



PHW250B Week 5 Reader

Topic 1: Cohort Studies

Lecture: Cohort Studies in Depth.....	2
Grimes & Schulz. Cohort studies: marching towards outcomes. Lancet 2002; 359: 341–45.	24
Samet & Muñoz. Evolution of the Cohort Study. Epidemiologic Reviews 1998; 20(1)1.	29

Topic 2: Evaluating and Reporting Cohort Studies

Lecture: Evaluating and Reporting Cohort Studies.....	43
STROBE Checklist for Cohort Studies.	52

Podcast

Episode 1: Interview with Ben Arnold – Surfer Health Study.	54
--	----

Journal Club

Arnold et al. (2017).....	63
---------------------------	----

Lecture: Cohort Studies in Depth



Cohort studies in depth

PHW250 B - Andrew Mertens



In this video, we'll talk about methods for designing cohort studies in depth.

Quick review: cohort studies

- • The term “cohort” has military roots.
- • **Military:** A cohort was a 300–600-man unit in the Roman army.
- • **Epidemiology:** A cohort study consists of bands or groups of persons marching forward in time from an exposure to one or more outcomes.



Let's start with a quick review of what cohort studies are. The term cohort has military roots. It's a term that was used to describe 300 to 600 soldiers in the Roman army. In epidemiology, we use this term to describe a type of study that consists of bands or groups of people who we can think of as marching forward in time.

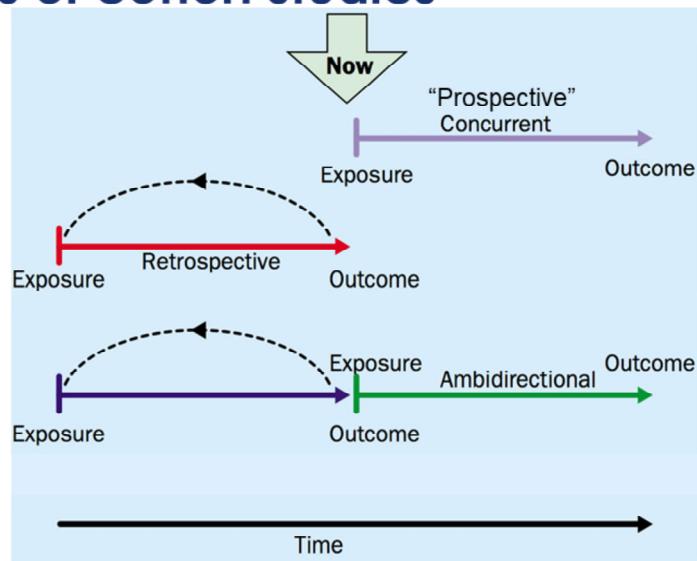
And as they are marching forward in time they're experiencing different exposures and potentially different outcomes as well.

Quick review: types of cohort studies

The defining characteristic of all cohort studies is that they track people forward in time from exposure to outcome.

A cohort study moves in one direction, although gathering data might not.

Exposure → outcome



Grimes et al. 2002

And to quickly review, here's a schematic of different types of cohort studies. The defining characteristic that is shared across all these types of cohort studies is that they track what people forward in time *from an exposure to an outcome. The cohort study itself is moving forward in one direction in time, but that doesn't mean that the data collection is moving forward in time.

If we look at the first type of cohort study at the top of this figure, it's a prospective or concurrent cohort study, meaning that at the beginning of the study marked by this now arrow, we measure the exposure status of people in the cohort and then we follow them forward in time and assess their outcomes status later on in time. The second type of cohort study in this figure is a retrospective cohort study, where we often know the outcome status of individuals in the study population at the beginning of the study, and then we move backwards in time to assess their exposure. And this is why it's called a retrospective study.

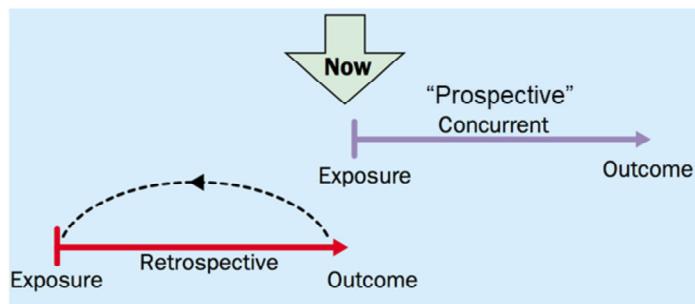
And then finally, we can combine these two types-- prospective and retrospective-- and conduct an ambidirectional cohort study, where some information is collected as we wait for outcomes to develop in the future, but we also assess exposures at the present time and in the past.

Quick review: types of cohort studies

Example: study of whether assistive reproductive technologies are associated with multiple births (e.g., twins)

Prospective: Track (a) women exposed to these technologies and (b) a similar group who conceived naturally. Monitor the frequency of multiple births.

Retrospective: Use existing medical records and go back in time several years to identify women exposed and not exposed to these technologies. Then track them forward through records to note the birth outcomes.



Berkeley School of Public Health Grimes et al. 2002

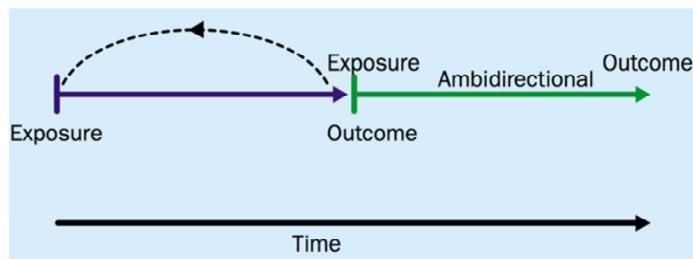
Let's go over an example. Let's say we're interested in studying whether assisted reproductive technologies such as in vitro fertilization, are associated with multiple births, such as having twins or triplets. In a prospective cohort study, we could track women who were exposed to these kinds of assisted reproductive technologies, and then we could also track women who conceived naturally without these technologies.

And then we would monitor the frequency of multiple versus single births in these two different groups and compare them. If we wanted to conduct a retrospective study of this question, we could use existing medical records to assess the status of births to see if they are multiple births or single births, and then we can go back in time several years to identify the women who did and did not use reproductive technologies. We would then retrace the outcomes of women in the cohort and associate those outcomes with the particular exposures that each woman experienced.

Quick review: types of cohort studies

Example: Are assisted reproductive technologies associated with multiple births and with ovarian cancer in later life.

Ambidirectional: Look back through records for women who had multiple births and also start to follow these women into the future for ovarian cancer occurrence.



In this example, we use an ambidirectional cohort design to answer a slightly different research question, are assistive reproductive technologies associated with multiple births and with ovarian cancer after the child is delivered or the children are delivered.

In this design, we could look back through records to see which women had multiple births or single births, and then we could also look into the future to see if women developed ovarian cancer. We would also need to track any reproductive technologies used to assist with conception.

Quick review: types of populations

- **Closed population:**
 - There are no losses to follow-up
 - No migration in to the population
 - No competing risks
 - No change in exposure status over time
- **Open population:**
 - Migration in and out of the population may occur
 - Loss to follow-up may occur
 - Competing risks may occur
 - Exposure status may change over time

Before we go on, let's also quickly review two types of populations that will affect the appropriate measures of disease that we can estimate in different types of cohort studies. First is a closed population, and this is a simpler population for us to think about as epidemiologists. In this type of population, there is no losses to follow-up. This means every who we start tracking at the beginning of our cohort study participates in the study the whole time, and we have information about them throughout the entire follow up period.

Nobody moves out of the population. Nobody moves into the population, and there are no competing risks, which means if our outcome of interest is ovarian cancer, for example, there is no other competing outcomes like other types of cancer, or heart disease, or any other health outcome that could affect our population, doesn't exist during our follow up period. There's also no change in exposure status over time. An open population, on the other hand, allows for migration in and out of the population to occur. Loss to follow-up and competing risks may occur, and exposure status may change over time.

In most cases in epidemiology, especially if we're not dealing with a very short follow up period, we need to make an assumption that the population is an open rather than a closed population.

What measures of disease can be estimated in a cohort study?

- If there is **no loss to follow-up** and **no competing risks** (i.e. in a **closed population**) we can directly estimate
 - Cumulative incidence
 - Incidence density
 - Occurrence times
 - Prevalence

Now let's think about what measures of disease can be estimated in cohort studies, depending on the type of population that we have. If we're talking about a closed population with no loss to follow-up and no competing risks, then we can directly estimate cumulative incidence, incidence density, occurrence times, and prevalence. What do I mean by occurrence times?

Well, often we're interested in the average age of onset of disease or average time of onset since the beginning of follow-up. So I'll get back to what we mean by directly in just a moment.

What measures of disease can be estimated in a cohort study?

