b) Provide in the abstract an informative and balanced summary of what was done and what was found
<b>Introduction</b>	
Background/rationale	2 Explain the scientific background and rationale for the investigation being reported
Objectives	3 State specific objectives, including any prespecified hypotheses
<b>Methods</b>	
Study design	4 Present key elements of study design early in the paper
Setting	5 Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6 (a) Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls (b) For matched studies, give matching criteria and the number of controls per case
Variables	7 Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable

The STROBE statement is a checklist. It has broad categories that you see on the next few slides. I'll just highlight a couple of them.

Title and abstract are commonly used terms about what the investigators are doing. Under methods, what are some of the key elements that were present in the study and are they presented early and clearly in the paper? Participants is a really important section. We want to know exactly what determine eligibility of cases and controls in a case control study and what was the rationale for that.

And if matching was used, we want to be clearly able to see what matching criteria were chosen and what were the number of controls per case.

Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	<ul style="list-style-type: none"> <li>(a) Describe all statistical methods, including those used to control for confounding</li> <li>(b) Describe any methods used to examine subgroups and interactions</li> <li>(c) Explain how missing data were addressed</li> <li>(d) If applicable, explain how matching of cases and controls was addressed</li> <li>(e) Describe any sensitivity analyses</li> </ul>



Bias is something we've talked a lot about in the course and earlier in the lecture today. We want to very clearly see and understand what the authors did to address potential sources of selection and information bias. How did the authors determine the study size that was used in the case control study that's being reported?

And then there are a number of details about statistical methods. I'll call your attention to item 12c, which we haven't discussed much about before, which is, what did the authors do when data were missing? Did they drop those participants? Did they do something special to impute the missing values based on averages? Well, there's lots of different techniques for missing data, but what did the authors do?

<b>Results</b>		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed  (b) Give reasons for non-participation at each stage  (c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders  (b) Indicate number of participants with missing data for each variable of interest
Outcome data	15*	Report numbers in each exposure category, or summary measures of exposure
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included  (b) Report category boundaries when continuous variables were categorized  (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period
Other analyses	17	Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses

5

With respect to the results of this study, you see the various categories here. I just want to draw your attention to the Main Results section where we talk about reporting unadjusted results first and then, if appropriate, confounder adjusted results, then recall that the confounder adjusted results arise from either multi-variate analysis or stratification and approaches such as the Mantel-Haenszel tests to come up with a summary adjusted estimate. We want to be clear we understand what confounders the authors adjusted for and why they were included.

And then were there other analyses done? Were there sub-groups studied? Were interactions studied? Were sensitivity analyses done to see how important certain measurements or exposures were in the final results?

<b>Discussion</b>		
Key results	18	Summarise key results with reference to study objectives
Limitations	19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias
Interpretation	20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence
Generalisability	21	Discuss the generalisability (external validity) of the study results
<b>Other information</b>		
Funding	22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based



And then in the discussion section, really well-written papers will show us very clearly what the authors see as the limitations of the study. We shouldn't have to come up with these ourselves, but we always need to try to see if the authors have included all the potential limitations that might have occurred. And the last comment would be about the funding. Is it very clear who funded the study and is there any potential for bias? You know, did the tobacco industry fund the study of smoking and lung cancer and so forth?

## Summary of key points

- You can use the principles outlined in the Wacholder et al. 1992 series when evaluating the quality of a case-control study.
- We recommend that you use the STROBE reporting checklist when publishing results of a case-control study.
- The article by Vandenbrouke et al. 2007 article provides detailed examples of how to use the STROBE checklist with an example paper.



In summary, we can use the principles outlined in the Wacholder 1992 series of papers that we've been working with in the course when evaluating the quality of the case control study. We personally recommend the use of the STROBE reporting checklist when publishing or reviewing the results of a case control study. And the article we provided you by Vandenbrook and colleagues from 2007 give detailed examples of how to use the STROBE checklist with an example paper. You might want to practice using that.

STROBE Statement—Checklist of items that should be included in reports of *case-control studies*

	<b>Item No</b>	<b>Recommendation</b>
<b>Title and abstract</b>	1	(a) Indicate the study's design with a commonly used term in the title or the abstract  (b) Provide in the abstract an informative and balanced summary of what was done and what was found
<b>Introduction</b>		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
<b>Methods</b>		
Study design	4	Present key elements of study design early in the paper
Setting	5	Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection
Participants	6	(a) Give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls  (b) For matched studies, give matching criteria and the number of controls per case
Variables	7	Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable
Data sources/ measurement	8*	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group
Bias	9	Describe any efforts to address potential sources of bias
Study size	10	Explain how the study size was arrived at
Quantitative variables	11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen and why
Statistical methods	12	(a) Describe all statistical methods, including those used to control for confounding  (b) Describe any methods used to examine subgroups and interactions  (c) Explain how missing data were addressed  (d) If applicable, explain how matching of cases and controls was addressed  (e) Describe any sensitivity analyses
<b>Results</b>		
Participants	13*	(a) Report numbers of individuals at each stage of study—eg numbers potentially eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed  (b) Give reasons for non-participation at each stage  (c) Consider use of a flow diagram
Descriptive data	14*	(a) Give characteristics of study participants (eg demographic, clinical, social) and information on exposures and potential confounders  (b) Indicate number of participants with missing data for each variable of interest
Outcome data	15*	Report numbers in each exposure category, or summary measures of exposure
Main results	16	(a) Give unadjusted estimates and, if applicable, confounder-adjusted estimates and their precision (eg, 95% confidence interval). Make clear which confounders were adjusted for and why they were included  (b) Report category boundaries when continuous variables were categorized  (c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period

---

Other analyses 17 Report other analyses done—eg analyses of subgroups and interactions, and sensitivity analyses

---

### **Discussion**

---

Key results 18 Summarise key results with reference to study objectives

---

Limitations 19 Discuss limitations of the study, taking into account sources of potential bias or imprecision.  
Discuss both direction and magnitude of any potential bias

---

Interpretation 20 Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence

---

Generalisability 21 Discuss the generalisability (external validity) of the study results

---

### **Other information**

---

Funding 22 Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based

---

\*Give information separately for cases and controls.

**Note:** An Explanation and Elaboration article discusses each checklist item and gives methodological background and published examples of transparent reporting. The STROBE checklist is best used in conjunction with this article (freely available on the Web sites of PLoS Medicine at <http://www.plosmedicine.org/>, Annals of Internal Medicine at <http://www.annals.org/>, and Epidemiology at <http://www.epidem.com/>). Information on the STROBE Initiative is available at <http://www.strobe-statement.org>.



## Epidemiology Case Studies – Episode 3: Risk Factors for Menstrual Toxic Shock Syndrome - Multistate Case-Control Study

**MICHELLE RUIZ:** Hello welcome to the third episode of Epidemiology Case Studies. In this episode Dr. Jack Colford interviews Dr. Art Reingold about the case-control study performed to asses risk factors for developing toxic shock syndrome during menstruations.

**JACK COLFORD:** Hi. This is Jack Colford, a professor of epidemiology here at UC Berkeley, continuing our series of interviews on research studies. And today, it's my great pleasure to interview professor Art Reingold, the chairman of our division of epidemiology and a longtime member of our faculty.

We're going to talk about his case control study related to toxic shock syndrome that you've read. So, Art, perhaps first I'll start with asking you to just tell us what toxic shock syndrome is.

**ART REINGOLD:** Sure, well, as any woman who uses tampons should know, if she's reading her package insert, toxic shock syndrome is a potentially lethal, serious infection, in this case caused by the bacterium *Staph aureus*, producing a toxin-- a poisonous substance, which then gets into the system and lowers your blood pressure, gives you a fever, produces a rash, causes organ damage. And if it's not well-treated appropriately, it can be fatal or produce very serious injury to the body.

**JACK COLFORD:** Thanks. Tell us some of the hypotheses you had going into the study that you were trying to test.

**ART REINGOLD:** So this particular study was done once we already knew a fair bit about toxic shock syndrome occurring during menstruation. Early on, there were lots of hypotheses about whether tampon use was a risk factor, but also whether things such as over-the-counter medications to treat menstrual cramps, whether oral contraception or exercise or sexual activity during menstruation might be risk factors.

By the time this study was done, while many of those questions had been laid to rest, there were still a couple of issues beyond the role of tampons. But the primary question was whether the tampons then on the market, which had lower absorbency and different composition than a few years earlier, were still associated with an increased risk of developing this disease.

**JACK COLFORD:** Great. So you chose to do a case-control study. Why, from all the different kind of study designs that our students have studied, did you pick a case-control study?

**ART REINGOLD:** It was really the only feasible study design. Toxic track syndrome at its peak, at its worst, is a very rare disease. So under the worst of circumstances, the rate was perhaps 10 to 15 per 100,000 per year. So cohort studies are simply not possible with such a rare condition.

Randomized trials obviously off the table. So it really was the only plausible study design.

**JACK COLFORD:** Some of our students realize that the selection of controls is really the tough thing to do in a case-control study, or one of the tough things to do. Can you tell us a little bit about how you picked your controls, how you selected them?

**ART REINGOLD:** So because the cases were being identified and enrolled across a number of different states, doing real population-based sampling-- a random sample of women the population was really not an option. There are no lists of such women to sample from.

So but we were also very concerned about whether the right control group would be, for example, friend controls, or whether that might be too similar to the cases with regard to the exposures we are studying-- or something close to a population control. So we did both. We enrolled friend controls. And we enrolled controls from the community through random digit dialing.

**JACK COLFORD:** And in the paper, you mentioned that you also matched on age and area residents. Can you tell us a little bit about your thinking there?

**ART REINGOLD:** Well, I think the primary concern of matching on an age was that tampon use varies considerably by age of the woman. And so we wanted to make sure that we were not comparing 45-year-old women to 20-year-old women. Of course, we could have dealt with that in the analysis. But the decision was done to match on that. I think matching on geographic area was, frankly, a little more based on simple convenience and ease of sampling strategy than it was on concerns about confounding.

**JACK COLFORD:** And after all was said and done and used these procedures, how comparable were the cases and controls for these other things?

**ART REINGOLD:** The case and controls or the two groups of controls? The two different types of controls?

**JACK COLFORD:** No, the cases and controls.

**ART REINGOLD:** So the cases and controls were very similar with regard to age. And obviously, they were identical with regard to which state they lived in or what geographic area they lived in. So they were very well-matched with regard to the matching criteria. And also, it turned out to be relatively similar around a number of other variables.

**JACK COLFORD:** And tell us more about how you specifically designed the exposure that was being studied.

**ART REINGOLD:** Well, so as I said, the primary question being answered in this study was how much risk was associated, depending on which type of tampon a woman used-- chemical composition and it's absorbency-- how much liquid it could absorb-- as well as the features of whether they were used continuously, intermittently, how long the tampons were left in. So those were the key issues.

And so the questionnaire was designed to focus primarily on questions of a tampon use during two menstrual periods. In the case of the women, the menstrual period when had onset of her illness and the one before that. And with regard to the controls, two menstrual periods during the same time period.

**JACK COLFORD:** Great, OK. So when the researchers assess the exposure of the women to the tampons, were they blinded?

**ART REINGOLD:** So the researchers, the interviewers were blinded to the case-control status of the women, but not to the hypothesis. They certainly knew the hypothesis. We also, as is shown in the article, tried to assess whether the interviewers remained blinded at the end of the interview. Because sometimes during an interview, someone may say something that clearly gives away--

**JACK COLFORD:** Reveals it, yeah.

**ART REINGOLD:** --what they were. And if I recall correctly, the blinding was maintained through the interview for about 80% of the individuals interviewed. But in some instances, blinding was undone during the course of the conversation.

**JACK COLFORD:** And tell us about the team that conducted the study. How big was the team? How spread apart were they? How did you interact with each other as a team-- the research team?

**ART REINGOLD:** Well, I was working at CDC at the time. So I had staff--

**JACK COLFORD:** In Atlanta.

**ART REINGOLD:** --there in Atlanta, working for me. But we also were enrolling in these cases and controls across a number of different states scattered across the country. So the identification of the cases and the enrolling was occurring at states around the country. But all the interviews were being done by telephone from Atlanta.

**JACK COLFORD:** How big was your team?

**ART REINGOLD:** Boy, I don't remember exactly. We had one full-time staff person, Suzanne Gaventa. And I don't remember how many people she had helping her. But there were probably at least one or two additional people helping with the interviews.

**JACK COLFORD:** Great, great. And were all of these epidemiologists, or were there other people on the research team?

**ART REINGOLD:** They were pretty much all epidemiologists-- some master's level and doctoral level.

**JACK COLFORD:** Mm-hmm, great. OK, so that concludes our discussion of the kind of basic design of the study, Art. Thanks very much.

**ART REINGOLD:** OK.

**JACK COLFORD:** So I know the camera's still rolling, Michelle. We're going to turn now to the next course. You have the next course for later in the semester.

So welcome again. We're talking with Professor Art Reingold about his study of the association between tampon use and toxic shock syndrome in the *JAMA* article that you've all read. And I'd like to begin, Art, today talking about some of the potential biases in a study like this and how the case-control study you used might address things like, for instance, selection bias, information bias.

**ART REINGOLD:** So I would say of the three main areas where there'd be concern about inference-- confounding information bias and selection bias-- we were fairly comfortable that uncontrolled confounding was not a major problem. We certainly collect either matched on, as we already indicated, age and geographic area of residence. We collected information about other potential confounders. There weren't big differences, but they were controlled for in the analysis.

So I think confounding was not so much an issue, but there are certainly concerns in this study or in similar prior studies about information bias and selection bias. And so one question at the heart of any of these studies of menstrual toxic shock is whether the cases are representative of all the cases occurring during menstruation, or whether there might be a greater diagnosis in reporting of cases in tampon users than in non-tampon users.

And once the association had been reported and was known to women, was known to health care providers, it's certainly possible that cases that were diagnosed and reported and included in the study were more likely to be tampon users than non-tampon users. To the extent that that was true in our study, that would have biased the results toward an association with tampon use in general. And there's not much we can do to evaluate that, although we did, in fact, using a lot of resources, look through medical records of people with overlapping clinical diagnoses to see if any of them were undiagnosed cases of toxic shock syndrome.

**JACK COLFORD:** OK.

**ART REINGOLD:** A huge amount of work that didn't produce very much. But I would point out that within this study, the primary question wasn't whether tampon use increased the risk, but whether use of one particular brand--

**JACK COLFORD:** Or type.

**ART REINGOLD:** --or style or absorbency increased the risk, compared to another brand or style or absorbency. So for selection bias to have influenced that, we would have needed patients or their providers to be more suspicious of and report a case in users of one brand--

**JACK COLFORD:** Particular type, yeah.

**ART REINGOLD:** --versus another.

**JACK COLFORD:** Sure, sure.

**ART REINGOLD:** Higher absorbency versus lower absorbency. And I, personally, don't think that was very likely. But I suppose one could argue about that. So certainly, questions about whether there could have been some selection bias. So the odds ratio for tampon use versus no tampon use could certainly, possibly have been influenced by that.

In the era of information bias, in all of these case-control studies of menstrual toxic shock syndrome, there have been concerns about whether women accurately report on their tampon use or their other exposures. And we actually had the women look for their package of tampons that had been used. About 60% of them had the package.

**JACK COLFORD:** Yup.

100

**ART REINGOLD:** They could read the label. They could describe the package and tell us the color.

**JACK COLFORD:** Sure.

**ART REINGOLD:** We were pretty sure that they were giving us accurate information. I would also assert that the average woman can reliably report whether she used tampons or not during her last month period. You know, it's not a subtle thing to report on.

So when you start asking questions about sexual activity and other things where there might be some sensitivity, I suppose you could question whether women would report honestly about something like that. But we have reason to think that women were being pretty candid about the things we were asking them about.

**JACK COLFORD:** Great. So you estimated odds ratios in this study. Can you tell us kind of the statistical methods you used to estimate those?

**ART REINGOLD:** Well, it was pretty straightforward, case-control studies 101 in terms of looking at basically what proportion of the cases were exposed, what portion of the controls were exposed, and initially using univariate analyses and then, ultimately, logistic regression, in this case because they were matched to conditional logistic regression to control for possible confounders and generate an odds ratio that took into account the matching.

**JACK COLFORD:** The matching, great.

**ART REINGOLD:** So there was nothing very fancy about the analysis.

**JACK COLFORD:** Conditional on the two factors you matched on.

**ART REINGOLD:** Exactly.

**JACK COLFORD:** So perfect. OK. So what associations did you find between the various tampon use and types of tampons and toxic shock syndrome?

**ART REINGOLD:** So beyond the association of tampon use--

**JACK COLFORD:** In general.

**ART REINGOLD:** --versus no tampon use, which, as I've said, there certainly could be some selection bias in that, fundamentally we confirmed our hypothesis that some brands and styles were higher risk compared to other brands and styles. And so they are one particular brand, Tampax Regular,<sup>101</sup> which was thought to be, if anything, the lowest risk product. When that was the standard of

comparison, higher absorbency tampons were associated with a higher odds of disease compared to Tampax Regular use.

And we also actually produced an odds ratio for increasing absorbency because tampons can be measured in vitro or in vivo for how many grams of liquid they absorb and produce an odds ratio per gram of absorbency. And showed that for every increase in gram of absorbency of a tampon product, there was a 34% increase in the odds of toxic shock syndrome.

**JACK COLFORD:** My recollection of that figure in the article was it was very linear, that relation.

**ART REINGOLD:** Yeah, that was pretty much a straight, linear relationship. So again, I have a hard time believing that was due to bias in terms of what people were reporting, and certainly suggest that absorbency was the critical factor in determining a risk of toxic shock syndrome.

Now, I would say that two things remain unresolved by that-- one is the role of chemical composition. Because not all tampons are made of the same chemicals-- cotton versus rayon being a major distinction. We didn't really end up having a large enough number of cases to look at that factor. The good news is, by the time we were doing that study, the rate of this disease declined and we didn't have enough cases.

But the other real question is, what is it in a tampon, or what is it about absorbency that changes your risk of this disease? And there are lots of theories. So one is, how much oxygen is introduced when you introduce a tampon. Because oxygen is being introduced into a normally oxygen-less or anaerobic environment.

So maybe the amount of oxygen introduced is important. Maybe these products bind certain cations, like calcium and magnesium that, in vitro, is important in terms of toxin production. So our studies really couldn't begin to tease apart those questions about what is it about absorbency and is there an added effect of chemical composition.

**JACK COLFORD:** How generalizable do you think your study in a particular population that you used is to other populations, either in the US or in other countries or?

**ART REINGOLD:** Well, I don't have any reason to think it isn't generalizable. Now, the risk of this disease has always been substantially higher in younger women than in older women. And that's consistently been found to be the case.

is not as much of a risk factor when you're an older woman than a younger woman? That's possible. One reason might be that older women are more likely to have antibodies to the toxin and basically be immune to the disease--

**JACK COLFORD:** Sure.

**ART REINGOLD:** --as opposed to younger women, who are not. So it's possible that the results don't extrapolate well to all age groups.

**JACK COLFORD:** Mm-hmm.

**ART REINGOLD:** In terms of racial groups or geographic groups, I honestly don't see any reason to think they wouldn't be generalizable.

**JACK COLFORD:** Be any different. Sure, sure. How would you say your study-- this particular case-control study-- improved upon or built upon prior studies of the toxic shock syndrome in risk factors?

**ART REINGOLD:** Well, I think, first of all, it quantified the relationship with absorbency in a way that earlier studies had not.

**JACK COLFORD:** Hadn't done. Yup.

**ART REINGOLD:** Secondly, as I said, because in response to the earlier studies, manufacturers had substantially reduced the absorbency of products they were selling and changed the chemical composition. The real question was whether the risk continued, given that the products had changed. So I think it contributed to our understanding that, yes, the risk was still there, in addition to giving us a pretty good quantitative estimation of what that risk was.

I think it also provided some reassurance that some of the other variables we looked at, like contraception and the like, were not important. Or over-the-counter medication use didn't contribute to the risk of this disease.

**JACK COLFORD:** Do you have any specific stories or tales of manufacturers' responses to your study or?

**ART REINGOLD:** Well, to our particular study, I think the industry was basically fairly quiescent. They didn't raise a fuss. They didn't start pointing to a lot of flaws. I don't think they made a lot in the way of further changes to their products. I think they pretty much accepted it at face value and went on with their jobs.

**JACK COLFORD:** What happened to the national incidence of the syndrome as time went on?

**ART REINGOLD:** Well, there's a lot of controversy about that, even to this day. And several studies, including one that I did with colleagues at Kaiser Permanente, would suggest that the rate of menstrual toxic shock syndrome in some areas did go down during this time period. In other areas, it didn't go down very much. And so while the number of reported cases seem to have gone down after the early 1980s, these other studies that pretty much remove either diagnostic bias or reporting bias completely sort of give a mixed picture about that.

So that's somewhat unclear, how much the rate declined at that point. I would say at this point in time, in 2018, it remains an exceedingly rare disease. So the estimates are at this point well under 1 in 100,000 menstruating women per year, and perhaps more like 1 in 500,000.

**JACK COLFORD:** OK.

**ART REINGOLD:** And so is it a common problem? No. And at this point, it's very hard to study because it's so rare.

**JACK COLFORD:** So rare. Yeah. Yeah.

**ART REINGOLD:** So there are not ongoing studies that I know of.

**JACK COLFORD:** So if you could go back and do anything different about the design and conduct of this particular study, what would you have done?

**ART REINGOLD:** Well, that's a good question. You know, I think we certainly did our best around control selection and trying to make sure the cases were representative of all the cases that were out there. I can't think of any hypotheses that I would have tried to study that would seem plausible that we would have looked at.

So I guess, maybe if we'd carried it on longer and enrolled enough more cases, we'd have had a better understanding about the role of chemical composition. But the decision was, we would only enroll patients for 18 months. And when that time was over, we stopped.

**JACK COLFORD:** You were done. Great.

**ART REINGOLD:** [CHUCKLES]

**JACK COLFORD:** Well, thanks a lot, Art, for talking to us about your case-control study of tampons and toxic

shock syndrome.

**ART REINGOLD:** Great, happy to do it.

**JACK COLFORD:** Thank you.

# Effects of water quality, sanitation, handwashing, and nutritional interventions on diarrhoea and child growth in rural Kenya: a cluster-randomised controlled trial

Clair Null, Christine P Stewart, Amy J Pickering, Holly N Dentz, Benjamin F Arnold, Charles D Arnold, Jade Benjamin-Chung, Thomas Clasen, Kathryn G Dewey, Lia C H Fernald, Alan E Hubbard, Patricia Kariger, Audrie Lin, Stephen P Luby, Andrew Mertens, Sammy M Njenga, Geoffrey Nyambane, Pavani K Ram, John M Colford Jr



## Summary

**Background** Poor nutrition and exposure to faecal contamination are associated with diarrhoea and growth faltering, both of which have long-term consequences for child health. We aimed to assess whether water, sanitation, handwashing, and nutrition interventions reduced diarrhoea or growth faltering.

**Methods** The WASH Benefits cluster-randomised trial enrolled pregnant women from villages in rural Kenya and evaluated outcomes at 1 year and 2 years of follow-up. Geographically-adjacent clusters were block-randomised to active control (household visits to measure mid-upper-arm circumference), passive control (data collection only), or compound-level interventions including household visits to promote target behaviours: drinking chlorinated water (water); safe sanitation consisting of disposing faeces in an improved latrine (sanitation); handwashing with soap (handwashing); combined water, sanitation, and handwashing; counselling on appropriate maternal, infant, and young child feeding plus small-quantity lipid-based nutrient supplements from 6–24 months (nutrition); and combined water, sanitation, handwashing, and nutrition. Primary outcomes were caregiver-reported diarrhoea in the past 7 days and length-for-age Z score at year 2 in index children born to the enrolled pregnant women. Masking was not possible for data collection, but analyses were masked. Analysis was by intention to treat. This trial is registered with ClinicalTrials.gov, number NCT01704105.

**Findings** Between Nov 27, 2012, and May 21, 2014, 8246 women in 702 clusters were enrolled and randomly assigned an intervention or control group. 1919 women were assigned to the active control group; 938 to passive control; 904 to water; 892 to sanitation; 917 to handwashing; 912 to combined water, sanitation, and handwashing; 843 to nutrition; and 921 to combined water, sanitation, handwashing, and nutrition. Data on diarrhoea at year 1 or year 2 were available for 6494 children and data on length-for-age Z score in year 2 were available for 6583 children (86% of living children were measured at year 2). Adherence indicators for sanitation, handwashing, and nutrition were more than 70% at year 1, handwashing fell to less than 25% at year 2, and for water was less than 45% at year 1 and less than 25% at year 2; combined groups were comparable to single groups. None of the interventions reduced diarrhoea prevalence compared with the active control. Compared with active control (length-for-age Z score -1.54) children in nutrition and combined water, sanitation, handwashing, and nutrition were taller by year 2 (mean difference 0.13 [95% CI 0.01–0.25] in the nutrition group; 0.16 [0.05–0.27] in the combined water, sanitation, handwashing, and nutrition group). The individual water, sanitation, and handwashing groups, and combined water, sanitation, and handwashing group had no effect on linear growth.

**Interpretation** Behaviour change messaging combined with technologically simple interventions such as water treatment, household sanitation upgrades from unimproved to improved latrines, and handwashing stations did not reduce childhood diarrhoea or improve growth, even when adherence was at least as high as has been achieved by other programmes. Counselling and supplementation in the nutrition group and combined water, sanitation, handwashing, and nutrition interventions led to small growth benefits, but there was no advantage to integrating water, sanitation, and handwashing with nutrition. The interventions might have been more efficacious with higher adherence or in an environment with lower baseline sanitation coverage, especially in this context of high diarrhoea prevalence.

**Funding** Bill & Melinda Gates Foundation, United States Agency for International Development.

**Copyright** © The Author(s). Published by Elsevier Ltd. This is an Open Access article under the CC BY 4.0 license.

## Introduction

An estimated 156 million children worldwide suffer from stunting (linear growth faltering) and are unlikely to reach their full potential as adults.<sup>1</sup> Linear growth faltering

is the most apparent sign of chronic undernutrition and is the physical manifestation of combined physiological and developmental insults. Early-life stunting leads to poor cognitive development in childhood, reduced

*Lancet Glob Health* 2018;  
6: e316–29

Published Online  
January 29, 2018  
[http://dx.doi.org/10.1016/S2214-109X\(18\)30005-6](http://dx.doi.org/10.1016/S2214-109X(18)30005-6)  
See [Comment](#) page e236  
See [Articles](#) page e302  
Innovations for Poverty Action, Kakamega, Kenya (C Null PhD, H N Dentz MPH, G Nyambane MA); Center for International Policy Research and Evaluation, Mathematica Policy Research, Washington, DC, USA (C Null); Rollins School of Public Health, Emory University, Atlanta, GA, USA (C Null, Prof T Clasen PhD); Department of Nutrition, University of California, Davis, CA, USA (C P Stewart PhD, H N Dentz, C D Arnold MS, Prof K Dewey PhD); Department of Civil and Environmental Engineering (A J Pickering PhD), and Department of Infectious Diseases and Geographic Medicine (S P Luby MD), Stanford University, Stanford, CA, USA; Department of Civil and Environmental Engineering, Tufts University, Medford, MA, USA (A J Pickering); Division of Epidemiology (B F Arnold PhD, J Benjamin-Chung PhD, A Lin PhD, A Mertens MS, Prof J M Colford Jr MD), Division of Community Health Sciences (Prof L C H Fernald PhD, P Kariger PhD), and Division of Biostatistics (A E Hubbard PhD) School of Public Health, University of California, Berkeley, CA, USA; Eastern and Southern Africa Centre of International Parasite Control, Kenya Medical Research Institute, Nairobi, Kenya (S M Njenga PhD); Department of Epidemiology and Environmental Health, School of Public Health and Health

Professions, University at  
Buffalo, Buffalo, NY, USA  
(P K Ram MD)

Correspondence to:  
Dr Clair Null, Center for  
International Policy Research and  
Evaluation, Mathematica Policy  
Research, Washington,  
DC 20002, USA  
cnull@mathematica-mpr.com

### Research in context

#### Evidence before this study

Malnutrition and enteric infection are thought to act together to impair child health and survival, yet there is limited evidence of low cost, scalable interventions effective at breaking this cycle. A 2008 meta-analysis by Dewey and Adu-Afarwuah found that interventions offering nutrient supplementation or counselling on complementary feeding could result in modest improvements to child growth. Another meta-analysis by Waddington and Snilsveit in 2009 showed that water treatment or handwashing could prevent diarrhoea, but there had not been any randomised trials of the effect of sanitation on diarrhoea. During this study, five other randomised trials of the effects of sanitation on diarrhoea and growth were published, but three were limited by low adherence. Whether combining water, sanitation, handwashing, or nutrition interventions could result in added benefits for health and growth was not known.

#### Added value of this study

This trial is one of the first to provide experimental evidence on whether individual and combined water, sanitation, or handwashing interventions improve growth; combined water, sanitation, and handwashing interventions are more effective at reducing diarrhoea and growth faltering than any intervention alone; and nutrition counselling and supplementation are more effective when combined with

improved water, sanitation, and handwashing. This is the first rigorous evaluation of upgrading from unimproved to improved latrines in sub-Saharan Africa. None of the interventions reduced diarrhoea, and only the interventions that included nutrition counselling and nutrient supplementation improved growth.

#### Implications of all the available evidence

Our results on growth effects are consistent with those from previous research on the combination of nutrition counselling and nutrient supplementation, finding modest effects on linear growth. It is possible that more intensive promotion and higher adherence would have resulted in larger effects, especially in this context of high diarrhoea prevalence, but few programmes are likely to be able to afford sustaining a more ambitious behaviour change programme than was included in this trial. In a context where most households already had an unimproved sanitation facility, provision of technologically simple interventions including chlorination for household treatment of drinking water, improved pit latrines, and handwashing stations—standard for most WASH programmes in rural areas of low-income countries—might not be sufficient to improve growth. By contrast with previous studies, this trial provided evidence that technologically simple water, sanitation, and handwashing interventions with adherence rates at least as high as most programmes achieve might not reduce childhood diarrhoea in all situations.

economic productivity in adulthood, and increased risk of morbidity and mortality.<sup>2,3</sup> Because nutrient supplementation and counselling interventions for maternal, infant, and young child feeding have been only marginally successful at preventing growth faltering, exposure to faecal contamination in the environment has recently been hypothesised to lead to environmental enteric dysfunction, which features chronic immune stimulation and impaired nutrient absorption, thereby constraining a growth response to improved nutrition.<sup>4</sup> In addition to the detrimental effects on growth and development, undernutrition was estimated to cause 45% of all child deaths in 2011, and it has long been recognised that undernutrition is an important determinant of susceptibility to infectious disease.<sup>5,6</sup> Diarrhoea is the second leading cause of death in children aged 1–59 months, contributing to almost 500 000 deaths in children younger than 5 years in 2015.<sup>7</sup> Frequent diarrhoea is also associated with linear growth faltering.<sup>8</sup> If there is a pathway independent of symptomatic diarrhoea linking environmental contamination to growth faltering, the benefits of improving water safety, sanitation, and handwashing could be underestimated because studies have generally focused on diarrhoea. It is unclear whether combined water, sanitation, handwashing, and nutritional interventions reduce diarrhoea or improve growth more than single interventions.