- If there is **no loss to follow-up** and **no competing risks** (i.e. in a **closed population**) we can directly estimate
 - Cumulative incidence
 - Incidence density
 - Occurrence times
 - Prevalence
- If there is **loss to follow-up** and/or **competing risks** (i.e., in an **open population**) we do not know the final outcome status of some individuals in the study population, so:
 - Cumulative incidence: must use the Actuarial, Kaplan Meier, or Density methods
 - Incidence density: can still be calculated directly
 - Occurrence times: cannot be estimated directly
 - Prevalence: cannot be estimated directly
- This is the case in nearly all epidemiologic studies with the exception of efficacy trials with short follow-up periods.

Berkeley School of Public Health
Rothman ME3, Chapter 7

If there is loss to follow-up and or competing risks, we're dealing with an open population. And as a result, we don't know the final outcome status of some people in the study population. To estimate cumulative incidence, we need to use indirect methods, such as the actuarial, Kaplan Meier, or density methods that you learned about earlier in this course.

We can still calculate incidence density directly, but we can't directly calculate occurrence times or prevalence without making some assumptions. This is really the case in nearly all epidemiologic studies with the exception of some efficacy trials with short follow-up periods. It's good to become really familiar with the methods that we need to use in this scenario for this type of population.

Example of direct cumulative incidence calculation

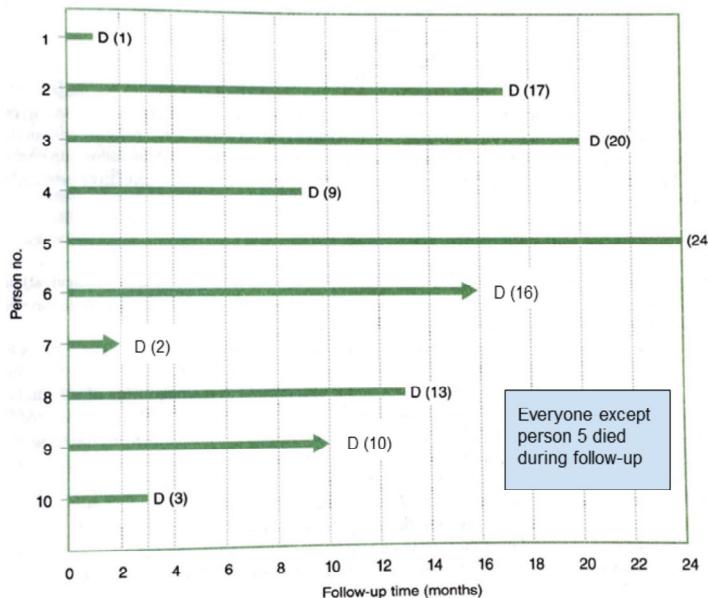
- No loss to follow-up or competing risks - we know the time of death for everyone in the population.
- Direct **cumulative incidence** from 0-24 months:

Number of deaths
Initial population at risk

$$= 9 / 10 = 0.9$$

- Direct average time of occurrence among those who died during follow-up:

$$= \\ (1+17+20+9+16+2+13+10+3)/9 \\ = 10.1$$



Szklo & Nieto 2019; Figure 2.2



Let's take a closer look at what we mean by direct versus indirect methods of calculating cumulative incidence in cohort studies. On the right, we have a figure where we're following 10 people over 24 months.

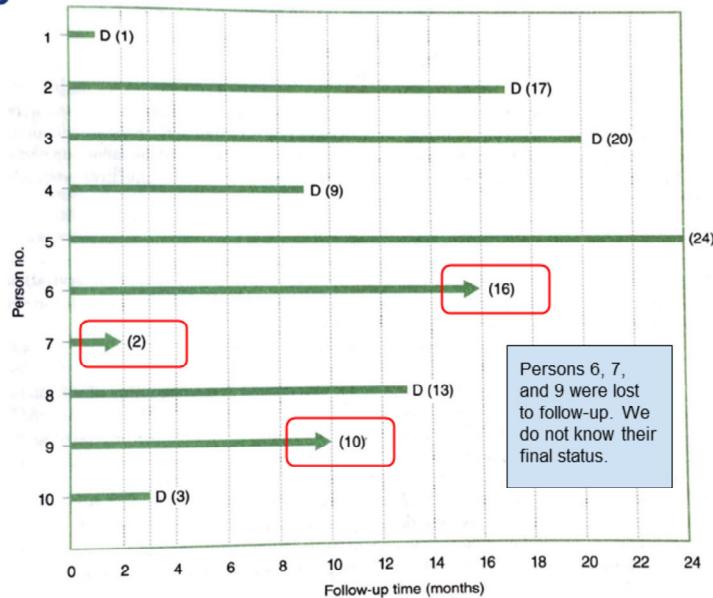
The green lines indicate the duration of follow-up. And in this particular example, I'm using a d with a number in parentheses to indicate whether the person passed away during follow-up and if so, at which month. In this example, there's no loss to follow-up, and there are no competing risks. We know the time of death for everyone in the population, and everyone except person five died during follow-up. Because we have information on everyone's outcome status, we can directly calculate the cumulative incidence from zero to 24 months using the simple formula for cumulative incidence.

We just take the number of deaths during follow-up and divide that by the initial population at risk at time zero. Our numerator is nine, because everyone except person five died during follow-up, and our denominator is 10, the number of people we started with at follow-up, which gives us 0.9. We can also directly calculate the average time of death or occurrence among those who died during follow-up. We'll exclude person five from this analysis, because they survived. Then we sum up the time till death of all the other people.

We add one plus 17 plus 20 plus nine and so on for everyone but person five, and then we divide by nine, since there were nine people who died during follow-up. And that gives us an average of 10 months until death.

Example of loss to follow-up / competing risks

- We don't know the final death/survival status for 3 people in the population, so we can't directly calculate
- For example, person 6 could have died between months 16 and 24 or later, so we don't know if they should be in the numerator of the cumulative incidence.
 - The Kaplan-Meier, Actuarial, or density method.
- We also can't calculate average time to death because we don't know what values to use for persons 6, 7, and 9.



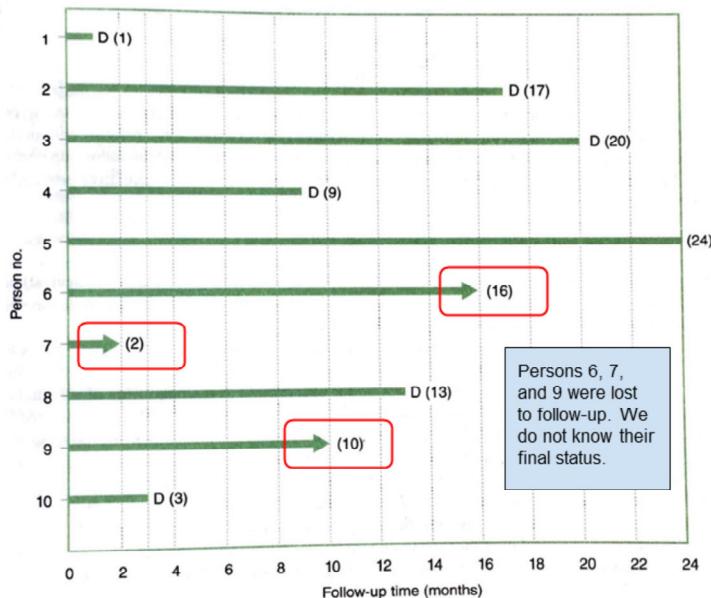
Szklo & Nieto 2019; Figure 2.2

Now, let's go over an example where there's loss to follow-up or competing risks in the population. So much of this figure is the same. Now, I've circled the outcomes status for person six, seven, and nine. In this example, the arrows indicate that they were lost to follow-up, so we don't know if they died at this time. We just don't have any information, for example for person six, beyond month 16.

So they could have died at month 16. They also could have died later, and we just don't know the answer. And as a result, we can't use the simple cumulative incidence formula, because we don't know the final outcome status for these individuals. We need to use other methods that we've previously learned, like the Kaplan Meier, actuarial, or density methods. We also can't calculate the average time to death, because we don't know what the values are for person six, seven, and nine. In other words, we don't know in what month they died, or if they died at all during the follow-up period.

Advantages of estimating incidence density in cohort studies

- The denominator is person-time instead of individuals which allows for a more flexible classification into exposed and unexposed groups.
- A single person can contribute person-time to multiple exposure groups in a single cohort study.
- An individual whose exposure experience changes with time can contribute person-time to either the exposed or unexposed group.



Szklo & Nieto 2019; Figure 2.2

DEI KE
10

This is why there are some advantages to estimating the incidence density in cohort studies, instead of the cumulative incidence. This is because the denominator is person time instead of individuals. For incidence density we're dividing by person time, and for cumulative incidents, we're dividing by the number of individuals at the beginning of the follow-up interval of interest. By using person time as the denominator, it allows us to have a more flexible classification of person time into exposed and unexposed groups when there's varying lengths of follow-up for different individuals in the population.