See Online for appendix

We aimed to investigate whether individual water, sanitation, handwashing, or nutrition interventions can reduce linear growth faltering; to assess whether combined water, sanitation, and handwashing interventions are more effective at reducing diarrhoea than individual interventions; and to investigate whether the combination of water, sanitation, handwashing, and nutrition interventions reduces growth faltering more than each individual intervention. A companion trial<sup>9</sup> in Bangladesh evaluated the same objectives.

### Methods

#### Study design

The Kenya WASH Benefits study was a cluster-randomised trial done in rural villages in Bungoma, Kakamega, and Vihiga counties in Kenya's western region (appendix p 11). We used a cluster design to facilitate the logistics of the behaviour change component of the interventions and minimise contamination between intervention and comparison households. We hypothesised that the interventions would improve the health of the index child in each household. We optimised the trial design to measure group-level differences in primary outcomes by including a large number of clusters, each comprising relatively few children (12 on average) with infrequent measurement. Each measurement round lasted roughly 1 year and was balanced across treatment

groups and geography to minimise seasonal or geographic confounding when comparing outcomes across groups.

With active and passive control groups and six intervention groups (water; sanitation; handwashing; combined water, sanitation, and handwashing; nutrition; and combined water, sanitation, handwashing, and nutrition), the design enabled 11 comparisons of each intervention group with the active control; combined water, sanitation, and handwashing with each intervention alone; and combined water, sanitation, handwashing, and nutrition with nutrition alone, and combined water, sanitation, and handwashing. A double-sized active control group was used to increase power because there were six separate intervention comparisons against control.<sup>10</sup> Households in the active control and all intervention groups were visited by community-based health promoters monthly to measure the child's mid-upper arm circumference. Health promoters did not visit households in passive control clusters. Measurement of outcomes, as well as water, sanitation, handwashing, and nutrition characteristics were measured in the passive control group at the same times as in other groups. The study design and rationale have been published previously.<sup>10</sup>

The study protocol was approved by the Committee for the Protection of Human Subjects at the University of California, Berkeley (protocol number 2011-09-3654), the institutional review board at Stanford University (IRB-23310), and the scientific and ethics review unit at the Kenya Medical Research Institute (protocol number SSC-2271). Under direction of the study investigators, Innovations for Poverty Action (IPA) was responsible for intervention delivery and data collection.

## Participants

Villages were eligible for selection into the study if they were rural, most of the population relied on communal water sources and had unimproved sanitation facilities, and there were no other ongoing water, sanitation, handwashing, or nutrition programmes. Participants were identified through a complete census of eligible villages. Within selected villages, women were eligible to participate if they reported that they were in their second or third trimester of pregnancy, planned to continue to live at their current residence for the next 2 years, and could speak Kiswahili, Luhya, or English well enough to respond to an interviewer administered survey. IPA staff formed clusters from one to three neighbouring villages to have six or more pregnant women per cluster after the enrolment survey. Outcomes were assessed in the children born from these pregnancies (index children), including twins. Although the study area is one of the areas with the highest HIV prevalence in Kenya, according to the 2012 Kenya AIDS Indicator Survey, the prevalence in women aged 15–64 years in the study area was below 0.8% (that survey did not include testing of children). Because there would not have been sufficient

sample size to allow for subgroup analysis by HIV status, no attempt was made to identify participants who were HIV positive. Participants gave written informed consent before enrolment.

## Randomisation and masking

Clusters were randomly allocated to treatment using a random number generator with reproducible seed at the University of California, Berkeley. Groups of nine geographically-adjacent clusters were block-randomised into a double-sized active control; passive control; water; sanitation; handwashing; water, sanitation, and handwashing; nutrition; or water, sanitation, handwashing, and nutrition. Allocation by cluster identification number was communicated directly to the field team; investigators remained blinded to treatment assignments. Blinding of participants was not possible. Participants were informed of their treatment assignment after baseline data collection and might have known the treatment assignment of nearby villages. The health promoters and staff who delivered the interventions were not involved in data collection, but the data collection team could have inferred treatment status if they saw intervention materials in study communities.

## Procedures

The interventions were designed to maximise adherence to behaviours that could protect children from exposure to pathogens in their environment and improve diet quality. Formative research in the study area concluded that the health benefits of target behaviours were already well understood, but this knowledge was not sufficient to lead to action. As such, the behaviour change strategy and intervention materials were selected to create enabling environments, build supportive social norms, and target emotional drivers of decision making. The messages and delivery modes for the behaviour change strategy drew from existing information, education, and communication materials from organisations such as WHO, the Kenyan Government, UNICEF, and the Alive and Thrive network, and extensive previous qualitative work on the drivers of handwashing behaviours. Monthly visit modules were developed and pilot-tested to provide behavioural recommendations to mothers and other caregivers using key thematic constructs of convenience, nurturing care, and aspiration. We did a pilot randomised controlled trial<sup>11</sup> to test the feasibility and acceptability of all the interventions and to collect data that allowed us to optimise the ratio of community-based promoters to study participants. To identify and correct systematic problems with adherence, staff confirmed that intervention materials were delivered to all study participants at the outset of the trial, and collected monitoring data on availability of intervention materials and recommended behaviours during unannounced visits to a random sample of at least 20% of participants in intervention groups 2, 6, 10, and 19 months after the interventions began.

Community-based promoters for intervention and active control groups were nominated by study mothers and other mothers of children younger than 3 years in the community. A second promoter was added if there were more than ten participants (single groups) or more than eight participants (combined groups) in the cluster, giving a total of 1031 promoters. Promoters attended 2 days (active control), 6 days (single groups), or 7 days (combined groups) of initial training led by study staff on how to measure mid-upper-arm circumference, communication skills, intervention-specific behaviour change messages and intervention materials, and the information they were expected to report to IPA. Refresher trainings were done 6, 12, and 18 months after the initial training. At 2, 4, 9, 15, and 21 months, study staff met with promoters in their clusters to observe visits and offer supportive supervision. Study staff called promoters monthly to collect information on their activities, intervention adherence in the households they visited, referrals to health centres, and births or deaths of study children. Promoters received a branded T-shirt, a mobile phone, job aids and intervention materials, and compensation of approximately US\$15 per month for the first 6 months when they had more intensive engagement with the study participants, and \$9 per month thereafter (the prevailing daily wage for unskilled labour in the study area is \$1–2). Promoters were instructed to visit all participants in their cluster monthly and measure the child's arm circumference or the pregnant mother's abdomen.

In intervention groups, promoters engaged study participants and other compound members through interactive activities such as guided discussions using visual aids, song, and storytelling; resupplied consumable intervention materials; encouraged consistent practice of targeted behaviours; and helped troubleshoot barriers to adherence, including problems with intervention hardware and behavioural barriers. Promoters were provided with detailed plans for every visit, including key messages, scripts for discussing visual aids, and instructions for activities that emphasised the learning objectives. Visits lasted about 10 min in the active control group and 45–60 min in intervention groups during the first year when the key messages were conveyed. In the second year, promoters reinforced messages to maintain habits. All groups used messages on themes of nurture, aspiration, and self-efficacy, particularly in the context of a new birth. Interventions used convenience and social norms to encourage target behaviours.

In the three intervention groups that included water, promoters advocated treatment of drinking water with sodium hypochlorite. Chlorine dispensers for convenient water treatment at the point of collection were installed at an average of five communal water sources in the cluster and refilled as needed. Every 6 months, households in study compounds were given a 1 L bottle of chlorine for point-of-use water treatment in case households collected rainwater or used a source without

a dispenser. Promoters used chlorine test strips during their regular visits to determine if the household was using chlorine, and negative results stimulated conversation about addressing barriers to chlorination.

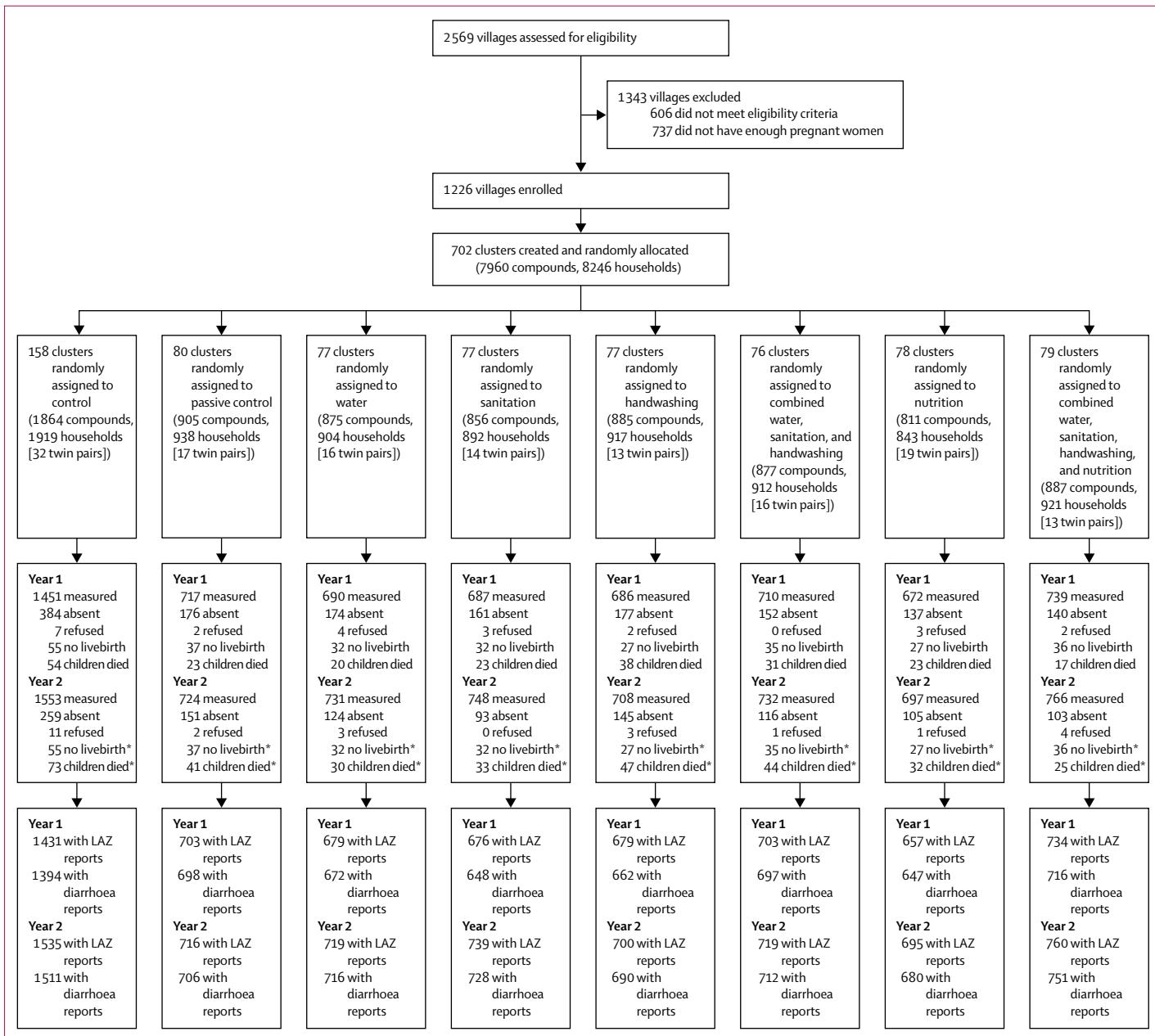
In the three intervention groups that included sanitation, promoters advocated using latrines for defecation and safe disposal of children's and animals' faeces into a latrine. Existing unimproved latrines in study households were upgraded to improved latrines by installing a plastic slab, which also had a tight-fitting lid over the hole. New latrines were constructed for study households that did not have a latrine or whose latrine was unlikely to last for 2 years. All households in study compounds received a sani-scoop with a paddle as a dedicated faeces-removal tool. Finally, all households with children younger than 3 years in study compounds received plastic potties to facilitate toilet training and transfer of child faeces to the latrine.

In the three intervention groups that included handwashing, promoters advocated handwashing with soap before handling food and after defecation (including assisting a child). Study compounds were given two permanent, water-frugal handwashing stations intended to be installed near the food preparation area and the latrine. Handwashing stations were constructed of painted metal, with two foot-pedal-operated jerry-cans that dispensed a light flow of rinse water and soapy water. Promoters added chunks of bar soap to the soapy water container quarterly.

In the two intervention groups that included nutrition, a set of ten age-targeted modules were developed to enable promoters to advocate for best practices in maternal, infant, and young child feeding: recommendations for dietary diversity during pregnancy and lactation, early initiation of breastfeeding, exclusive breastfeeding until 6 months, introduction of appropriate and diverse complementary foods at 6 months, and continued breastfeeding through 24 months. Facilitators and barriers to behaviour change were elicited using formative research and health promoter guides were developed to address common barriers and questions. Study mothers with children between 6–24 months were provided with two 10 g sachets per day of a small quantity of lipid-based nutrient supplement (LNS; Nutriset; Malauny, France) that could be mixed into the child's food. LNS provided 118 kcal per day and 12 essential vitamins and ten minerals. Promoters explained that LNS was not to replace breastfeeding or complementary foods.

Promoters and intervention materials were introduced at community meetings roughly 6 weeks after enrolment. All interventions were delivered within 3 months of enrolment (appendix p 1). LNS was introduced to each child when they turned 6 months old. All handwashing stations and latrines were inspected within a month of construction, and a subset of households was periodically visited to observe group-specific indicators of intervention adherence. These data alerted study investigators to any

For intervention-specific training materials see  
<https://osf.io/fs23x>

**Figure 1: Trial profile and analysis populations for primary outcomes**

LAZ=length-for-age Z scores. \*Stillbirth and child death counts are cumulative.

issues with intervention implementation so they could be addressed consistently across all clusters and groups.

The enrolment survey included baseline demographics; assets; water, sanitation, and handwashing infrastructure; and target behaviours. Follow-up at 1 year and 2 years after intervention delivery consisted of an unannounced visit to study compounds to observe objective indicators of target behaviours (in all groups other than the passive control) and, on the following day, growth and health outcome measurements at a central location in the cluster (eg, a church or school).

Children identified as possibly malnourished (mid-upper-arm circumference <11.5 cm), either by the promoter during routine visits or by study staff during follow-up measurements, were referred to health facilities for treatment.

### Outcomes

Adherence to the interventions was assessed using objective, observable indicators where possible (appendix pp 2, 3). We calculated Z scores for length for age, weight for length, weight for age, and head circumference for

age using the WHO 2006 child growth standards. All child deaths reported by the health promoters were confirmed by a staff nurse who visited households. All outcomes were prespecified. Primary outcomes were caregiver-reported diarrhoea in the past 7 days (based on all data from year 1 and year 2) and length-for-age Z score at year 2 in index children. Secondary and tertiary outcomes reported in this paper are length-for-age Z score at year 1; weight-for-length Z score, weight-for-age Z score, head circumference-for-age Z score at year 1 and year 2; prevalence of stunting (length-for-age Z score less than -2), severe stunting (length-for-age Z score less than -3), wasting (weight-for-length Z score less than -2), and underweight (weight-for-age Z score less than -2); and all-cause mortality. We excluded children from Z-score analyses if their measurements were outside biologically plausible ranges following WHO recommendations. More details on exclusion criteria, measurement protocols, and outcome definitions are in the appendix (p 1).