A single person can contribute person time to multiple exposure groups within one cohort study. So there may be periods of time when a person's person time is classified as exposed and another period of time when their person time is classified as unexposed, and the person time varies from person to person, which allows us to cleanly account for loss to follow up and competing risks.

Exposure classification in cohort studies

- Changes in exposure status at different points in follow-up can affect how we estimate measures of disease.
- The sequence of exposure and timing of exposure relative to age or disease onset could be important.
- Most epidemiologic studies assume that it is a numerical summary of exposure history determines the risk of disease. For example:
 - Current level of exposure
 - Average exposure
 - Cumulative exposure
 - Maximum exposure
- Often exposure is lagged, so that only exposure up to a certain time point before the time of disease measurement is classified as “exposed”.
- For chronic diseases, the time when exposure occurs may differ from the time when exposures have an effect.

Rothman ME3, Chapter 7

Let's talk more about how to appropriately classify exposure status in cohort studies. When we expect that exposure status can change at different points in the follow-up period, we need to think carefully about the appropriate way to measure exposure. In some cases, the sequence of exposure and timing of exposure relative to the age or timing of disease onset is important.

We'll go over a few examples of this. In most cases, in epidemiologic studies we make an assumption that we can use a numerical summary of exposure history to determine the risk of disease rather than measuring accurately the exact exposure history. For example, we may measure the current level of exposure and use that to assess the risk of disease, and that definition would ignore prior levels of exposure. We could average the level exposure over a period of time, or we could sum up the total exposure, which would give us a cumulative exposure, or we could take the maximum level of exposure.

Each of these different measures of exposure classification would lead to a different measure of association with disease.

Sometimes it's appropriate to lag an exposure. This means that there's only exposures up to a certain time point before disease that are classified as exposed, and the other remaining time would be classified as unexposed. We'll talk about this in more detail in a minute. And then finally, for chronic diseases the time when an exposure occurs may differ from the time when exposures have an effect, and we'll come back to this as well.

Chronic exposures in cohort studies

- Accumulation of exposure experience may be a complex function of the intensity of the exposure and time
- **Example of a chronic exposure:** smoking and lung cancer
- Exposure could be defined by “pack-years”, the number of cigarette packs smoked per year
 - Composite of both duration and intensity of smoking
 - 20 pack-years =
 - Half a pack a day for 40 years
 - 1 pack a day for 20 years
 - 2 packs a day for 10 years
 - Using composite measures can conceal biological effects associated with intensity vs. duration
- Other factors to consider: age of smoking initiation, age of smoking cessation, timing of exposure relative to disease



Let's look at the example of smoking and lung cancer, and we'll talk more how the accumulation of exposure experience can be complex. What's often done in studies of smoking is investigators will define something called pack years, which is the number of cigarette packs smoked per year.

This is a composite measure of both the duration and the intensity of smoking, and much like person years, pack years can be interpreted in many different ways. So if we say that somebody had 20 pack years of smoking, it could be equal to half a pack a day for 40 years, a pack a day for 20 years, or two packs a day for 10 years. What's tricky about using this kind of composite measure that blends duration and intensity is that it can conceal biological effects that are associated with either intensity or duration but not both. There's other elements of exposure that may be important to consider.

For example, the age when someone started smoking, the age when they stopped smoking, and then the timing of their smoking habit relative to the onset of their disease. The main takeaway is that for this kind of exposure where we hypothesize that the biological process that leads to disease may vary depending on the duration and the intensity of exposure, we need to think really carefully about exactly how we define exposure in our study.

Unexposed time for exposed subjects

- In most cohort studies, individuals can be unexposed and exposed at different points during follow-up.
- How should exposure be classified in the analysis?
 - (1) Consider any time not related to the exposure as unexposed time
 - Ignores possible threshold effects
 - Ignores exposure accumulation or induction period
 - (2) Omit from the study the experience of exposed subjects that is not at risk of exposure effects
 - Requires a larger sample size
- Improper definition of the induction period may lead to misclassification.

Here's a few more considerations about exposure classification, specifically about how to classify unexposed versus exposed time.

In many cohort studies, people can fall into both unexposed and exposed categories at different times during follow-up. Two possible approaches to this are first, we could consider any time not related to the exposure as unexposed time. This would ignore possible threshold effects and any exposure accumulation or induction periods. It's a pretty simplistic way of classifying exposure. So in the case of smoking, if we use this first approach, anytime that someone wasn't actively smoking we would classify them as unexposed, even if they had been a heavy smoker immediately before that period of time. And a threshold effect would be an effect where we expect disease to occur once someone has smoked a certain number of packs or a certain number of packs by a certain age.

If we simply classify the time when they're not smoking as unexposed, we wouldn't be able to correctly assess these kinds of threshold effects. A second approach is to omit from the study the experience of exposed subjects that is not at risk of exposure effects. This requires us to have a pretty careful definition of what we think the induction period is, and we'll go over that in a moment. It may also require a larger sample size, because we are fully omitting those people's person time from the study. Instead of classifying them as unexposed, we're just not including them in the denominator at all for those periods of time.

Example of a brief exposure and chronic disease

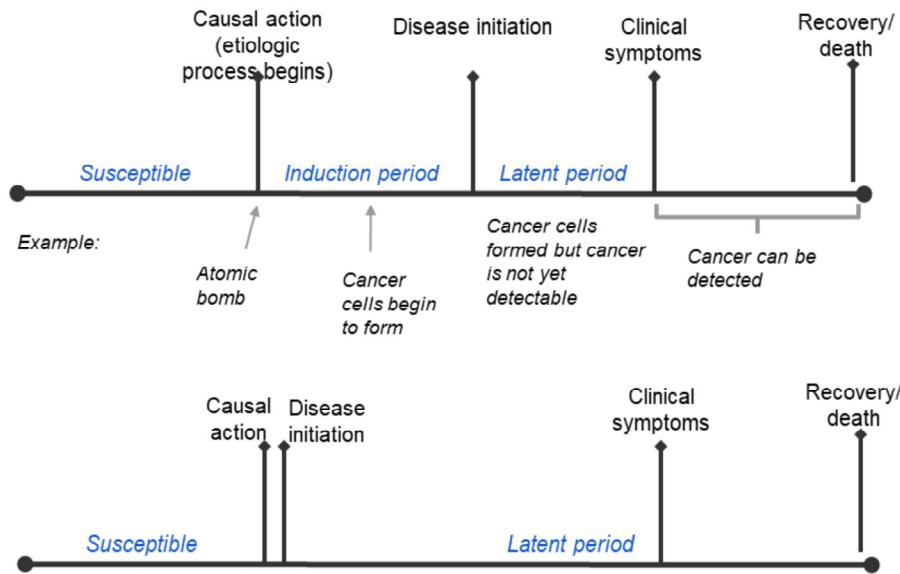
- Is radiation emitted from an atomic bomb associated with cancer risk?
- Exposure is nearly instantaneous, but the risk of disease does not increase immediately after exposure.
 - Some diseases occurring right after exposure can't have been caused by exposure, so are counted as having occurred during unexposed time.
- When we decide what time to include as exposed for an individual in the denominator of a rate, we are implicitly using a definition of the induction period.
- In this example, if we define follow-up time immediately after exposure as exposed, this implies a short induction period.
- The rate will be different with a short induction period than with a long induction period, so the choice of which time period to include in the denominator is important.



Here's an example of a very brief exposure and a chronic disease where decisions about which time counts as exposed can have a large impact. And the research question is, is radiation emitted from an atomic bomb associated with later cancer risk? The exposure is nearly instantaneous, because when an atomic bomb goes off, people are exposed to radiation immediately. But the disease develops slowly, and it may be a number of years before it's detectable. When we do a cohort study with this kind of brief exposure and a chronic disease outcome, we have to make a decision about whether we want to classify all of the follow-up time between the exposure and the outcome as exposed, or if we want to classify some of that follow up time as unexposed. That's because, for example, if an exposed person was diagnosed with leukemia 2 weeks after exposure, it is highly unlikely that the leukemia was a consequence of the radiation exposure. So the decision about what follow-up time is classified as exposed implies a certain definition of an *induction period.

In this example, if we , define follow-up time starting immediately after exposure, it implies a short induction period between the exposure and the onset of disease biologically. The rate will be different with a short induction period than with a long induction period because we classify a longer period of follow-up time as exposed. So the choice of which time period to include in the denominator for person time is important.

Example: atomic bomb & cancer



In this example, a certain induction period is assumed, and follow-up only begins at the point when disease initiation is hypothesized to have occurred.

When follow-up starts immediately after the causal action, a short induction period and a long latent period is implied. This increased the amount of time a person is considered exposed.

15

To go into a little more detail, here's a figure that you've seen before. It shows the progression of disease for a person. The person starts off as susceptible, and *an etiologic process that eventually leads to disease begins. Here the etiological process is the near instantaneous radiation exposure. This is the start of the induction period, because the cancer does not occur immediately upon radiation exposure. *The disease needs time to develop.

Eventually, *there is biological disease initiation in the body. And that's the start of the latent period. But often clinical symptoms are not present, and the disease is not immediately detectable. *Once the disease can be detected and there's clinical symptoms, that's the end of the latent period and the person is classified as diseased, and then they may recover or pass away. As we were discussing, the definition of induction period will affect the length of follow-up time that's classified as exposed.

Induction period as unexposed time

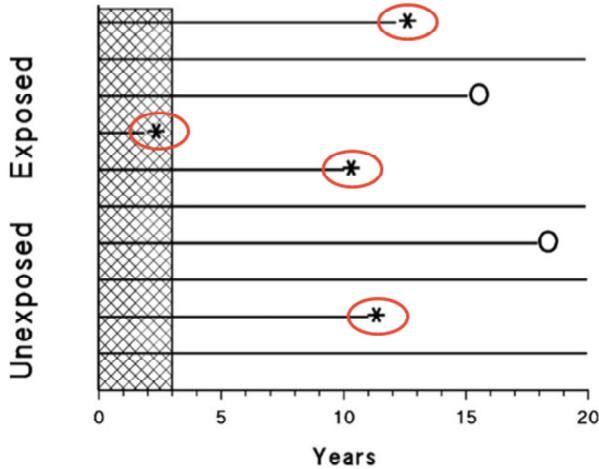
- Assume that the minimum time it takes for cancer to develop after radiation exposure is 3 years.
 - Any event occurring within the 3-year induction period counts towards the unexposed group.

3-year induction period:

$$\begin{aligned} \text{Exposed} &= \frac{2 \text{ events}}{45 \text{ years pt}} = 2.3 \text{ IRR} \\ \text{Unexposed} &= \frac{2 \text{ events}}{103 \text{ years pt}} \end{aligned}$$

No induction period:

$$\begin{aligned} \text{Exposed} &= \frac{3 \text{ events}}{59 \text{ years pt}} = 4.6 \text{ IRR} \\ \text{Unexposed} &= \frac{1 \text{ events}}{89 \text{ years pt}} \end{aligned}$$

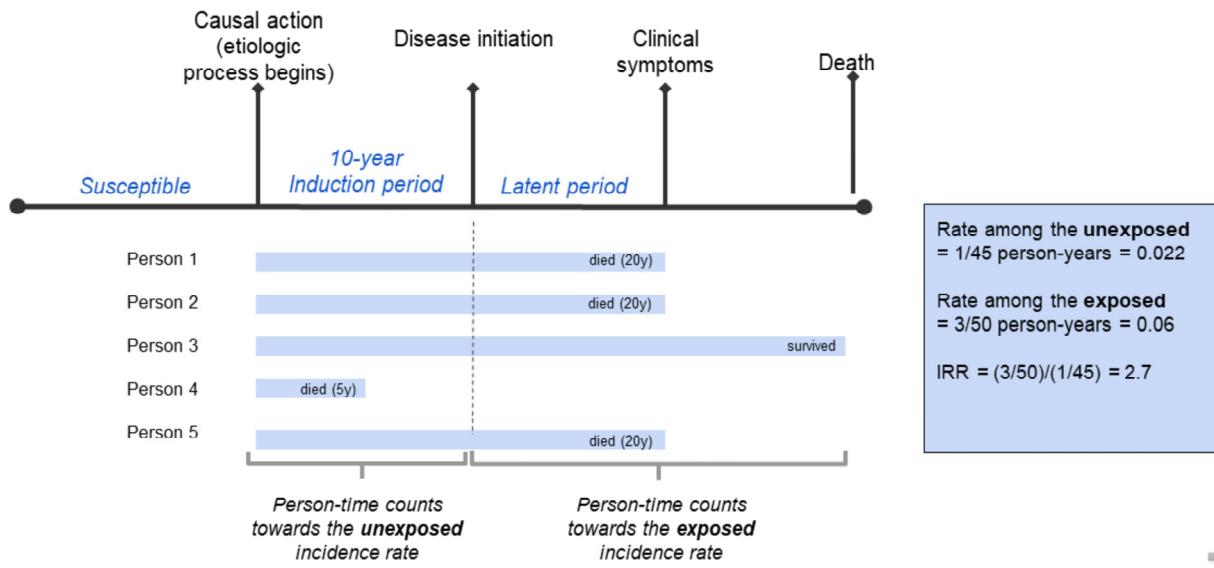


So here's an illustrative example. If you believe that nuclear radiation cannot cause cancer to develop in under 3 years, you define a 3-year induction period. Any event occurring within the 3-year induction period counts towards the unexposed group. So here on the right, there is a cohort of 5 exposed and 5 unexposed individuals, and a 3-year induction period that is shaded.

So two people develop cancer after the induction period in the exposed group. One person develops cancer in the unexposed group, but one exposed subject develops cancer during the induction period, so it counts towards the unexposed sum. So in the exposed group there are 2 events and 45 years of person time, for a rate of 4.4 per 100 years of person-time. And in the unexposed group, there are two events and 103 years of person-time (with 14 of them from the exposed group's induction period), for a rate of 1.9 per 100 years of person time. So the incidence rate ratio is 2.3.

Contrast that with the assumption of no induction period. Now there are 3 events in the exposed group and 59 years of person-time, for a rate of 5.1 cases per 100 person-years. In the unexposed group, the follow-up times total 89 years, and there was only one disease event, for a rate of 1.1/100 years of person time, and an incidence rate ratio of 4.6. So including the induction period as unexposed will always affect the incidence rate ratio by decreasing the denominator in the calculation of the exposed rate and increasing it in the unexposed, and sometimes moving cases from the exposed to unexposed, but the direction of change will depend on the specific amounts of person-time and cases changed from the exposed to unexposed.

Induction period definition affects potential misclassification

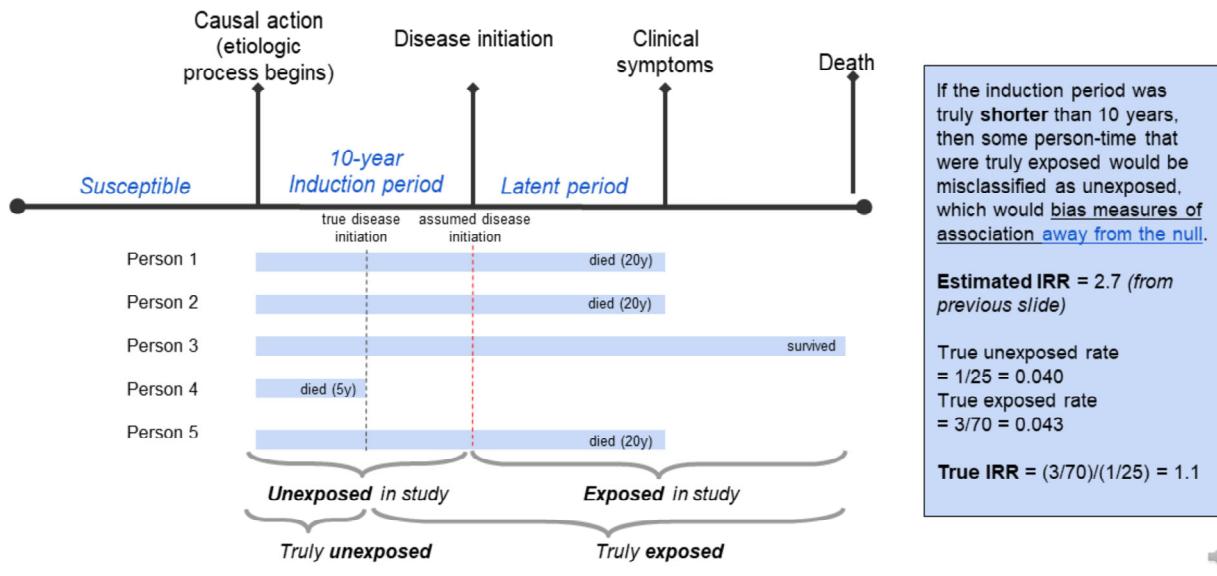


So let's talk about how the definition of the induction period effects potential misclassification. Here's a similar figure to before, where we start off susceptible. Here, there's then a 10-year induction period following the causal action in the exposure. Then disease initiation and a latent period, and then clinical symptoms develop. Here, we're following five people over time, and we'll do similar calculations as before, but here we are only contrasting exposure after the induction period to being unexposed before.

So if we just look within the first 10 years of follow-up, the induction period, we can see that persons one, two, three, and five survived the entire induction period, but one person died during the induction period.

To calculate the rate among the unexposed, there's one person who died, so our numerator is one. And then if we count the total person years of follow-up, person one, two, three, and five had 10 years of follow-up, that's 40 person years, and then there was five more person years from person four, so a total of 45 person years. One over 45 gives us 0.022 for the rate among the unexposed. And then we can do the same thing in the exposed group and get a rate of 0.06. And then we can calculate an incidence rate ratio, which is 2.7.