### Statistical analyses

Sample size calculations for the two primary outcomes were based on a minimum detectable effect of 0·15 in length-for-age Z score (intraclass correlation of 0·02 in our pilot study) and a relative risk of diarrhoea of 0·7 or smaller (assuming a 7-day prevalence of 12% in the active control group based on a pilot study to inform this trial) for a comparison of any intervention with the double-sized control group, assuming a type I error ( $\alpha$ ) of 0·05 and power ( $1-\beta$ ) of 0·8, a one-sided test for a two-sample comparison of means, and 10% loss to follow-up.<sup>10,11</sup> Sample size calculations indicated 80 clusters per group, each with ten children.

Two biostatisticians, blinded to treatment assignment, independently replicated the analyses following the prespecified analysis plan with minor updates.<sup>10</sup> We analysed participants according to their randomised assignment (intention to treat), regardless of adherence to the intervention, using the active control group as the comparator. We used paired *t* tests for unadjusted length-for-age Z score comparisons and the Mantel-Haenszel prevalence ratio and difference for unadjusted diarrhoea and stunting comparisons, with randomisation block defining matched pairs or stratification. In secondary analyses, we estimated prevalence ratios and differences, adjusting for baseline covariates using targeted maximum likelihood estimation.<sup>12</sup> Analyses were done in R (version 3.2.3). We tested for the presence of between-cluster spillover effects using a non-parametric method described in the prespecified analysis plan, which tested whether primary outcomes were the same in control households with more versus fewer households receiving interventions within a 2 km radius. In an analysis that was not prespecified, we tested for intervention effects on diarrhoea using only year 1 data.

The trial is registered at ClinicalTrials.gov, number NCT01704105. IPA convened a data and safety monitoring board.

### Role of the funding source

The funders of the study approved the study design, but had no role in data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

### Results

2569 villages were assessed for eligibility, of which 606 were excluded on the basis of village-level characteristics (primarily not meeting the study's rural criteria). 1226 villages were grouped into 702 clusters that had six or more pregnant women (figure 1). Between Nov 27, 2012, and May 21, 2014, 8246 pregnant women were enrolled in the study. 281 women did not have a livebirth and 140 women delivered twins. After at least three attempts to measure each child, 6659 (86%) of 7780 surviving children were measured at year 2, with diarrhoea reports for 6494 children and length-for-age Z score measures for 6583 children. Children were aged 2–18 months (median 12 months) at 1-year follow-up (January, 2014, to June, 2015) and aged 16–31 months (median 25 months) at 2-year follow-up (February, 2015, to July, 2016), but 11184 (87%) of 12841 children were in the target age ranges of 9–15 months at year 1 and 21–27 months at year 2 (appendix p 12).

Household characteristics were similar across groups at enrolment (table 1). Roughly three-quarters of participants collected drinking water from an improved source, but had to walk at least 10 min on average to the source. Over 80% of households owned a latrine, but less than 20% had access to an improved latrine. Less than 15% of households had soap available at a handwashing location. The prevalence of moderate-to-severe household hunger was 12% or lower.

Around 75% of households were visited by their promoter within the past month at year 1, but frequency of contact fell by year 2, with 40% or fewer households reporting a visit in the past month in each group (monitoring data suggest that most households were still visited at least every other month during the second year of the trial; see details in the appendix p 2, and table 2). Slightly less than half of households had detectable free chlorine in stored drinking water in the water group. Around 40% of drinking water samples tested in the water, sanitation, handwashing, and nutrition group had detectable free chlorine at year 1, which fell to around 20% by year 2. A high proportion of households (75%) had improved latrine access, which remained stable in year 1 and year 2 in households in the sanitation groups, increasing by more than 50% compared with the active control group. Reported safe disposal of children's faeces into a latrine fell by roughly half in all

For more on the updates to the analysis plan see <https://osf.io/7urqa/>

	Active control (N=1919)	Passive control (N=938)	Water (N=904)	Sanitation (N=892)	Handwashing (N=917)	Water, sanitation, and handwashing (N=912)	Nutrition (N=843)	Water, sanitation, handwashing, and nutrition (N=921)
<b>Maternal</b>								
Age (years)	26 (6)	26 (7)	26 (6)	26 (7)	26 (6)	26 (6)	26 (6)	26 (6)
Completed at least primary education	916 (48%)	441 (47%)	447 (50%)	430 (48%)	402 (44%)	430 (47%)	409 (49%)	438 (48%)
Height (cm)	160 (6)	160 (7)	160 (6)	160 (6)	160 (6)	160 (6)	160 (7)	160 (7)
Study child is firstborn	490 (26%)	237 (25%)	205 (23%)	222 (25%)	208 (23%)	191 (21%)	206 (24%)	225 (25%)
<b>Paternal</b>								
Completed at least primary education	1098 (62%)	521 (60%)	532 (64%)	482 (58%)	500 (59%)	521 (61%)	491 (64%)	526 (62%)
Works in agriculture	749 (41%)	376 (43%)	378 (44%)	362 (43%)	363 (42%)	374 (43%)	343 (43%)	372 (43%)
<b>Household</b>								
Number of households per compound	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)	2 (1)
Number of people per compound	8 (5)	8 (6)	8 (6)	8 (5)	8 (6)	8 (5)	8 (7)	8 (5)
Number of children <18 years in the household	3 (2)	3 (2)	3 (2)	3 (2)	3 (4)	3 (2)	3 (2)	3 (4)
Has electricity	122 (6%)	51 (5%)	60 (7%)	73 (8%)	67 (7%)	64 (7%)	58 (7%)	67 (7%)
Has a cement floor	107 (6%)	50 (5%)	71 (8%)	48 (5%)	41 (4%)	50 (5%)	48 (6%)	55 (6%)
Has an iron roof	1302 (68%)	600 (64%)	610 (68%)	587 (66%)	581 (63%)	574 (63%)	580 (69%)	615 (67%)
Owns a mobile phone	1526 (80%)	742 (79%)	705 (78%)	690 (77%)	722 (79%)	722 (79%)	685 (81%)	730 (79%)
Owns a motorcycle	185 (10%)	75 (8%)	81 (9%)	72 (8%)	91 (10%)	72 (8%)	81 (10%)	71 (8%)
<b>Drinking water</b>								
Primary drinking water source is improved*	1446 (76%)	699 (75%)	679 (75%)	675 (76%)	708 (78%)	624 (69%)	603 (72%)	697 (76%)
One-way walking time to primary water source (min)	11 (12)	12 (16)	12 (30)	10 (10)	11 (13)	11 (13)	11 (12)	11 (12)
Reported treating stored water	196 (13%)	92 (12%)	81 (11%)	94 (13%)	96 (13%)	97 (13%)	79 (12%)	106 (14%)
<b>Sanitation</b>								
Always or usually use primary toilet for defecation								
Men	1778 (95%)	867 (95%)	828 (94%)	810 (94%)	845 (95%)	851 (95%)	785 (95%)	854 (95%)
Women	1822 (96%)	898 (96%)	868 (96%)	840 (94%)	871 (96%)	877 (96%)	812 (96%)	872 (95%)
Daily defecating in the open								
Children aged 3 to <8 years	145 (12%)	87 (14%)	74 (13%)	68 (13%)	81 (14%)	75 (13%)	82 (15%)	75 (12%)
Children aged 0 to <3 years	789 (78%)	378 (77%)	376 (80%)	370 (75%)	358 (76%)	394 (77%)	363 (79%)	388 (78%)
Latrine								
Own any latrine	1561 (82%)	774 (83%)	750 (83%)	722 (81%)	756 (83%)	754 (83%)	701 (83%)	764 (83%)
Access to improved latrine	309 (17%)	153 (17%)	150 (18%)	131 (16%)	157 (19%)	153 (18%)	119 (15%)	143 (16%)
Human faeces observed in the compound	163 (9%)	79 (8%)	66 (7%)	72 (8%)	84 (9%)	73 (8%)	73 (9%)	87 (9%)
<b>Handwashing location</b>								
Has water within 2 m of handwashing location	487 (25%)	236 (25%)	242 (27%)	245 (28%)	245 (27%)	251 (28%)	228 (27%)	249 (27%)
Has soap within 2 m of handwashing location	164 (9%)	94 (10%)	91 (10%)	75 (8%)	83 (9%)	115 (13%)	90 (11%)	87 (9%)
<b>Food security</b>								
Prevalence of moderate-to-severe household hunger†	203 (11%)	113 (12%)	106 (12%)	91 (10%)	92 (10%)	101 (11%)	98 (12%)	104 (11%)

Data are n (%) or mean (SD). Percentages were calculated from smaller denominators than those shown at the top of the table for all variables because of missing values. \*Defined by WHO UNICEF Joint Monitoring Program's definition for an improved water source. †Assessed by the Household Food Insecurity Access Scale.

Table 1: Baseline characteristics by intervention group

groups between year 1 and year 2, although the practice remained over twice as likely in the groups that included sanitation compared with other groups at year 1 and year 2. More than 75% of households in the intervention groups that included handwashing had water and soap available at a handwashing location at year 1, but this indicator also fell to about 20% by year 2. Adherence to LNS

recommendations was high ( $\geq 95\%$ ) at year 1 and year 2, with children consuming a few more LNS sachets per month on average than would be expected at year 2. Across all indicators, adherence was comparable between the water, sanitation, and handwashing group and the water, sanitation, handwashing, and nutrition group compared with single intervention groups.

	Active Control (N=1919)	Passive Control (N=938)	Water (N=904)	Sanitation (N=892)	Handwashing (N=917)	Water, sanitation, and handwashing (N=912)	Nutrition (N=843)	Water, sanitation, handwashing, and nutrition (N=921)
<b>Number of compounds assessed</b>								
Enrolment	1913/1919 (100%)	936/938 (100%)	902/904 (100%)	890/892 (100%)	914/917 (100%)	912/912 (100%)	843/843 (100%)	918/921 (100%)
Year 1	1043/1919 (54%)	..	477/904 (53%)	473/892 (53%)	501/917 (55%)	536/912 (59%)	454/843 (54%)	493/921 (54%)
Year 2	1458/1919 (76%)	..	696/904 (77%)	712/892 (80%)	690/917 (75%)	675/912 (74%)	650/843 (77%)	735/921 (100%)
<b>Visited by promoter in past month</b>								
Enrolment	..	..	..	..	..	..	..	..
Year 1	666/980 (68%)	..	338/445 (76%)	333/445 (75%)	333/480 (69%)	386/512 (75%)	344/433 (79%)	388/474 (82%)
Year 2	492/1412 (35%)	..	255/680 (37%)	278/692 (40%)	228/678 (34%)	241/649 (37%)	251/635 (40%)	259/710 (36%)
<b>Stored drinking water has detectable free chlorine</b>								
Enrolment	44/1529 (3%)	24/736 (3%)	20/720 (3%)	20/715 (3%)	30/743 (4%)	29/711 (4%)	14/661 (2%)	26/729 (4%)
Year 1	25/847 (3%)	..	151/385 (39%)	18/367 (5%)	20/417 (5%)	180/424 (42%)	9/392 (2%)	156/367 (43%)
Year 2	38/1365 (3%)	..	144/637 (23%)	17/641 (3%)	16/648 (2%)	112/598 (19%)	15/614 (2%)	128/652 (20%)
<b>Access to improved latrine</b>								
Enrolment	309/1788 (17%)	153/878 (17%)	150/844 (18%)	131/836 (16%)	157/847 (19%)	153/867 (18%)	119/794 (15%)	143/872 (16%)
Year 1	178/993 (18%)	..	74/461 (16%)	409/458 (89%)	65/486 (13%)	472/526 (90%)	63/424 (15%)	425/477 (89%)
Year 2	271/1381 (20%)	..	128/664 (19%)	534/683 (78%)	119/654 (18%)	529/644 (82%)	99/613 (16%)	561/706 (79%)
<b>Child faeces safely disposed of</b>								
Enrolment	114/721 (16%)	51/323 (16%)	53/310 (17%)	67/347 (19%)	54/319 (17%)	65/369 (18%)	33/310 (11%)	56/353 (16%)
Year 1	338/903 (37%)	..	158/424 (37%)	317/412 (77%)	157/431 (36%)	326/463 (70%)	155/391 (40%)	287/432 (66%)
Year 2	136/1320 (10%)	..	52/625 (8%)	240/643 (37%)	62/616 (10%)	205/597 (34%)	52/578 (9%)	219/657 (33%)
<b>Handwashing location has water and soap</b>								
Enrolment	96/1913 (5%)	58/936 (6%)	56/902 (6%)	42/890 (5%)	52/914 (6%)	64/912 (7%)	57/843 (7%)	53/918 (6%)
Year 1	124/1043 (12%)	..	53/477 (11%)	49/473 (10%)	381/501 (76%)	416/536 (78%)	61/454 (13%)	381/493 (77%)
Year 2	127/1458 (9%)	..	49/696 (7%)	57/712 (8%)	159/690 (23%)	130/675 (19%)	76/650 (12%)	152/735 (21%)
<b>LNS sachets consumed (% expected)*</b>								
Enrolment	..	..	..	..	..	..	..	..
Year 1	..	..	..	..	..	..	5264/5558 (95%)	5583/5838 (96%)
Year 2	..	..	..	..	..	..	3577/3136 (114%)	4028/3458 (116%)

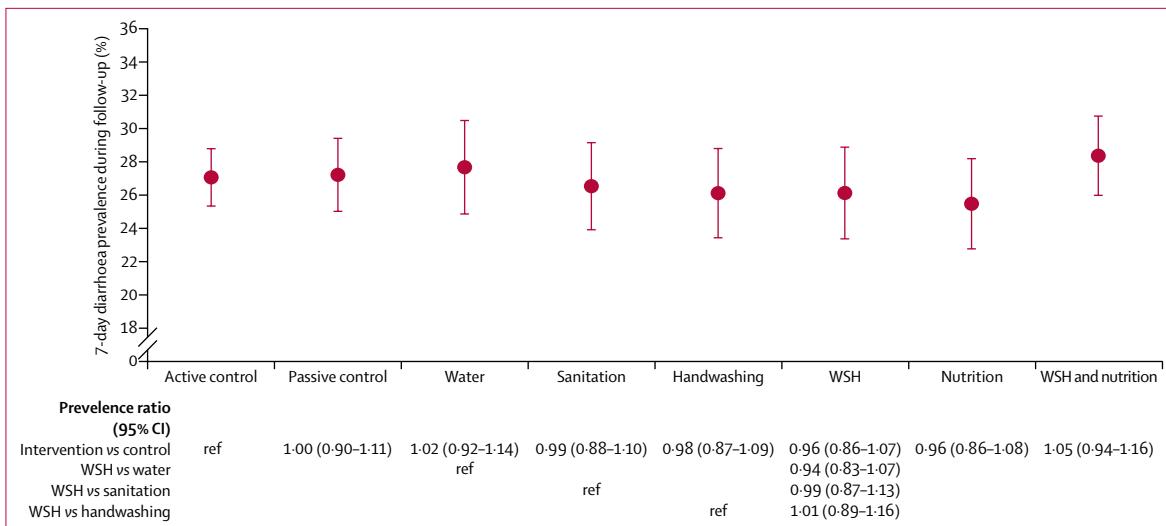
Data are n (%), or %. Free chlorine in drinking water and LNS consumption were not measured at enrolment and were only measured in a subset of groups. LNS=lipid-based nutrient supplement. \*LNS adherence measured as reported proportion of 14 sachets consumed in the past week in index children aged 6–24 months.

Table 2: Measures of intervention adherence by study group at enrolment, 1-year follow-up, and 2-year follow-up

Diarrhoea prevalence over the past 7 days (combining data from year 1 and year 2) was 27·1% in children in the active control group (figure 2, table 3). The intracluster correlation for diarrhoea was 0·012. Compared with the active control group, the diarrhoea prevalence ratios across all groups were not significantly different from one and differences were not significantly different from zero (figure 2, table 3). Diarrhoea prevalence was the same in the combined water, sanitation, and handwashing group and the individual water, sanitation, and handwashing groups. Although adherence to the water and handwashing interventions was higher in year 1 than in year 2, in an analysis that was not prespecified, diarrhoea prevalence was not significantly lower in any of the intervention groups at year 1 (appendix p 12). The high diarrhoea prevalence was fairly stable over 2 years of follow-up and there were no apparent seasonal trends (appendix p 13). Although we had prespecified a sensitivity analysis by age group of child at year 2, we did not complete this

analysis because sample sizes in the age group strata were smaller than expected.