Induction period definition affects potential misclassification



Now, we assumed that disease initiation occurred 10 years after the exposure and the causal action. But what if we were wrong, and what if the true end of the induction period was at five years instead of at 10 years? Well, if that's the case, then our measure of association would be biased away from the null.

Let's go over why that is.

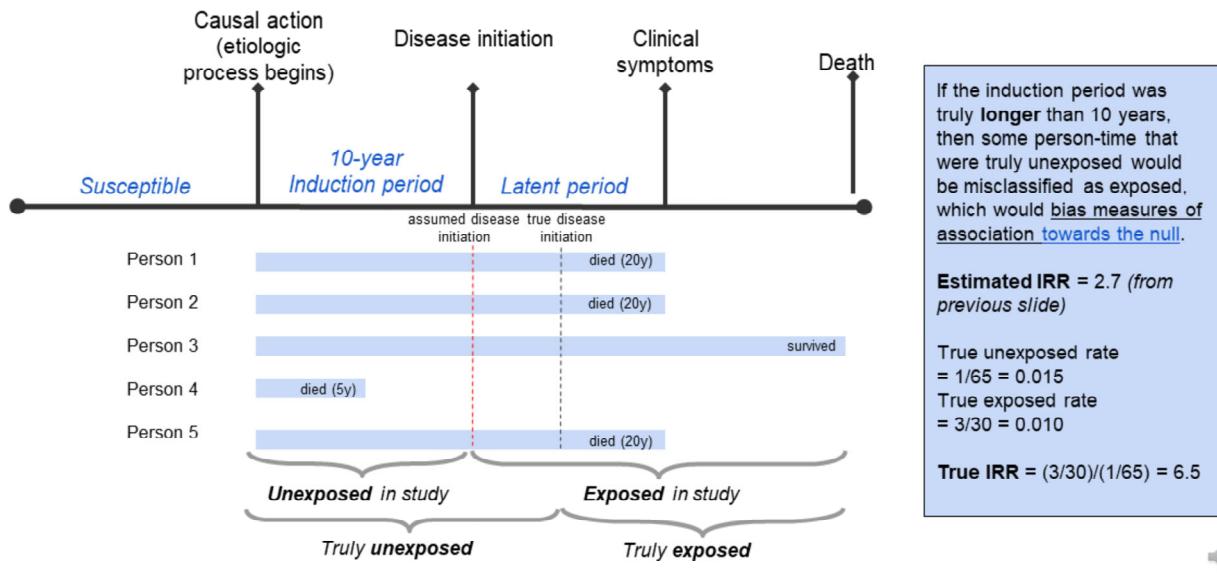
The outcome status is the same for each of these five people in this example. We're just moving the exposure classification for the truth back by five years. From years zero to five, people were truly unexposed, and from years five through 30, people were truly exposed. On the previous slide, we calculated an incidence rate ratio of 2.7, using the 10-year cutoff for the induction period. Now, let's calculate the incidence ratio using the five year induction period.

We'll start by calculating the true unexposed rate. So one person died within the first five years. They were followed up all for five years of time, so five people times five years of follow-up is 25 person years, and one over 25 is 0.04.

And then we can calculate the true exposed rate. If we look at the data from the black dashed line forward, we still have three people who died, but now we have a total of 70 person years, so our true exposed rate is three over 70, which is 0.043, and our new, true incident rate ratio is 1.1.

Our previously-estimated incident rate ratio is 2.7, so what we can see is that this misclassification of exposure caused our measure of association to be biased away from the null. That's because if the induction period was truly shorter than 10 years, that would mean that some person time that was truly exposed would be misclassified as unexposed, increased the denominator in the unexposed rate and decreasing it in the exposed rate, which inflates the ratio.

Induction period definition affects potential misclassification



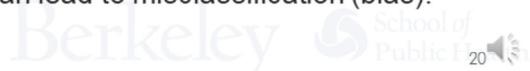
Let's go over the same thing with an example in which the true disease initiation was after 10 years. So the induction period was truly longer than 10 years, and in this case, we will see bias towards the null. Our true unexposed rate again has one case in the numerator, but our person time changes to 65 person years, and that gives us an unexposed rate of 0.015. And then we can do the same thing for the true exposed rate, and we get three out of 30 person years, which is 0.01.

And our true incident rate ratio is then 6.5. When we compare our estimated and our true incident rate ratio, what we see is our estimated incident rate ratio is closer to the null than our true incident rate ratio, which means that our measure of association is biased towards the null. I've gone into quite a bit of detail in this example mainly to show you that we have to be very careful about the way that we define our induction period and the choices that we make in how we decide to classify person time as exposed or unexposed, because in this very simple example, you can see that the amount of bias is actually quite large depending on the decision that we make.

So what's often done in practice is to calculate incidence rates among the exposed and unexposed with different induction times following the exposure, with a range guided by biological plausibility based on prior research. This will lead to different results, but that can be used to assess the sensitivity of the results to a range of induction times.

Summary of key points

- In most cohort studies, there is some loss to follow-up and some competing risks.
- As a result, to calculate cumulative incidence we must use the Actuarial, Kaplan Meier, or Density methods. We can calculate incidence density directly.
- Most cohort studies define a numerical summary of exposure history that is expected to determine the risk of disease.
 - The precise definition of the summary of exposure will affect measures of association as well as interpretation of the study.
- When exposure is time-varying, we can classify it by:
 - (1) Considering any time not related to the exposure as unexposed time
 - (2) Omitting from the study the experience of exposed subjects that is not at risk of exposure effects
- An incorrect definition of the induction period can lead to misclassification (bias).



To summarize the points from this video, first in most cohort studies there's some loss to follow-up and at least some competing risks. And as a result, to calculate the cumulative incidence, we usually need to use the actuarial, Kaplan Meier, or density methods instead of using the simple direct formula for cumulative incidence. But we can calculate the incidence density directly, and that's because the denominator uses person time instead of persons.

Second, most cohort studies define a numerical summary of exposure history that's expected to determine the risk of disease, but the exact definition that we choose to use to summarize exposure will affect our final measures of association and how we interpret the study.

So we need to be very careful about how we make this definition. When our exposure varies over time, which is very common, we talked about two different options for classifying exposed versus unexposed person time. First is considering any time not related to the exposure as unexposed time, which allows us to include all the person time that we've followed up, but may lead to misclassification depending on how we define our induction period, and the second is omitting from the study the experience of exposed subjects that is not at risk of exposure effect.

This is essentially throwing away some information we calculated during follow-up, and as a result, it may require us to enroll a larger sample size.

Lastly, an incorrect definition of the induction period can lead to misclassification bias, which we showed in the last few slides.

Epidemiology series

Cohort studies: marching towards outcomes

David A Grimes, Kenneth F Schulz

A cohort study tracks two or more groups forward from exposure to outcome. This type of study can be done by going ahead in time from the present (prospective cohort study) or, alternatively, by going back in time to comprise the cohorts and following them up to the present (retrospective cohort study). A cohort study is the best way to identify incidence and natural history of a disease, and can be used to examine multiple outcomes after a single exposure. However, this type of study is less useful for examination of rare events or those that take a long time to develop. A cohort study should provide specific definitions of exposures and outcomes: determination of both should be as objective as possible. The control group (unexposed) should be similar in all important respects to the exposed, with the exception of not having the exposure. Observational studies, however, rarely achieve such a degree of similarity, so investigators need to measure and control for confounding factors. Reduction of loss to follow-up over time is a challenge, since differential losses to follow-up introduce bias. Variations on the cohort theme include the before-after study and nested case-control study (within a cohort study). Strengths of a cohort study include the ability to calculate incidence rates, relative risks, and 95% CIs. This format is the preferred way of presenting study results, rather than with p values.

The term cohort has military, not medical, roots. A cohort was a 300–600-man unit in the Roman army; ten cohorts formed a legion (figure 1). The etymology of the term provides a useful mnemonic: a cohort study consists of bands or groups of persons marching forward in time from an exposure to one or more outcomes.

This analogy might be helpful, since cohort studies have a bevy of confusing synonyms: incidence, longitudinal, forward-looking, follow-up, concurrent, and prospective study.^{1,2} Although the terminology can seem daunting, the cohort study is easy for clinicians to understand, since it flows in a logical direction (unlike the case-control study). Here, we explain the terminology, describe the strengths and weaknesses of cohort studies, consider several logistical concerns, mention two permutations of cohort studies, and summarise their analysis.

Data collection: forwards and backwards

A cohort study follows-up two or more groups from exposure to outcome. In its simplest form, a cohort study compares the experience of a group exposed to some factor with another group not exposed to the factor. If the former group has a higher or lower frequency of an outcome than the unexposed, then an association between exposure and outcome is evident.