By year 2, when children were between 16 and 31 months old (median 25 months), mean length-for-age Z score in children in the active control group was -1·54 (SD 1·11; figure 3). The intracluster correlation for length-for-age Z score was 0·037. Compared with the active control group, only nutrition and combined water, sanitation, handwashing, and nutrition had higher length-for-age Z score (mean difference in score 0·13 [95% CI 0·01–0·25] for nutrition; 0·16 [0·05–0·27] for combined water, sanitation, handwashing, and nutrition; figure 3). Children in the combined water, sanitation, handwashing, and nutrition group were not significantly taller than children in the nutrition group (mean difference 0·04 [95% CI -0·11 to 0·19]; figure 3). Most length-for-age Z score gains in these two groups were already apparent by year 1 (0·11 [-0·01 to 0·22] for nutrition; 0·12 [0·01–0·22] for combined water, sanitation, handwashing, and nutrition; appendix p 14).



**Figure 2:** Intervention effects on diarrhoea prevalence 1 and 2 years after intervention

Data are mean (95% CI). ref=reference. WSH=water, sanitation, and handwashing.

Mean weight-for-age Z score at year 2 was higher in children in the nutrition and combined water, sanitation, handwashing, and nutrition groups than the mean of -0.72 (SD 1.01) in the active control group (table 4). Children in the active control group were close to WHO standards for weight-for-length Z score; however, weight-for-length Z score at year 2 was higher in the combined water, sanitation, handwashing, and nutrition group (table 4). There were no differences in mean head circumference for age Z score at year 2 between children in any of the intervention groups and those in the active control group. Results were similar at year 1, with the exception that differences in mean weight-for-length Z score between the active control and two groups with the nutrition intervention appear to have been numerically larger at year 1 (appendix p 15).

Compared with the active control group, a smaller proportion of children in the combined water, sanitation, handwashing, and nutrition group were stunted (too short for their age; -5.4 percentage points [95% CI -9.4 to -1.4]), severely stunted (-2.7 percentage points [-5.1 to -0.2]), or underweight (-3.0 percentage points [-5.4 to -0.6]; table 5); no other groups appeared to affect these outcomes. Notably, there were no significant differences between the combined water, sanitation, handwashing, and nutrition and nutrition groups for any growth outcomes. 1% of active control children were wasted and the proportions were similar across all groups.

Differences in growth outcomes between the active control and intervention groups were similar in magnitude and precision when estimated using adjusted models (appendix pp 16–19). We found no evidence of between-cluster spillover effects (appendix p 20).

The cumulative incidence of all-cause mortality was 3.9% in the active control and ranged from 5.3% in the handwashing group to 2.8% in the combined water,

	Mean* prevalence	Unadjusted† prevalence difference (95% CI)	Adjusted‡ prevalence difference (95% CI)
<b>Intervention vs active control</b>			
Active control	27.1%	..	..
Passive control	27.2%	-0.0 (-2.9 to 2.9)	-0.4 (-3.3 to 2.4)
Water	27.7%	0.7 (-2.3 to 3.6)	0.4 (-3.2 to 4.0)
Sanitation	26.5%	-0.3 (-3.3 to 2.6)	-0.3 (-3.2 to 2.6)
Handwashing	26.1%	-0.6 (-3.5 to 2.3)	-1.1 (-4.0 to 1.8)
Water, sanitation, and handwashing	26.1%	-1.2 (-4.1 to 1.7)	-1.1 (-4.3 to 2.0)
Nutrition	25.5%	-1.0 (-4.0 to 2.0)	-0.6 (-4.0 to 2.7)
Water, sanitation, handwashing, and nutrition	28.4%	1.2 (-1.7 to 4.1)	0.7 (-2.4 to 3.7)
<b>Water, sanitation, and handwashing vs single groups</b>			
Water, sanitation, and handwashing	26.1%	..	..
Water	27.7%	-1.6 (-5.1 to 1.9)	-2.1 (-6.0 to 1.8)
Sanitation	26.5%	-0.2 (-3.6 to 3.2)	-0.8 (-4.5 to 2.9)
Handwashing	26.1%	0.4 (-3.2 to 3.9)	0.5 (-3.6 to 4.5)

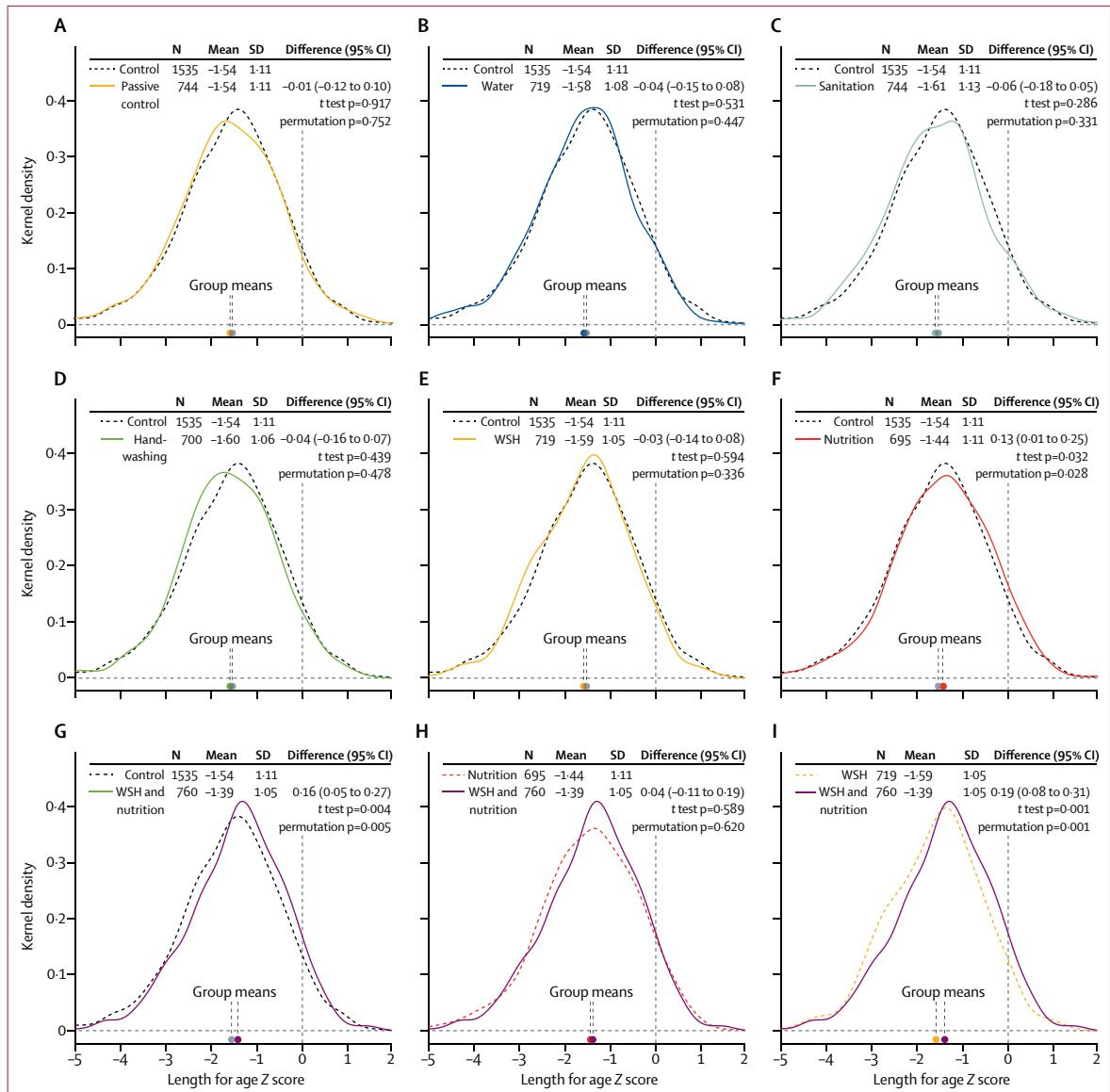
\*Post-intervention measurements in years 1 and 2 combined. †Unadjusted estimates were estimated using a pair-matched Mantel-Haenszel analysis. ‡Adjusted for prespecified covariates using targeted maximum likelihood estimation with data-adaptive model selection: field staff who collected data, month of measurement, household food insecurity, child age, child sex, mother's age, mother's height, mothers education level, number of children <18 years in the household, number of individuals living in the compound, distance in minutes to the primary water source, household roof, floor, wall materials, and household assets.

**Table 3:** Diarrhoea prevalence from 1 and 2 years (combined) after intervention

sanitation, handwashing, and nutrition group; none of the differences between intervention groups and the active control were statistically significant at  $\alpha=0.05$  (figure 1, appendix p 21).

## Discussion

In the WASH Benefits cluster-randomised controlled trial, we found no effect of any interventions (improved



**Figure 3: Intervention effects on length-for-age Z scores in 6583 children after 2 years of intervention**

Kernel density plots show the distribution of length-for-age Z scores; dashed lines are the comparison group distribution and solid lines are the active comparator distribution. (A) Passive control vs active control. (B) Water vs active control. (C) Sanitation vs active control. (D) Handwashing vs active control. (E) WSH vs active control. (F) Nutrition vs active control. (G) WSH and nutrition vs active control. (H) WSH and nutrition vs nutrition. (I) WSH and nutrition vs WSH. p values for t test are for differences in group means from zero; permutation p values test the null hypothesis of no difference between groups using a Wilcoxon signed-rank test statistic. WSH=water, sanitation, and handwashing.

water quality, safe sanitation, handwashing, nutrition, or combinations of the interventions) on caregiver-reported diarrhoea prevalence during the first 2 years of life, and improvements in growth were only observed in groups including the nutrition intervention (maternal, infant, and young child feeding counselling and LNS distribution). With a large sample size and high-quality anthropometric measurements, this trial was powered to detect small effects in diarrhoea prevalence and length-for-age Z score had they been present. Lower adherence to the water and handwashing interventions

by the end of the 2 years of intervention does not seem to be the only explanation for the absence of benefits: there were also no reductions in diarrhoea or improvements in growth in children in the water, handwashing, sanitation, or combined water, sanitation, and handwashing groups even in the first year (a typical measurement point in previous trials), when community-based promoters were most active and adherence was higher, whereas almost all of the growth benefits in the nutrition group and combined water, sanitation, handwashing, and nutrition group were already manifest in the first year. Adherence

	N	Mean (SD)	Difference vs active control (95% CI)	Difference vs nutrition (95% CI)	Difference vs water, sanitation, and handwashing (95% CI)
<b>Weight-for-age Z score</b>					
Active control	1548	-0.72 (1.01)	..	..	..
Passive control	721	-0.76 (0.97)	-0.04 (-0.13 to 0.05)	..	..
Water	727	-0.73 (1.00)	0.00 (-0.10 to 0.10)	..	..
Sanitation	747	-0.80 (1.05)	-0.07 (-0.19 to 0.04)	..	..
Handwashing	706	-0.77 (1.01)	-0.05 (-0.15 to 0.05)	..	..
Water, sanitation, and handwashing	725	-0.77 (0.98)	-0.02 (-0.12 to 0.08)	..	..
Nutrition	698	-0.65 (0.98)	0.11 (0.00 to 0.21)	..	..
Water, sanitation, handwashing, and nutrition	765	-0.60 (0.96)	0.14 (0.04 to 0.25)	0.04 (-0.07 to 0.15)	0.17 (0.05 to 0.30)
<b>Weight-for-length Z score</b>					
Active control	1536	0.11 (0.94)	..	..	..
Passive control	717	0.08 (0.92)	-0.04 (-0.13 to 0.05)	..	..
Water	719	0.14 (0.95)	0.04 (-0.06 to 0.13)	..	..
Sanitation	740	0.05 (0.97)	-0.05 (-0.14 to 0.05)	..	..
Handwashing	700	0.09 (0.93)	-0.02 (-0.11 to 0.06)	..	..
Water, sanitation, and handwashing	714	0.08 (0.92)	-0.02 (-0.10 to 0.07)	..	..
Nutrition	695	0.14 (0.92)	0.04 (-0.05 to 0.14)	..	..
Water, sanitation, handwashing, and nutrition	762	0.18 (0.90)	0.09 (0.00 to 0.19)	0.04 (-0.05 to 0.13)	0.12 (0.00 to 0.23)
<b>Head circumference-for-age Z score</b>					
Active control	1545	-0.27 (1.02)	..	..	..
Passive control	719	-0.27 (1.05)	0.00 (-0.10 to 0.10)	..	..
Water	727	-0.27 (1.03)	0.02 (-0.08 to 0.12)	..	..
Sanitation	745	-0.27 (1.04)	0.01 (-0.09 to 0.11)	..	..
Handwashing	705	-0.29 (0.99)	0.00 (-0.10 to 0.10)	..	..
Water, sanitation, and handwashing	729	-0.30 (0.96)	-0.03 (-0.12 to 0.06)	..	..
Nutrition	695	-0.23 (0.99)	0.05 (-0.05 to 0.15)	..	..
Water, sanitation, handwashing, and nutrition	763	-0.22 (0.99)	0.05 (-0.04 to 0.15)	-0.02 (-0.14 to 0.10)	0.08 (-0.05 to 0.20)
Median child age at 2-year follow-up was 2.05 years (IQR 1.93–2.16). All three secondary outcomes were prespecified.					

Table 4: Child growth Z scores at 2-year follow-up

to the interventions was comparable to or better than what a government or large non-governmental organisation might hope to achieve at scale (appendix p 22), with increases in adherence indicators of 30 percentage points or higher in all intervention groups relative to the control in the first year.

These findings contrast with several systematic reviews<sup>13–15</sup> that have found significant protective benefits of water, sanitation, and hygiene interventions (including handwashing) on diarrhoea in efficacy trials, although most of these studies were shorter and had higher adherence. Results from other trials<sup>16–18</sup> also showed no effect of improved sanitation on diarrhoea, although differences in contexts and interventions complicate comparisons between these trials. Our trial differed from previous trials in that the intervention shifted households from unimproved sanitation (rather than open defecation) to improved sanitation. Additionally, the prevalence of diarrhoea in this study population was high, consistent with prevalence in 12–23-month-old infants measured in the 2014 Kenya Demographic and Health Survey.<sup>19</sup>

A systematic review and meta-analysis<sup>20</sup> of the effects of water quality and supply, sanitation, and hygiene interventions to improve growth identified only five randomised controlled trials of water or handwashing interventions, which did not suggest strong effects on growth, perhaps in part because the interventions lasted only 9–12 months. Since then, five more randomised trials of sanitation interventions have generated mixed evidence on child growth effects: two trials done in India and one in Indonesia had low adherence and no effect, and two done in settings with high rates of open defecation in India and Mali showed improvements in length-for-age Z score of 0.18–0.40 in children younger than 5 years.<sup>16–18,20–22</sup> The sanitation intervention in our trial was aligned with the focus on improved latrines initiated under the Millennium Development Goals, and the Sustainable Development Goals' recognition that children's faeces also need to be safely disposed of. This trial and its companion trial<sup>9</sup> in Bangladesh suggest that a compound-level approach to upgrading existing latrines and safely disposing of children's faeces is not sufficient

	n/N (%)	Difference vs active control (95% CI)	Difference vs nutrition (95% CI)	Difference vs water, sanitation, and handwashing (95% CI)
<b>Stunting*</b>				
Active control	483/1535 (31%)	..	..	..
Passive control	223/716 (31%)	-1.7 (-5.9 to 2.5)	..	..
Water	233/719 (32%)	0.1 (-4.2 to 4.3)	..	..
Sanitation	255/739 (35%)	2.3 (-2.0 to 6.6)	..	..
Handwashing	235/700 (34%)	0.8 (-3.5 to 5.1)	..	..
Water, sanitation, and handwashing	236/719 (33%)	1.3 (-3.0 to 5.6)	..	..
Nutrition	201/695 (29%)	-3.2 (-7.5 to 1.1)	..	..
Water, sanitation, handwashing, and nutrition	203/760 (27%)	-5.4 (-9.4 to -1.4)	-2.3 (-7.1 to 2.5)	-5.8 (-10.6 to -1.0)
<b>Severe stunting†</b>				
Active control	143/1535 (9%)	..	..	..
Passive control	62/716 (9%)	-0.8 (-3.3 to 1.8)	..	..
Water	69/719 (10%)	-0.5 (-3.2 to 2.2)	..	..
Sanitation	77/739 (10%)	1.0 (-1.8 to 3.7)	..	..
Handwashing	59/700 (8%)	-1.1 (-3.7 to 1.5)	..	..
Water, sanitation, and handwashing	65/719 (9%)	0.2 (-2.4 to 2.8)	..	..
Nutrition	55/695 (8%)	-1.6 (-4.2 to 1.0)	..	..
Water, sanitation, handwashing, and nutrition	55/760 (7%)	-2.7 (-5.1 to -0.2)	-0.9 (-3.7 to 2.0)	-2.7 (-5.6 to 0.2)
<b>Wasting‡</b>				
Active control	22/1536 (1%)	..	..	..
Passive control	10/717 (1%)	0.0 (-1.1 to 1.1)	..	..
Water	9/719 (1%)	-0.2 (-1.3 to 0.8)	..	..
Sanitation	19/740 (3%)	1.1 (-0.3 to 2.4)	..	..
Handwashing	6/700 (1%)	-0.5 (-1.5 to 0.4)	..	..
Water, sanitation, and handwashing	10/714 (1%)	0.2 (-0.9 to 1.2)	..	..
Nutrition	8/695 (1%)	-0.3 (-1.3 to 0.8)	..	..
Water, sanitation, handwashing, and nutrition	11/762 (1%)	-0.1 (-1.2 to 1.0)	0.2 (-1.0 to 1.4)	0.0 (-1.2 to 1.1)
<b>Underweight†</b>				
Active control	148/1548 (10%)	..	..	..
Passive control	70/721 (10%)	-0.4 (-3.0 to 2.2)	..	..
Water	76/727 (10%)	-0.1 (-2.8 to 2.7)	..	..
Sanitation	87/747 (12%)	1.6 (-1.2 to 4.4)	..	..
Handwashing	71/706 (10%)	0.5 (-2.2 to 3.3)	..	..
Water, sanitation, and handwashing	72/725 (10%)	0.5 (-2.3 to 3.2)	..	..
Nutrition	59/698 (8%)	-1.2 (-3.9 to 1.5)	..	..
Water, sanitation, handwashing, and nutrition	52/765 (7%)	-3.0 (-5.4 to -0.6)	-1.8 (-4.7 to 1.1)	-3.3 (-6.2 to -0.5)

Median child age at 2-year follow-up was 2.05 years (IQR 1.93–2.16). \*Prespecified secondary outcome. †Prespecified tertiary outcome.

Table 5: Proportion of children stunted, severely stunted, wasted, and underweight at 2-year follow-up

to improve child growth, and neither are water and handwashing interventions.

Conversely, counselling and LNS provided in the nutrition group improved length-for-age Z score by year 2. Compared with randomised controlled trials of LNS during complementary feeding, our finding of length-for-age Z score improvements of 0.13–0.16 in the nutrition groups falls in the middle of the spectrum between four trials: one from Malawi<sup>23</sup> that reported no effect on length-for-age Z score, one from Haiti<sup>24</sup> and one from Bangladesh<sup>25</sup> that reported an effect on length-for-age Z score comparable to this study, and one from Burkina Faso<sup>26</sup> that reported a

larger effect on length-for-age Z score. Thus, there appears to be consistent evidence that LNS distribution together with some promotion of improved infant and young child feeding can reduce growth faltering, although this approach falls far short of eliminating the problem. Interventions will likely need to address the complex set of underlying determinants of growth faltering, including prenatal or preconception factors. Future analyses will explore changes in feeding practices that resulted from the intervention.

Although there were more improvements in anthropometric measures in the combined water, sanitation, handwashing, and nutrition group versus active control

than in the nutrition versus active control group, the differences were of little clinical or statistical significance. We conclude that combining nutrition with water, sanitation, and handwashing did not provide additional growth benefits beyond nutrition alone. Although the effect of water, sanitation, handwashing, and nutrition on mortality was not significant, the lower mortality in that group is consistent with the statistically significant effect of water, sanitation, handwashing, and nutrition on mortality in the Bangladesh trial.<sup>9</sup> Pending analyses will evaluate potential differences in effects on other child health outcomes.

It is possible that the water, sanitation, and handwashing interventions delivered in this trial did not sufficiently address important transmission routes for enteric pathogens.<sup>11</sup> Although the sanitation intervention included a sani-scoop and messages about preventing children from being exposed to domestic animal faeces, the emphasis was mostly on behaviours related to human faeces and might not have protected children from zoonotic pathogens.<sup>27</sup> Although chlorination of water has the advantage of providing residual protection against recontamination, it is not effective against protozoa such as *Giardia lamblia* and *Cryptosporidium* spp, the latter of which was identified as one of the most common causes of moderate-to-severe diarrhoea in children 0–23 months in a neighbouring part of Kenya.<sup>28</sup> Other limitations of this trial include the inability to mask the interventions; the absence of observable indicators of actual behaviour for the handwashing, sanitation, and nutrition interventions; lower adherence to the water and hygiene interventions during the second year of the trial than in the first year; and the use of a compound-level sanitation intervention, as opposed to community-level. Because masking was not possible, we focused on objective, observable indicators whenever possible rather than self-reported behaviours, recognising that the availability of a latrine or handwashing station stocked with water and soap does not necessarily imply that the materials were used. Despite an intensive design process that drew heavily on best practices in behaviour change, incorporation of lessons learned from the pilot randomised controlled trial, thorough verification of availability of the intervention materials, and periodic monitoring of indicators of recommended behaviours, adherence to the water and handwashing interventions appeared to reduce sharply in the last months of the trial. The waning intensity of promotion activities after a reduction in the stipend given to the health promoters could at least partly explain the drop in adherence. Finally, by contrast with water, handwashing, and nutrition interventions that directly benefit households that adhere to the intervention, a sanitation intervention in only a subset of compounds might not be sufficient to protect against exposure to faecal contamination in the environment<sup>28</sup> that originates from other compounds in the community. We decided, however, to deliver

compound-level interventions based on evidence that child exposure to enteric pathogens during the first 2 years of life occurs predominantly within the household compound.<sup>29</sup> Because environmental contamination and disease transmission pathways could be different in densely populated contexts, similar studies in urban areas would complement this rural trial.

Additional outcome measures collected in this trial will help to elucidate potential mechanisms for the observed effects, including indicators of environmental contamination, environmental enteric dysfunction, anaemia, enteric parasite infection, and child development. Molecular measurement of infections in the laboratory with stored stool specimens collected as part of this trial offer an opportunity for unbiased indicators of pathogen burden.

More intensive promotion and higher adherence could have resulted in larger effects than those reported, but our findings are relevant for large-scale programmes that struggle to achieve adherence rates as high as those of efficacy studies. The potential for water, sanitation, hygiene, and nutrition interventions to reduce diarrhoea and improve growth might be highly context-dependent. In our rural setting, water was plentiful but rarely available on premises, susceptible to contamination at the source and in storage, and rarely treated despite introduction of a nearly-universal filter distribution programme;<sup>30</sup> unimproved latrine coverage was high and there was a culture of using sanitation facilities for defecation by human beings, but there was probably persistent exposure to animal faeces; handwashing was not a common practice; breastfeeding was common, but exclusive breastfeeding was not, and most people had enough food, but not a diverse diet; diarrhoea prevalence was high; and many children had low length-for-age Z score, but not weight-for-length Z score. Our findings call into question the ability of large-scale water, sanitation, and handwashing interventions to reduce diarrhoea or improve growth. Our results suggest that integrated water, sanitation, and handwashing and nutrition programmes are no more effective than nutrition programmes at reducing diarrhoea or improving growth, and that nutritional interventions that include counselling and LNS can modestly reduce growth faltering, but fall short of eliminating it, even when LNS adherence is high.

#### Contributors

CN and CPS contributed equally to the manuscript. CN drafted the research protocol and manuscript with input from all listed coauthors and oversaw all aspects of the trial. CPS led the nutrition intervention and protocols for anthropometry data collection. KGD contributed to the nutrition intervention and interpretation of results. PK assisted with oversight of anthropometry data collection. CN, AJP, HND, TC, and PKR developed the water, sanitation, and handwashing intervention. CN, CPS, AJP, HND, and GN oversaw piloting and subsequent study implementation, contributed to refinements in interventions and measurements, and responded to threats to validity. CN, CPS, BFA, CDA, JB-C, AEH, AM, and JMC Jr developed the analytical approach, did the statistical analysis, and constructed the tables and figures. AL, SPL, and JMC Jr advised on harmonising the trials between Kenya

and Bangladesh and SMN helped adapt the trial to the Kenyan context. CN, CPS, AJP, BFA, JB-C, LCHF, AL, SPL, SMN, and JMC Jr secured funding for the trial. All authors have read, contributed to, and approved the final version of the manuscript.

#### Declaration of interests

All authors received funding for either salary or consulting fees through a grant from the Bill & Melinda Gates Foundation for this study.

#### Acknowledgments

We thank the study participants and promoters who participated in the trial, the fieldworkers who delivered the interventions and collected the data for the study, and the managers who ensured that everything ran smoothly. This research was financially supported in part by Global Development grant OPPCD759 from the Bill & Melinda Gates Foundation to the University of California, Berkeley, CA, USA, and grant AID-OAA-F-13-00040 from United States Agency for International Development (USAID) to Innovations for Poverty Action. This manuscript was made possible by the generous support of the American people through the USAID. The contents are the responsibility of the authors and do not necessarily reflect the views of USAID or the US Government.

#### References

- 1 United National Children's Fund, WHO, World Bank Group. Levels and trends in child malnutrition. 2016. [http://www.who.int/nutgrowthdb/jme\\_brochure2016.pdf?ua=1](http://www.who.int/nutgrowthdb/jme_brochure2016.pdf?ua=1) (accessed March 18, 2017).
- 2 Sudfeld CR, McCoy DC, Danaei G, et al. Linear growth and child development in low- and middle-income countries: a meta-analysis. *Pediatrics* 2015; **135**: e1266–75.
- 3 Prendergast AJ, Humphrey JH. The stunting syndrome in developing countries. *Paediatr Int Child Health* 2014; **34**: 250–65.
- 4 Humphrey JH. Child undernutrition, tropical enteropathy, toilets, and handwashing. *Lancet* 2009; **374**: 1032–35.
- 5 Black RE, Victora CG, Walker SP, et al. Maternal and child undernutrition and overweight in low-income and middle-income countries. *Lancet* 2013; **382**: 427–51.
- 6 Scrimshaw NS, Taylor CE, Gordon JE. Interactions of nutrition and infection. *Monogr Ser World Health Organ* 1968; **57**: 3–329.
- 7 GBD 2015 Mortality and Causes of Death Collaborators. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* 2016; **388**: 1459–544.
- 8 Richard SA, Black RE, Gilman RH, et al. Diarrhoea in early childhood: short-term association with weight and long-term association with length. *Am J Epidemiol* 2013; **178**: 1129–38.
- 9 Luby SP, Rahman M, Arnold BF, et al. Effects of water quality, sanitation, handwashing and nutritional interventions on diarrhoea and child growth in rural Bangladesh: a cluster randomised trial. *Lancet Glob Health* 2018; published online Jan 29. [http://dx.doi.org/10.1016/S2214-109X\(17\)30490-4](http://dx.doi.org/10.1016/S2214-109X(17)30490-4).
- 10 Arnold BF, Null C, Luby SP, et al. Cluster-randomised controlled trials of individual and combined water, sanitation, hygiene and nutritional interventions in rural Bangladesh and Kenya: the WASH Benefits study design and rationale. *BMJ Open* 2013; **3**: e003476.
- 11 Christensen G, Dentz HN, Pickering AJ, et al. Pilot cluster randomized controlled trials to evaluate adoption of water, sanitation, and hygiene interventions and their combination in rural western Kenya. *Am J Trop Med Hyg* 2015; **92**: 437–47.
- 12 Balzer L, van der Laan M, Petersen M. Adaptive pre-specification in randomized trials with and without pair-matching. *Stat Med* 2016; **35**: 4528–45.
- 13 Wolf J, Prüss-Ustün A, Cumming O, et al. Assessing the impact of drinking water and sanitation on diarrhoeal disease in low- and middle-income settings: systematic review and meta-regression. *Trop Med Int Health* 2014; **19**: 928–42.
- 14 Clasen TF, Alexander KT, Sinclair D, et al. Interventions to improve water quality for preventing diarrhoea. *Cochrane Database Syst Rev* 2015; **10**: CD004794.
- 15 Ejemot-Nwadiaro RI, Ehiri JE, Arikpo D, Meremikwu MM, Critchley JA. Hand washing promotion for preventing diarrhoea. *Cochrane Database Syst Rev* 2015; **9**: CD004265.
- 16 Pickering AJ, Djebbari H, Lopez C, Coulibaly M, Alzuza ML. Effect of a community-led sanitation intervention on child diarrhoea and child growth in rural Mali: a cluster-randomised controlled trial. *Lancet Glob Health* 2015; **3**: e701–11.
- 17 Clasen T, Boisson S, Routray P, et al. Effectiveness of a rural sanitation programme on diarrhoea, soil-transmitted helminth infection, and child malnutrition in Odisha, India: a cluster-randomised trial. *Lancet Glob Health* 2014; **2**: e645–53.
- 18 Patil SR, Arnold BF, Salvatore AL, et al. The effect of India's total sanitation campaign on defecation behaviors and child health in rural Madhya Pradesh: a cluster randomized controlled trial. *PLoS Med* 2014; **11**: e1001709.
- 19 Kenya National Bureau of Statistics, Ministry of Health, National AIDS Control Council, Kenya Medical Research Institute, National Council for Population and Development, ICF International. Kenya Demographic and Health Survey 2014. Rockville, MD, USA: ICF International, 2015.
- 20 Dangour AD, Watson L, Cumming O, et al. Interventions to improve water quality and supply, sanitation and hygiene practices, and their effects on the nutritional status of children. *Cochrane Database Syst Rev* 2013; **8**: CD009382.
- 21 Cameron L, Shah M, Olivia S. Impact evaluation of a large-scale rural sanitation project in Indonesia. World Bank Policy Research Working Paper 6360. Washington, DC: World Bank, 2013.
- 22 Hammer J, Spears D. Village sanitation and children's human capital: evidence from a randomized experiment by the Maharashtra government. World Bank Policy Research Working Paper 6580. Washington, DC: World Bank, 2013.
- 23 Maleta KM, Phuka J, Alho L, et al. Provision of 10–40 g/d lipid-based nutrient supplements from 6 to 18 months of age does not prevent linear growth faltering in Malawi. *J Nutr* 2015; **145**: 1909–15.
- 24 Iannotti LL, Dulience SJ, Green J, et al. Linear growth increased in young children in an urban slum of Haiti: a randomized controlled trial of a lipid-based nutrient supplement. *Am J Clin Nutr* 2014; **99**: 198–208.
- 25 Christian P, Shaikh S, Shamim AA, et al. Effect of fortified complementary food supplementation on child growth in rural Bangladesh: a cluster-randomized trial. *Int J Epidemiol* 2015; **44**: 1862–76.
- 26 Hess SY, Abbedou S, Jimenez EY, et al. Small-quantity lipid-based nutrient supplements, regardless of their zinc content, increase growth and reduce the prevalence of stunting and wasting in young Burkinafaso children: a cluster-randomized trial. *PLoS One* 2015; **10**: e0122242.
- 27 Harris A, Pickering AJ, Harris M, et al. Ruminants contribute fecal contamination to the urban household environment in Dhaka, Bangladesh. *Environ Sci Technol* 2016; **50**: 4642–49.
- 28 Kotloff KL, Nataro JP, Blackwelder WC, et al. Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* 2013; **382**: 209–22.
- 29 Ngure F, Reid BM, Humphrey JH, Mbuya MN, Peletto G, Stoltzfus RJ. Water, sanitation, and hygiene (WASH), environmental enteropathy, nutrition, and early child development: making the links. *Ann NY Acad Sci* 2014; **1308**: 118–28.
- 30 Pickering AJ, Arnold BF, Dentz HN, Colford JM, Null C. Climate and health co-benefits in low-income countries: a case study of carbon-financed water filters in Kenya and a call for independent monitoring. *Environ Health Perspect* 2017; **125**: 278–83.