The defining characteristic of all cohort studies is that they track people forward in time from exposure to outcome. Researchers doing this kind of study must, therefore, go forward in time from the present or go back in time to choose their cohorts (figure 2). Either way, a cohort study moves in the same direction, although gathering data might not. For example, an investigator who wants to study the epidemic of multiple births stemming from assisted reproductive technologies³ could begin a cohort study now. Women exposed to these technologies and a similar group who conceived naturally



Figure 1: An early cohort in search of favourable outcomes

could be tracked forward through their pregnancies to monitor the frequency of multiple births (a concurrent cohort study). Alternatively, the investigator might use existing medical records and go back in time several years to identify women exposed and not exposed to these technologies. He would then track them forward through records to note the birth outcomes. Again, the study moves from exposure to outcome, though the data collection occurred after the fact.

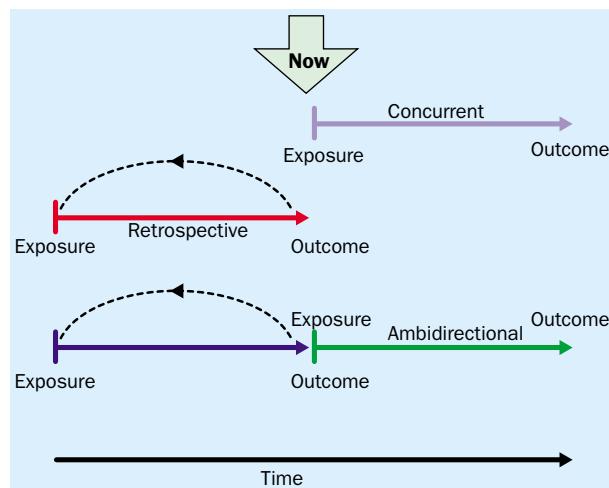


Figure 2: Schematic diagram of concurrent, retrospective, and ambidirectional cohort studies

Lancet 2002; 359: 341–45

Family Health International, PO Box 13950, Research Triangle Park, NC 27709, USA (D A Grimes MD, K F Schulz PhD)

Correspondence to: Dr David A Grimes
(e-mail: dgrimes@fhi.org)

Yet a third variation exists: ambidirectional.⁴ As the name implies, data collection goes in both directions. This approach can be useful for exposures that have both short-term and long-term outcomes. In this hypothetical example, assisted reproductive technologies might be associated with multiple births and with ovarian cancer in later life.⁵ The investigator might, therefore, look back through records for multiple births and also start to follow-up these women into the future for ovarian cancer occurrence.

Advantages of cohort studies

Cohort studies have many appealing features. They are the best way to ascertain both the incidence and natural history of a disorder.⁶ The temporal sequence between putative cause and outcome is usually clear: the exposed and unexposed can often be seen to be free of the outcome at the outset. By contrast, this chicken-egg question often frustrates cross-sectional and case-control studies. For example, in a case-control study, patients with chronic widespread pain were more likely to have mental illness than controls.⁷ Do mood and anxiety disorders increase this risk, or do patients with chronic pain develop mood and anxiety disorders as a result of their disorder?

Cohort studies are useful in investigation of multiple outcomes that might arise after a single exposure. A prototype would be cigarette smoking (the exposure) and stroke, emphysema, oral cancer, and heart disease (the outcomes). Although assessment of many outcomes is often cited as a positive attribute of cohort studies, this feature can be abused. For example, testing the associations between exposure and many outcomes, but only reporting the significant ones, represents misleading science. Investigators should preferably have planned primary and secondary associations to examine (sometimes called hypothesis confirmation). Although investigators can look at other outcomes (hypothesis generation), they should report the findings of all examinations, not just significant ones, so that readers can correctly interpret the results.

The cohort design is also useful in the study of rare exposures: a researcher can often recruit people with uncommon exposures—eg, to ionising radiation or chemicals—in the workplace. A hospital or factory might provide a large number of individuals with the exposure of interest, which would be rare in the general population. Since the investigator does not assign exposure, no ethical concerns arise.

Cohort studies also reduce the risk of survivor bias.⁶ Diseases that are rapidly fatal are difficult to study because of this factor. For example, a hospital-based case-control study of the link between snow-shovelling and myocardial infarction would miss all those who died in the driveway. A cohort study would be a less biased (but more cumbersome) approach: compare rates of myocardial infarction among those who shovel and those who do not shovel. Finally, cohort studies allow calculation of incidence rates, relative risks, and confidence intervals.² Other outcome measures in cohort studies include life-table rates, survival curves, and hazard ratios (panel 1).^{8–10} By contrast, case-control studies cannot provide incidence rates; at best, odds ratios approximate relative risks only when the outcome is uncommon.

Disadvantages of cohort studies

Cohort studies have important limitations too. Selection bias is built into cohort studies. For example, in a cohort study investigating effects of jogging on cardiovascular

Panel 1: Reporting time-to-event in cohort studies

Survival analysis

Survival analysis is useful when lengths of follow-up vary substantially or when participants enter a study at different times.⁸ The Kaplan-Meier method provides a more sophisticated expression of the risk of the outcome over time than does a simple dichotomous outcome.⁹ It can determine the probability (P) of the outcome at any point in time; this result is graphed as a step function (which jumps at every event). A complementary, mirror-image graph portrays the likelihood of avoiding the outcome (1-P) as a function of time (Kaplan-Meier survival curve). The log-rank test compares survival curves of different groups.¹⁰

Proportional hazard model

Another approach to different lengths of follow-up is the Cox proportional hazard model. It is a multivariate technique that has time-to-event (such as illness) as the dependent variable. By contrast, multiple logistic regression has "yes-no" as the dependent variable.⁸ Coefficients from this model can be used to calculate the risk ratio (hazard ratio) of the outcome, after controlling for other covariates in the equation. The hazard ratio (with 95% CIs) is interpreted in the same way as a relative risk for dichotomous outcomes.¹⁰

disease, those who choose to jog probably differ in other important ways (such as diet and smoking) from those who do not exercise.¹¹ In theory, both groups should be the same in all important respects, except for the exposure of interest (jogging), but this seldom occurs. The cohort design is not optimum for rare diseases—eg, scleroderma—or those that take a long time to develop—eg, cancer. However, several large (and thus expensive) cohort studies have made landmark contributions to our knowledge of uncommon diseases. Examples include the Royal College of General Practitioners' Oral Contraceptive Study,¹² the Framingham Heart Study,¹³ the Nurses Health Study,¹⁴ and the British Physicians' Study.¹⁵

Loss to follow-up can be a difficulty, even at 1 month, and particularly so with longitudinal studies that continue for decades. Differential losses to follow-up between those exposed and unexposed can bias results. Over time, the exposure status of study participants can change. For example, a proportion of women who use oral contraceptives will switch to an intrauterine device, and vice versa.¹⁶ Partitioning might be needed to avoid a blurring of exposure, sometimes termed contamination.

What to look for in cohort studies

Who is at risk?

All participants (both exposed and unexposed) in a cohort study must be at risk of developing the outcome.⁶ For example, since women who have had a tubal sterilisation operation have almost no risk of salpingitis,¹⁷ they should not be included in cohort studies of pelvic inflammatory disease.

Who is exposed?

Cohort studies need a clear, unambiguous definition of the exposure at the outset. This definition sometimes involves quantifying the exposure by degree, rather than just yes or no. For example, the minimum exposure might have to be 14 cigarettes per day or less,¹⁸ or 3–6 months of oral contraceptives.¹⁹ Definition of exposure levels in this way can result in more than two groups—eg, non-smokers, light smokers, and heavy smokers.¹⁸

Who is an appropriate control?

The key notion is that controls (the unexposed) should be similar to the exposed in all important respects, except for the lack of exposure. If so, the unexposed group will reveal the background rate of the outcome in the community.

The unexposed group can come from either internal (persons from the same time and place, such as a hospital ward) or external sources. Internal comparisons are most desirable. In a particular population, individuals segregate by themselves (or through medical interventions) into exposure status—eg, cigarette smoking, occupation, contraception. For example, in a cohort study, 138 patients with HIV-1-associated Kaposi's sarcoma were divided into two groups: those with oral and those with cutaneous lesions. The presence of oral lesions (the exposure) had a poorer prognosis, with a median survival (the outcome) one-third that of the other group.²⁰

If satisfactory internal controls are not available, researchers look elsewhere (sometimes termed a double-cohort study).⁶ In a trial of an occupational exposure, finding an adequate number of employees in the factory without the exposure might be difficult. Hence, one might choose workers in a similar factory in the same community. This choice assumes that workers in the other factory have the same baseline risk of the outcome in question, which might not be the case. Even less desirable is use of population norms; disease-specific mortality rates are an example. A researcher might compare lung-cancer death rates among workers in the factory with rates of persons of the same age and sex in the population. Bias inevitably creeps into such comparisons because of the healthy worker effect: those who work are healthier, in general, than those who do not (or cannot) work.^{4,9} Additionally, work reaps economic benefits which might further bias comparisons.

Have outcomes been assessed equally?

Outcomes must be defined in advance; they should be clear, specific, and measurable. Identification of outcomes should be comparable in every way for the exposed and unexposed to avoid information bias. Failure to define objective outcomes leads to uninterpretable results. This challenge relates not only to subjective syndromes such as Gulf War,²¹ chronic fatigue,^{22,23} and premenstrual,²⁴ but also to more mundane health problems such as endometritis. Just how tender must a uterus be? Keeping those who judge outcomes unaware of the exposure status of participants (blinding) in a cohort study is important for subjective outcomes, such as tenderness or erythema. By contrast, with objective outcome measures, such as fever or death, blinding the exposure status is less important.

Outcome information can come from many sources. For mortality studies, the death certificate is often used. Although convenient, the validity of the clinical information is highly variable. For non-fatal outcomes, sources include hospital charts, insurance records, laboratory records, disease registries, hospital discharge logs, and physical examination and measurement of participants. Optimally, the person who judges outcomes should be unaware of the exposure. When diagnoses vary in their confidence, assignment of levels of assurance might be helpful, such as definite, probable, and suspect.⁹

Tracking participants over time**Have losses been minimised?**

Although loss of participants damages the power and precision of a study, differential loss to follow-up is more sinister. Bail-outs are not random events. If the likelihood

of bailing out is related both to exposure and outcome, then bias can result.²⁵ For example, some participants given a new antibiotic might have such poor outcomes that they are unable to complete questionnaires or to return for examination.²⁶ Their disappearance from the cohort would make the new antibiotic look better than it is.

The best way of dealing with loss to follow-up is to avoid it. For example, restrict participation to only those judged likely to complete the study. Additionally, several safeguards are customary. Obtaining the names of several family members or friends who do not live with the respondent is often helpful at the start of such studies. The participant's family doctor might also be helpful. Should the respondent move, these contacts would probably know their new address. Motor vehicle registration records can be useful too. Furthermore, national vital statistics registries, such as the National Death Index in the USA, facilitate follow-up. Participants can be offered financial compensation for their time lost from work as a result of the study. Diligent tracking of participants is hard work, and might require hiring personnel for this task alone.

Reporting cohort studies

Many researchers who do cohort studies report their findings in an unsatisfactory way (panel 2).²⁷ An investigator's first challenge is to convince the editor (then readers) that the exposed and unexposed groups were indeed similar in all important respects, except for the exposure. The first table in reports of cohort studies customarily provides demographic and other prognostic factors for both groups with hypothesis testing (p values) to show the likelihood that observed differences could be due to chance.

For dichotomous outcome measures, such as sick or well, the investigator should provide raw data sufficient for the reader to confirm the results. For cumulative incidence, the investigator should calculate the proportion who developed the outcome during the specified study interval. For incidence rates, the value is expressed per unit of time.⁴ Then, relative risks and confidence intervals should be provided. Use of p values should not replace interval estimation (relative risks with confidence

Panel 2: Features to look for in a cohort study**How much selection bias was present?**

- 1 Were only people at risk of the outcome included?
- 1 Was the exposure clear, specific, and measurable?
- 1 Were the exposed and unexposed groups similar in all important respects except for the exposure?

What steps were taken to minimise information bias?

- 1 Was the outcome clear, specific, and measurable?
- 1 Was the outcome identified in the same way for both groups?
- 1 Was determination of outcome made by an observer blinded as to treatment?

How complete was the follow-up of both groups?

- 1 What efforts were made to limit loss to follow-up?
- 1 Was loss to follow-up similar in both groups?

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

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

intervals)²⁸ and should only be used as supplemental information.²⁵

Like other observational studies, cohort studies have built-in bias. Investigators should identify potential biases in their data and show how these might have affected results. Whenever possible, confounding should be controlled for in the analysis. These techniques are discussed in an earlier essay in this series.²⁹

Variations on the cohort theme

Before-after studies

Before-after studies (time series) have important limitations. Here, an investigator takes a measurement, exposes participants to an intervention (often a drug), repeats the measurements, then compares them. First, regression to the mean is often ignored. If admission to the cohort includes extreme measurements,³⁰ such as high laboratory values, then lower mean values will arise at follow-up, irrespective of treatment.³¹ Second, secular trends, such as seasonal changes in the frequency of pneumonia, can affect results. Third, washout periods are often needed to avoid a carryover effect of drugs given during the initial observation period.⁶

Nested case-control studies

Cohort studies sometimes spawn other studies. One of the most frequent is the nested case-control study.^{6,9,25} Why would an investigator carve out a case-control study in the midst of a cohort study? The answer often involves body fluids and a freezer. Some exposure or predictor variables are simply too expensive to determine on everyone in a study. A sophisticated blood test is the prototype. A clever way to skirt this financial obstacle is to do a cohort study that will yield a sufficient number of cases. All participants entering the cohort study have a tube of blood drawn at enrolment; serum is frozen until the study's conclusion. All those in the cohort study who develop the outcome of interest now become the cases for the nested study. The investigator then chooses a random sample of all participants who did not develop the outcome (controls). Next, the blood test is done on serum from only the cases and controls, not the whole group of exposed and unexposed. In this way, the laboratory cost is minimised while assuring that the exposure—eg, a positive laboratory test—was present before development of the outcome. Controls are generally matched to cases by important characteristics, such as age and sex.⁹

A nested case-control study, for example, examined the potential relation between body concentrations of organochlorines and non-Hodgkin's lymphoma. The blood samples were obtained on entry to a large cohort study started in Maryland, USA, in 1974. Blood samples were eventually analysed for only 74 individuals with lymphoma and 147 controls.³² Thus, instead of measuring organochlorine concentrations of the entire cohort of 25 802, the investigators incurred this laboratory expense for less than 1% of the cohort. In view of the availability of banked blood specimens around the world, this type of research design is likely to become popular. However, nested case-control studies might be useful for other studies that do not require blood tests but in which determination of the exposure is expensive or difficult⁹—eg, measurement of nerve conduction³³ or job stressors.³⁴

Conclusion

Cohort studies are common in medical research. Like other research designs, they entail important trade-offs. Readers should make sure that investigators provide clear,

specific, and measurable definitions of exposures and outcomes. The unexposed group should resemble the exposed group in all important respects, and determination of outcomes should be objective and, whenever possible, blinded. Results for dichotomous outcomes should be provided as rates, relative risks, and confidence intervals, which offer more information than do p values. Reports of cohort studies should identify and describe the potential effect of biases. Importantly, investigators should measure and control for potential confounding.

We thank Willard Cates and David L Sackett for their helpful comments on an earlier version of this report. Much of this material stems from our 15 years of teaching the Berlex Foundation Faculty Development Course.