For a list of managers see  
<http://washbenefits.net>

## Risk Factors for Menstrual Toxic Shock Syndrome: Results of a Multistate Case-Control Study

Arthur L. Reingold,\* Claire V. Broome,  
Suzanne Gaventa, Allen W. Hightower, and  
the Toxic Shock Syndrome Study Group†

From the Meningitis and Special Pathogens Branch and the Statistical Services Activity, Division of Bacterial Diseases, Center for Infectious Diseases, Centers for Disease Control, Atlanta, Georgia

For assessment of current risk factors for developing toxic shock syndrome (TSS) during menstruation, a case-control study was performed. Cases with onset between 1 January 1986 and 30 June 1987 were ascertained in six study areas with active surveillance for TSS. Age-matched controls were selected from among each patient's friends and women with the same telephone exchange. Of 118 eligible patients, 108 were enrolled, as were 185 "friend controls" and 187 telephone exchange-matched controls. Tampon use was a risk factor for developing TSS during menstruation (odds ratio = 29; 95% confidence interval = 7-120), and risk increased with increasing tampon absorbency (odds ratio = 1.34 per gram increase in absorbency; 95% confidence interval = 1.2-1.6). The role of tampon chemical composition could not be assessed because the number of cases was inadequate. Neither use of birth control pills for contraception nor use of medications for premenstrual or menstrual symptoms protected against or was a risk factor for the development of menstrual TSS.

Case-control studies conducted in the early 1980s demonstrated that tampon use was the major risk factor for the development of toxic shock syndrome (TSS) during menstruation and that risk varied with the brand and style of tampon used [1-6]. One of these studies further demonstrated that a tampon's absorbency and/or chemical composition was important in determining the risk associated with its use, although the relative importance of these two tampon characteristics remained uncertain [3]. Subsequent in vitro studies have suggested that the chemical composition of tampons may be the major de-

terminant of risk because of differences in the binding of magnesium and hence in the production of TSS toxin 1 [7-9]. However, a recent assessment of cases reported through a passive national-surveillance system suggests that both absorbency and chemical composition are important independent determinants of the risk of menstrual TSS [10].

In response to these findings and in an effort to minimize or eliminate the risk of menstrual TSS, manufacturers have both substantially altered the chemical composition and dramatically lowered the absorbency of the tampons they sell. As a result, the tampons that are available and being used today differ markedly from those in use in the early 1980s. In order to evaluate the risk of menstrual TSS associated with currently available tampons and to shed more light on the relative importance of tampon absorbency and chemical composition in determining that risk, we undertook a case-control study of menstrual TSS cases occurring in 1986-1987.

This study was supported by an interagency agreement of the Centers for Disease Control, the National Institute of Child Health and Human Development, and the U.S. Food and Drug Administration.

\* Present address: Department of Biomedical and Environmental Health Sciences, School of Public Health, University of California at Berkeley, Berkeley, California 94720.

† The Toxic Shock Syndrome Study Group includes S. Waterman and C. Hoppe (Los Angeles County); M. Spurrier and S. Sizte (Missouri); R. McCready, D. Cundiff, and M. Farrell (New Jersey); G. Istre and S. Makintubee (Oklahoma); L. Lefkowitz and J. Taylor (Tennessee); W. Lafferty and J. Harwell (Washington); Drs. M. Donawa and C. Gaffey (U.S. Food and Drug Administration); and Drs. J. Perlman and P. Wolf (National Institute of Child Health and Human Development).

Please address requests for reprints to the Meningitis and Special Pathogens Branch, Division of Bacterial Diseases, Center for Infectious Diseases, Centers for Disease Control, Atlanta, Georgia 30333.

### Methods

Patients with TSS and age-matched controls were sought in six study areas (Los Angeles County and the states of Missouri, New Jersey, Oklahoma, Tennessee and Washington) where active surveillance for TSS had been established. Details of the active surveillance methods used are presented elsewhere [11]. In brief, educational materials concerning TSS and

a request for reports of all suspected cases were distributed repeatedly to health care providers, infection control nurses, and medical records departments in the study areas. These materials stressed that TSS occurs in a variety of settings in patients of both sexes and all ages. Active surveillance for patients hospitalized with TSS was maintained by biweekly telephone calls to all hospitals in the study areas to ascertain the presence or absence of suspected cases.

All suspected cases in women 10–54 years of age with onset between 1 January 1986 and 30 June 1987 were assessed with regard to the case definition for TSS established by the Centers for Disease Control [12]. Cases meeting all of the criteria were considered definite cases, those lacking a single criterion were considered probable cases, and those lacking two or more criteria or having evidence of another cause of illness were considered not to be cases. All medical records were reviewed a second time by an individual blinded to the menstrual status and tampon use history of the patient. The few minor discrepancies in classification of cases were resolved by a second person blinded to menstrual status and tampon use history. Probable and definite cases with onset of symptoms during menstruation (i.e., during active bleeding) were eligible for inclusion in the study unless a focal site of infection outside the vagina was identified or a barrier contraceptive was used during the menstrual period.

For each patient who agreed to participate, two friends matched for age ( $\pm 3$  years if  $<25$  years of age;  $\pm 5$  years if  $\geq 25$  years of age) and two women matched for age and neighborhood of residence were sought as controls. Controls matched for neighborhood of residence were sought by taking the first five digits of the patient's phone number and randomly ordering the 99 other possible phone numbers with the same first five digits. These households matched by telephone exchange (and hence by neighborhood of residence) were called until two age-matched women were enrolled. Women with TSS and controls were interviewed by telephone concerning use of tampons and other catamenial products on each day of the menstrual period, use of medications for menstrual and premenstrual symptoms on each day for the 3 days before onset of menstruation and during menstruation, and use of contraceptives. Patients with TSS were asked about the menstrual period when they became ill (index menstrual period) and the preceding menstrual period; controls were asked about the two menstrual periods that coincided in

time with those of the respective case. While the interviewer was aware of the study hypotheses, she was blinded to the case/control status of participants at the time of the interviews. Tampon-using study participants were asked to find the box of tampons used during the most recent menstrual period and answer questions about its labeling and color.

Results were analyzed with conditional multivariate logistic regression models that took the matching into account [13]. Information concerning the chemical composition, oxygen content, and *in vivo* and *in vitro* absorbency of various tampon brands and styles was obtained from tampon manufacturers.

## Results

Altogether, 118 patients with TSS were eligible for enrollment in the study, and 108 of these patients were enrolled. Reasons for which patients were not enrolled included refusal (two patients) and loss to follow-up or inability to locate (eight patients). None of the 118 patients died. Of the 108 patients enrolled, 71 were classified as having definite and 37 as having probable TSS. Among the 37 probable cases, fever of  $\geq 102^{\circ}\text{F}$  was the criterion most often lacking (15 cases); desquamation was lacking in 14 cases, multisystem involvement in four, and hypotension in four. The characteristic rash of TSS was present in all probable cases. Onset of illness occurred most often on the third or fourth day of the menstrual cycle (day 1, 9%; day 2, 14%; day 3, 17%; day 4, 29%; day 5, 12%; day 6, 13%, day 7, 2%; and day 8, 4%).

Altogether, 372 age-matched controls were enrolled, including 185 friends of patients and 187 neighborhood residents. Four controls were enrolled for each of 71 cases (66%), three controls for each of 15 cases (14%), two controls for each of 21 cases (19%), and only one control for one case (1%). As expected, the patients and controls were similar in age, race, and marital status (table 1). "Friend controls" were somewhat more similar to patients than were "neighborhood controls" with regard to race and marital status, but these differences were not significant.

Of the 108 women with TSS, 106 (98%) were using tampons at the time of onset of illness; 88 women had been using a single brand and style of tampon during that menstrual period, whereas 18 had been using multiple brands and/or styles (table 2). Of the 372 control women, 244 (66%) had used tampons

**Table 1.** Characteristics of patients and controls enrolled in a multistate study of risk factors for menstrual toxic shock syndrome.

Characteristic (unit)	Value for indicated group			
	Patients	Friend controls	Neighborhood controls	Combined controls
Mean age (y)*	24.3 ± 8.1 (13–46)	24.8 ± 8.4 (11–48)	24.5 ± 8.1 (13–48)	24.6 ± 8.2 (11–48)
White (%)	94	94	89	91
Married (%)	44	39	36	37
Interval from onset of index menstrual period to interview (d)*	88 ± 50 (25–249)	...	...	87 ± 51 (17–281)
Interviews successfully completed with blinding to case/control status (%)	82	91	87	89

\* Values given are mean ± SD (range).

during their index menstrual period. Friend controls were more likely to have used tampons than were neighborhood controls (71% vs. 60%; odds ratio = 1.7; 95% confidence interval = 1.02–2.7; two-tailed  $P = .04$ , conditional logistic regression). Altogether, 44% of tampon-using patients and 62% of tampon-using controls were able during their telephone interview to find the box of tampons used.

Tampon use was associated with an increased risk of developing TSS during menstruation, regardless of which control group was used as a basis for comparison (friends, neighbors, or combined; table 3). Women who used multiple brands and/or styles were at greater risk than women who used a single brand and style (odds ratio = 2.3; 95% confidence interval = 1.2–4.6;  $P = .02$ ). However, this difference was due to the fact that users of multiple brands and/or styles tended to use more absorbent tampons. With control for absorbency, there was no difference

in risk between users of a single brand and users of multiple brands and/or styles.

Because there were overall no significant differences between friend and neighborhood controls regarding the brand or style of tampon used, these control groups were combined in studies of the risk of menstrual TSS associated with individual brands and individual brand/style combinations. The use of all major tampon brands was associated with an increased risk of developing TSS during menstruation, with odds ratios for individual brands ranging from 15 to 59 (table 4). Odds ratios for individual styles of each tampon brand were calculated in two ways; in comparison with the risk of TSS in women not using tampons and in comparison with the risk of TSS in users of Tampax Original Regular tampons. In comparison with women using no tampons, users of all assessed individual brands and styles (except Tampax Slender Regular and Tampax Original Regu-

**Table 2.** Tampon use during the index menstrual period.

Pattern of tampon use	Patients	No. (%) in indicated group with pattern of use		
		Friend controls	Neighborhood controls	Combined controls
None	2 (2)	54 (29)	74 (40)	128 (34)
Single brand and style	88 (81)	115 (63)	104 (56)	219 (59)
Multiple brands and/or styles	18 (17) {	15 (8) {	7 (4) {	22 (6) {
Unknown brand	... {	1 (<1) {	2 (1) {	3 (1) {
Total	108	185	187	372

\* Significant difference between friend and neighborhood controls (odds ratio = 1.7; 95% confidence interval = 1.02–2.7; two-tailed =  $P = .04$ ).

**Table 3.** Association between tampon use and risk of menstrual toxic shock syndrome.

Tampon use	Odds ratio*/95% confidence interval for patients vs. indicated control group		
	Friend	Neighborhood	Combined
Any tampon	19/5-78	48/7-362	29/7-120
Single brand and style	...	...	27/7-111
Multiple brand and/or style	...	...	62/13-291

\* Vs. no tampon use.

lar) were at increased risk of menstrual TSS (table 5). In comparison with users of Tampax Original Regular tampons, users of some but not all other brand/style combinations were demonstrated to be at increased risk.

We next analyzed risk of menstrual TSS as a function of various tampon characteristics, including measured in vitro and in vivo absorbency, weight, oxygen content, and chemical composition. There was a significant association between measured in vitro tampon absorbency and risk of menstrual TSS: the risk increased by 34% for every 1-g increase in absorbency (odds ratio per gram increase = 1.34; 95% confidence interval = 1.2-1.6). Tampon weight and in vivo absorbency were equally good predictors of the risk of menstrual TSS, while oxygen content correlated somewhat less well. After taking in vitro absorbency into account, we could detect no influence of oxygen content or of chemical composition (categorized either as the presence or absence of a given material or as the percentage comparison by weight) on the risk of menstrual TSS.

Analysis of tampon users revealed that patterns of tampon use differed between patients and controls (table 6). Tampon-using women with TSS used tampons on more days of the menstrual cycle, were more likely to use tampons continuously for at least 1 day, used tampons continuously on more days and on a higher percentage of days of the menstrual cycle, and left a single tampon in place for a longer mean maximum time. Patients and controls were similar, however, in the average number of tampons used per day and the total number of tampons used per menstrual period. Because many of these characteristics of tampon use were correlated with the absorbency of the tampon used, we also examined their effect on the risk of menstrual TSS after adjustment for absorbency. Using tampons continuously on at least 1 day of the menstrual cycle remained strongly correlated with the risk of menstrual TSS after adjustment for absorbency (odds ratio = 6.5; 95% confidence interval = 2.5-17.2). Once absorbency and continuous use of tampons were taken into account, none of the other tampon-use variables remained significantly associated with risk of menstrual TSS.

Neither increased nor decreased risk of menstrual TSS in association with the use of birth control pills or barrier contraception was found (table 7). Use of condoms for contraception was commoner, however, among women with TSS (odds ratio = 2.6; 95% confidence interval = 1.1-6.1). The use of medications for premenstrual and menstrual syndromes was not associated with either an increased or a decreased risk of developing TSS, whether examined by individual brand, by active ingredient, or by overall use/nonuse (table 8).

**Table 4.** Association between tampon brand and risk of menstrual toxic shock syndrome.

Tampon brand*	No. using brand in indicated group		Matched odds ratio	95% confidence interval
	Patients	Combined controls		
None	2	128	1	...
Tampax	23	128	15	3-64
OB	9	15	56	9-330
Playtex	46	63	59	13-265
Kotex	10	12	54	10-302
Other	0	1	0	...
Total	90	347		

\* Single brand and style use only.

**Table 5.** Risk of menstrual toxic shock syndrome among users of selected individual tampon brands and styles.

Brand and style of tampon	No. (%) using brand/style in indicated group		Odds ratio/95% confidence interval vs. indicated category	Use of Tampax Original Regular
	Patients	Controls		
No tampon	...	...	1/...	...
Tampax Original Regular	2 (2)	39 (18)	7/0.8-58	1/...
Tampax Slender Regular	4 (5)	27 (13)	6/1-35	0.98/0.1-8
Tampax Petal Soft Regular	2 (2)	11 (5)	22/2-212	3.2/0.4-30
Tampax Super	9 (11)	38 (18)	26/4-149	3.7/0.6-22
Tampax Super Plus	3 (4)	13 (6)	25/3-207	3.8/0.5-30
OB Regular	3 (4)	9 (4)	28/3-268	4.2/0.5-38
OB Super	4 (5)	5 (2)	86/9-862	13/1.4-122
OB Super Plus	2 (2)	1 (<1)	144/7-2,857	22/1.1-422
Playtex Slender Regular (D/ND)*	4 (5)	5 (2)	78/8-789	11/1.2-110
Playtex Regular (D/ND)	20 (24)	27 (13)	76/13-441	13/2.4-66
Playtex Super (D/ND)	16 (19)	25 (12)	74/13-429	11/2-58
Playtex Super Plus (D/ND)	6 (7)	6 (3)	79/10-612	12/1.6-83
Kotex Security Regular	2 (2)	6 (3)	21/1.7-253	2.9/0.2-40
Kotex Security Super	7 (8)	4 (2)	122/15-971	18/2.5-133

\* Deodorant and nondeodorant, combined.

## Discussion

The results presented here suggest that, despite marked changes in the absorbency and chemical composition of tampons in recent years, the use of many if not all tampons available in 1986-1987 is associated with an increased risk of menstrual TSS. Furthermore, while the measured absorbency of tampons has been reduced dramatically, there continues

to be a direct correlation between measured tampon absorbency and risk of menstrual TSS. Continuous use of tampons on at least 1 day of the menstrual cycle appears to increase a tampon user's risk of developing TSS, as has been noted previously [5]. We were unable to confirm the results of earlier studies that suggested a protective effect of oral contraceptive pills with regard to menstrual TSS [14].

**Table 6.** Univariate analyses of patterns of tampon use among toxic shock syndrome patients and controls who used tampons.

Variable	Mean $\pm$ SD for indicated group			95% confidence interval
	Patients (n = 106)	Controls (n = 244)	Odds ratio	
Mean average no. of tampons used per day	4.7 $\pm$ 4.1	4.3 $\pm$ 2.3	1.04/tampon	0.97-1.13
Mean total no. of tampons used per menstrual period	21.9 $\pm$ 21.6	18.3 $\pm$ 12.2	1.02/tampon	1.0-1.03
Mean no. of days on which tampons were used	4.5 $\pm$ 1.6	4.2 $\pm$ 1.5	1.22/day of use	1.03-1.44
Mean no. of days on which tampons were used continuously	4.0 $\pm$ 2.1	2.3 $\pm$ 2.3	1.46/day of continuous use	1.27-1.67
Mean percentage of days on which tampons were used continuously	83.8 $\pm$ 8	52.9 $\pm$ 47	1.02/percentage of days	1.01-1.03
Mean maximum time a single tampon was left in place (hours)	7.8 $\pm$ 2.1	6.6 $\pm$ 2.4	1.46/hour	1.21-1.75
Any day(s) of continuous tampon use	95 (90)*	141 (58)*	9.4	3.9-22.3

\* Values indicate number (percentage) of women.

**Table 7.** Use of contraceptives and risk of toxic shock syndrome.

Type of contraception	No. (%) using method in indicated group		Matched odds ratio	95% confidence interval
	Patients (n = 108)	Controls (n = 372)		
Condoms	10 (9)	15 (4)	2.6	1.1-6.1
Birth control pills	27 (25)	89 (24)	1.1	0.6-1.8
Any barrier contraception*	3 (3)	19 (5)	0.6	0.2-2.1
Diaphragm*	2 (2)	16 (4)	0.5	0.1-2.1
Contraceptive sponge*	1 (1)	2 (<1)	...	...
Any spermicide	6 (6)	22 (6)	...	...
Intrauterine device	2 (2)	7 (2)	...	...
Tubal ligation	6 (6)	31 (8)	...	...
Hysterectomy	1 (1)	1 (<1)	...	...
Rhythm	2 (2)	0	...	...
Withdrawal	2 (2)	1 (<1)	...	...
Cervical cap*	0	1 (<1)	...	...

\* All cases of menstrual and nonmenstrual toxic shock syndrome associated with the use of a diaphragm, contraceptive sponge, or cervical cap were excluded from this study.

The magnitude of the risk associated with tampon use in our study remains somewhat ill defined because of the different frequencies of tampon use observed among the two types of controls enrolled. Thus, depending on whether friend or neighborhood controls were used as the standard for comparison, the estimate of the risk varied between 19 and 48. While combining of the two control groups for this particular comparison is not valid because of their heterogeneity, it is likely that the resultant estimate of the frequency of tampon use among control women (66%) would yield a more accurate estimate of the risk associated with tampon use (odds ratio = 29) than does an analysis of either control group

alone. Data from national surveys conducted in 1985 suggest that ~65% of women with menstrual periods use tampons [10].

Two limitations to this study warrant discussion in an assessment of the results. First, it is possible that, despite all of our educational efforts and publicity, medical care providers were more likely to diagnose and/or report a case of menstrual TSS if the patient was a tampon user. Bias of this type would have resulted in overestimation of the risk associated with tampon use vs. no tampon use. We currently are reviewing ~12,000 medical records for all women 10–54 years of age who were discharged from hospitals in the study areas in 1986 with TSS or diagnoses likely to be confused with TSS in an effort to determine how many of these women had TSS that was undiagnosed and/or unreported. By ascertaining the menstrual status and pattern of tampon use for women with TSS that was unreported and/or misdiagnosed, we hope to assess the impact of diagnostic and reporting biases on our results. It should be noted, however, that these biases would not have affected our analysis of the risk associated with use of individual brands and styles of tampons vs. use of Tampax Original Regular tampons. Similarly, these biases would not have affected our analysis of the relation between measured tampon absorbency or tampon use patterns and risk of menstrual TSS.

The second limitation is the paucity of cases available for study. Because of the small number of cases studied, the confidence intervals around our point estimates are very wide; that is, our estimates of var-

**Table 8.** Use of medications for premenstrual and menstrual symptoms and risk of toxic shock syndrome.

Medication	No. (%) taking medication in indicated group		95% confidence interval	
	Patients (n = 108)	Controls (n = 372)	Odds ratio	confidence interval
Any	40 (37)	138 (37)	1.0	0.7-1.6
Midol	4 (4)	18 (5)	0.7	0.2-2.2
Aspirin	5 (5)	22 (6)	0.8	0.3-2.3
Tylenol	10 (9)	32 (9)	1.1	0.5-2.4
Motrin	3 (3)	14 (4)	0.7	0.2-2.6
Advil	7 (6)	13 (3)	2.1	0.7-6.1
Nuprin	0 (0)	8 (2)	...	...
Pamprin	4 (4)	12 (3)	1.1	0.3-3.6
Premesyn	3 (3)	2 (1)	5.0	0.8-30
Other	10 (9)	31 (8)	...	...

ious risks are imprecise. Furthermore, despite our efforts, there are insufficient cases to permit a meaningful assessment of the independent contributions of tampon absorbency, chemical composition, and other characteristics to the risk of menstrual TSS. Thus, it remains possible that one or more tampon characteristics other than measured *in vitro* absorbency could play an important role in determining the risk of menstrual TSS. Given the enormous effort and the size of the surveillance population required for the collection of the cases studied here, it seems unlikely that a prospective study that is based on active surveillance and is large enough to answer questions about the impact of tampon characteristics will be feasible.

While the observed incidence of nonmenstrual TSS in the study areas was approximately that predicted on the basis of findings from earlier studies, the incidence of menstrual TSS was substantially lower than that predicted from data gathered in other states during previous years [11]. Thus, while incidence rates in the range of 5–15 cases/100,000 menstruating women per year were observed in Wisconsin, Minnesota, Utah, and Colorado in 1980, the incidence rate of menstrual TSS observed in our six study areas in 1986 ranged between 1 and 2.5/100,000 menstruating women. Whether the incidence of menstrual TSS we observed was lower than expected because the incidence has dropped in recent years, because the areas under study always had lower incidences, because cases now are being recognized and treated earlier, or because other unknown factors are involved is unclear. However, even if the incidence of menstrual TSS has decreased in recent years, our data suggest that there is still a need for a uniform standard of tampon labeling with regard to measured absorbency.

#### References

- Davis JP, Chesney PJ, Wand PJ, LaVenture M, the Investigation and Laboratory Team. Toxic-shock syndrome: epidemiologic features, recurrence, risk factors, and prevention. *N Engl J Med* 1980;303:1429–35
- Helgerson SD, Foster LR. Toxic shock syndrome in Oregon: epidemiologic findings. *Ann Intern Med* 1982;96(Part 2):909–11
- Osterholm MT, Davis JP, Gibson RW, Mandel JS, Wintermeyer LA, Helms CM, Forfang JC, Rondeau J, Vergeront JM, and the Investigation Team. Tri-state toxic-shock syndrome study. I. Epidemiologic findings. *J Infect Dis* 1982;145:431–40
- Schlech WF III, Shands KN, Reingold AL, Dan BB, Schmid GP, Hargrett NT, Hightower A, Herwaldt LA, Neill MA, Band JD, Bennett JV. Risk factors for the development of toxic shock syndrome: association with a tampon brand. *JAMA* 1982;248:835–9
- Shands KN, Schmid GP, Dan BB, Blum D, Guidotti RI, Hargrett NT, Anderson RL, Hill DL, Broome CV, Band JD, Fraser DW. Toxic-shock syndrome in menstruating women: its association with tampon use and *Staphylococcus aureus* and the clinical features in 52 cases. *N Engl J Med* 1980;303:1436–42
- Kehrberg MW, Latham RH, Haslam BR, Hightower A, Tanner M, Jacobson JA, Barbour AG, Noble V, Smith CB. Risk factors for staphylococcal toxic-shock syndrome. *Am J Epidemiol* 1981;114:873–9
- Kass EH, Kendrick MI, Tsai Y-C, Parsonnet J. Interaction of magnesium ion, oxygen tension, and temperature in the production of toxic-shock-syndrome toxin-1 by *Staphylococcus aureus*. *J Infect Dis* 1987;155:812–5
- Mills JT, Parsonnet J, Kass EH. Production of toxic-shock-syndrome toxin-1: effect of magnesium ion [letter]. *J Infect Dis* 1986;153:993–4
- Mills JT, Parsonnet J, Tsai Y-C, Kendrick M, Hickman RK, Kass EH. Control of production of toxic-shock-syndrome toxin-1 (TSST-1) by magnesium ion. *J Infect Dis* 1985; 151:1158–61
- Berkley SF, Hightower AW, Broome CV, Reingold AL. The relationship of tampon characteristics to menstrual toxic shock syndrome. *JAMA* 1987;258:917–20
- Gaventa S, Reingold AL, Hightower AW, Broome CV, Schwartz B, Hoppe C, Harwell J, Lefkowitz LK, Mackintubee S, Cundiff D, Sitze S, the Toxic Shock Syndrome Study Group. Active surveillance for toxic shock syndrome in the United States, 1986. *Rev Infect Dis* 1989;11(Suppl 1):S28–34
- Reingold AL, Hargrett NT, Shands KN, Dan BB, Schmid GP, Strickland BY, Broome CV. Toxic shock syndrome surveillance in the United States, 1980 to 1981. *Ann Intern Med* 1982;92:875–80
- Breslow NE, Day NE. Statistical methods in cancer research. Lyon: International Agency for Research on Cancer, 1980
- Shelton JD, Higgins JE. Contraception and toxic-shock syndrome: a reanalysis. *Contraception* 1981;24(6):631–4

#### Discussion

**DR. EDWARD KASS.** Dr. Reingold, I find it difficult to match your second conclusion with your data. The only data that show a clear relation are those dealing with polyacrylate rayon. All of the rest are not statistically significant. Now, the same thing was true in the Tri-State Study. I do not understand how you can say there is a linear relation between risk and absorbency if all of the excess statistically significant cases occur in relation to only one fiber. This is particularly important because, as you know, there is a question of national policy. There is a question of labeling absorbency. Representations have been made to the U.S. Food and Drug Administration. I find it difficult to make national policy recommen-

dations based on data that seem to me not secure, and, by your own statement, the numbers other than those dealing with polyacrylate rayon are not secure.

**DR. ARTHUR REINGOLD.** This study was done in 1986–1987, and none of these tampons contained polyacrylate rayon. Polyacrylate rayon was removed from Playtex tampons in the spring of 1985. Therefore, we are not able to look at the risk associated with polyacrylate in these data. I am the first to admit that the numbers here are very sparse. The question of whether there is any increased risk associated with various brands and styles compared with no tampon use depends on how many cases of TSS in non-tampon-using women went undiagnosed. We hope to get at least some assessment of that through this enormous chart review. To the extent that there has been a lot of diagnostic bias and those cases have been missed, it is possible that the increased risk in comparison to non-tampon use is, in fact, erroneous. The real problem then comes in terms of comparing other tampons with the Tampax Original Regular in that we have few cases relative to what we would like to have. I am, in fact, somewhat pleased that we were able to find so few cases because it indicates to me that we have been going in the right direction in the last few years and that this disease has really decreased in incidence. On the other hand, it makes for difficulties in interpreting the results of the study.

**DR. JAMES TODD.** I hope your conclusion is correct. As you say, you will only know whether the incidence has decreased once you have ascertained your reporting bias and what effect it has on your statistics. Certainly, your data from California do not suggest that the incidence has decreased significantly in that area. To speculate a bit, let us assume that there is a direct risk associated with absorbency. It has been said that this risk is not a function of leaving tampons in longer, although from seeing cases clinically I am convinced that it is. My own experience suggests that the severity of illness seems to relate directly to how long the tampon was left in. What are the data to convince us that the increase in absorbency in tampons is not directly related to an increase in the length of time that the tampon is left in?

**DR. REINGOLD.** The data are not good. In this study we did look at the number of tampons used per day (as the best indicator we could come up with because we were interviewing between 1 and 2 months after the illness), and there is not a substantial differ-

ence between the patients and the controls, which is what has been found in similar case-control studies. As to the other point you raise, I do not understand the biologic way in which absorbency could affect risk. We have looked at the data, substituting oxygen content because there is some correlation between oxygen content and absorbency, and if anything, oxygen content is not as good a predictor of risk as absorbency. The weight of the tampon is as good an indicator as absorbency, but again, they are too closely correlated to be separable. I do not know what it is that measured absorbency is telling us or what it indicates.

**DR. KASS.** The most convincing data came from the Tri-State Study, which reported that if there was any kind of cross-over between length of time a tampon is worn and risk, it was at ~13 hours, and the effect was negligible. From that fairly large study, it did not appear that length of time was a great variable in rate of disease. Whether that has changed since then, I do not know. We have all seen cases of the kind that Dr. Todd mentioned, but I think that the length of time a tampon is kept in place has not been statistically significant in relation to risk.

Second, with respect to the point about oxygen, as you know, we published a paper on the effect of oxygen on toxin production, and, except at conditions of zero oxygen, there is toxin production, particularly when magnesium levels are low. I agree that it is unlikely that variation in oxygen is going to be a major significant variable if some oxygen is present.

Third, I hope people will keep in mind that most cotton-containing tampons, whether all cotton or partially cotton, have adherent magnesium that is not covalently linked. Cotton itself has no free carboxyl groups. Therefore, any salts that are in the cotton tampon are simply there as contaminants during the manufacturing process. The salts leach out easily, and the salt content varies immensely from batch to batch. Cotton-containing tampons will usually release magnesium and therefore counteract any other tendency toward increased toxin production, and this becomes an important variable in looking at the effect of different products. Unless each product is carefully examined to see how much this particular variable changes from product to product—and I can assure you it changes immensely from batch to batch—you will get peculiar and variable results, and this adds to the underlying argument that we are talking of a surrogate and not of absorbency itself.