References

- 1 Last JM, ed. A dictionary of epidemiology, 2nd edn. New York: Oxford University Press, 1988.
- 2 Lilienfeld AM, Lilienfeld DE. Foundations of epidemiology, 2nd edn. New York: Oxford University Press, 1980.
- 3 Anon. Contribution of assisted reproductive technology and ovulation-inducing drugs to triplet and higher-order multiple births—United States, 1980–1997. *MMWR Morb Mortal Wkly Rep* 2000; **49**: S35–38.
- 4 Hennekens CH, Buring JE. Epidemiology in medicine. Boston: Little, Brown and Company, 1987.
- 5 Whittemore AS, Harris R, Itnyre J. Characteristics relating to ovarian cancer risk: collaborative analysis of 12 US case-control studies, II: invasive epithelial ovarian cancers in white women. Collaborative Ovarian Cancer Group. *Am J Epidemiol* 1992; **136**: 1184–203.
- 6 Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman RB, eds. Designing clinical research: an epidemiologic approach, 2nd edn. Baltimore: Lippincott Williams and Wilkins, 2001.
- 7 Benjamin S, Morris S, McBeth J, Macfarlane GJ, Silman AJ. The association between chronic widespread pain and mental disorder: a population-based study. *Arthritis Rheum* 2000; **43**: 561–67.
- 8 Wassertheil-Smoller S. Biostatistics and epidemiology: a primer for health professionals, 2nd edn. New York: Springer-Verlag, 1995.
- 9 Kelsey JL, Whittemore AS, Evans AS, Thompson WD. Methods in observational epidemiology, 2nd edn. New York: Oxford University Press, 1996.
- 10 Lang TA, Seicic M. How to report statistics in medicine. Philadelphia: American College of Physicians, 1997.
- 11 Sackett DL. Bias in analytic research. *J Chron Dis* 1979; **32**: 51–63.
- 12 Beral V, Hermon C, Kay C, Hannaford P, Darby S, Reeves G. Mortality associated with oral contraceptive use: 25 year follow up of cohort of 46 000 women from Royal College of General Practitioners' oral contraception study. *BMJ* 1999; **318**: 96–100.
- 13 Seman LJ, DeLuca C, Jenner JL, et al. Lipoprotein(a)-cholesterol and coronary heart disease in the Framingham Heart Study. *Clin Chem* 1999; **45**: 1039–46.
- 14 Colditz GA, Rosner BA, Speizer FE. Risk factors for breast cancer according to family history of breast cancer. *J Natl Cancer Inst* 1996; **88**: 365–71.
- 15 Doll R, Peto R, Boreham J, Sutherland I. Smoking and dementia in male British doctors: prospective study. *BMJ* 2000; **320**: 1097–102.
- 16 Hannaford PC, Kay CR. The risk of serious illness among oral contraceptive users: evidence from the RCGP's oral contraceptive study. *Br J Gen Pract* 1998; **48**: 1657–62.
- 17 Levigur M, Duvivier R. Pelvic inflammatory disease after tubal sterilization: a review. *Obstet Gynecol Surv* 2000; **55**: 41–50.
- 18 Croft P, Hannaford P. Risk factors for acute myocardial infarction in women. *BMJ* 1989; **298**: 674.
- 19 Anon. The reduction in risk of ovarian cancer associated with oral-contraceptive use. The Cancer and Steroid Hormone Study of the Centers for Disease Control and the National Institute of Child Health and Human Development. *N Engl J Med* 1987; **316**: 650–55.
- 20 Rohrmos B, Thoma-Greber EM, Bogner JR, Rocken M. Outlook in oral and cutaneous Kaposi's sarcoma. *Lancet* 2000; **356**: 2160.
- 21 Fukuda K, Nisenbaum R, Stewart G, et al. Chronic multisymptom illness affecting Air Force veterans of the Gulf War. *JAMA* 1998; **280**: 981–98.
- 22 Schlueterberg A, Straus SE, Peterson P, et al. NIH conference. Chronic fatigue syndrome research: definition and medical outcome assessment. *Ann Intern Med* 1992; **117**: 325–31.
- 23 Armon C, Kurland LT. Chronic fatigue syndrome: issues in the diagnosis and estimation of incidence. *Rev Infect Dis* 1991; **13** (suppl 1): S68–72.

- 24 Haskett RF, Abplanalp JM. Premenstrual tension syndrome: diagnostic criteria and selection of research subjects. *Psychiatr Res* 1983; **9**: 125–38.
- 25 Rothman KJ. Modern epidemiology. Boston: Little, Brown and Company, 1986.
- 26 Sackett DL, Haynes RB, Guyatt GH, Tugwell P. Clinical epidemiology: a basic science for clinical medicine, 2nd edn. Boston: Little, Brown and Company, 1991.
- 27 Bracken MB. Reporting observational studies. *Br J Obstet Gynaecol* 1989; **96**: 383–88.
- 28 Gardner MJ, Altman DG. Confidence intervals rather than P values: estimation rather than hypothesis testing. *BMJ* 1986; **292**: 746–50.
- 29 Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet* 2002; **359**: 248–52.
- 30 Ewan PW, Clark AT. Long-term prospective observational study of patients with peanut and nut allergy after participation in a management plan. *Lancet* 2001; **357**: 111–15.
- 31 Bland JM, Altman DG. Regression towards the mean. *BMJ* 1994; **308**: 1499.
- 32 Rothman N, Cantor KP, Blair A, et al. A nested case-control study of non-Hodgkin lymphoma and serum organochlorine residues. *Lancet* 1997; **350**: 240–44.
- 33 Atterbury MR, Limke JC, Lemasters GK, et al. Nested case-control study of hand and wrist work-related musculoskeletal disorders in carpenters. *Am J Ind Med* 1996; **30**: 695–701.
- 34 Kawakami N, Araki S, Kawashima M. Effects of job stress on occurrence of major depression in Japanese industry: a case-control study nested in a cohort study. *J Occup Med* 1990; **32**: 722–25.

Uses of error

The error cascade

Neil Gittoes

Having just been appointed consultant physician, I found myself reflecting on my career and realising that I could at last stand alone and that finally the buck stops with me. There was a time that I wished it didn't. I was a medical senior house officer when I saw an elderly man who described a subacute onset of breathlessness and a dry cough. He had trouble speaking and was using his accessory muscles. He had initially received standard nebulised treatment for exacerbated chronic obstructive pulmonary disease, although his chest radiograph showed a small pneumothorax on the left. After conferring with senior colleagues, I inserted a chest drain on the left, and verified its position with a second radiograph. In the middle of the night the house officer saw the patient with worsening shortness of breath and surgical emphysema. He pushed the tube in further, but an hour later the arrest team were called because the patient had developed extreme respiratory distress and had become cyanosed. They thought that he had developed a contralateral

pneumothorax and proceeded to insert a chest drain on the right. Arriving on the ward the following morning, I was horrified to find my patient with bilateral chest drains and surgical emphysema from head to scrotum. However, at least he was alive. Chest radiographs and computed tomography showed bilateral pneumothoraces with both drains embedded deeply within the lung parenchyma, just short of the mediastinum on the right, and abutting the left ventricle on the left. I inserted bilateral anterior drains and cautiously removed the lateral ones. After a few days the right-sided pneumothorax resolved, although the left side needed surgical correction. He was finally discharged, and on reviewing the radiographs it was apparent that I had inserted the original drain where there was a small area of pleural adhesion. The two pleural surfaces remained contiguous, and the drain entered the lung parenchyma. The subsequent errors of management turned the situation rapidly into a life-threatening predicament.

Queen Elizabeth Hospital, Edgbaston, Birmingham, B15 2 TH UK (N Gittoes MRCP)

Evolution of the Cohort Study

Jonathan M. Samet and Alvaro Muñoz

INTRODUCTION

The occurrence of events over time unifies epidemiologic research. Regardless of the study hypothesis and design, the disease-causing actions of exposures and modifying factors are formulated as antecedent to the occurrence of the outcome. All study designs inherently acknowledge time and represent alternative approaches for sampling populations as exposed and nonexposed persons develop disease over time.

The cohort design explicitly incorporates the passage of time. In cohort studies, participants are followed over an interval defined by the study's beginning and end, and observations are made on outcome measures of interest: death, incidence of disease, change in a biologic measure, or health status. The study's purpose may be focused—to test a specific (or general) hypothesis—to gather data for descriptive purposes or to facilitate the testing of multiple hypotheses concerning disease. During the follow-up experience of participants in a cohort study, the factors determining the health of the participants may be continually changing as the participants age and exposures to environmental agents start, stop, increase, or decrease (figure 1). The dynamic nature of many risk factors and their relations in time to disease occurrence can only be captured in the cohort design; this temporal interplay is inherently absent from cross-sectional data and often investigated only with difficulty using the case-control design (1).

Epidemiologists may be challenged in investigating the relation between multiple risk factors, some changing in time, and disease occurrence, as displayed schematically in figure 1. During the follow-up of a cohort, participants age, temporal trends may affect the participants, and exposures to risk factors of primary

interest may change, as may potential confounding and modifying factors. Consider, for example, a prospective cohort study of cigarette smoking and lung cancer. Several determinants of lung cancer risk would change as the cohort was followed, including the participants' age (relevant because of the rise of lung cancer incidence with age), the cumulative amount smoked, the duration of smoking, and perhaps the characteristics of the cigarettes themselves. Some participants might choose to reduce the number of cigarettes smoked or to stop smoking, and some might start smoking. The study's data might be further complicated by a temporal trend in the validity of smoking information, if, for example, the social acceptability of smoking declined during the follow-up interval and participants began to underreport the extent of their smoking. Possible modifying or confounding factors (e.g., occupational exposures) might also change over time. An optimal study design would incorporate periodic assessment of smoking by the participants, as assessing smoking only at the study's start would not capture the temporally dynamic nature of the exposure.

Many examples of contemporary cohort studies show that the challenge of temporally varying exposures and disease risk can now be satisfactorily met using modern epidemiologic approaches for the design and analysis of cohort studies. In parallel with the increasing design sophistication of many cohort studies, new biostatistical methods now make possible longitudinal analyses that can incorporate temporally-varying exposures. Application of these analytic methods has been facilitated by the availability of hardware and software, which make possible analyses that could not have been contemplated one or two decades previously.

In spite of the central role of the cohort design in epidemiologic research, it has been the focus of few monographs. The book by Breslow and Day (2), published in 1987, represents a pioneering synthesis on design and analysis of cohort studies on cancer. A number of statistical texts address analysis of longitudinal data (3–5). The history of the cohort study was comprehensively addressed in a 1988 review by

Received for publication December 29, 1997, and accepted for publication June 12, 1998.

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

From the Department of Epidemiology, School of Hygiene and Public Health, The Johns Hopkins University, Baltimore, MD.

Reprint requests to Dr. Jonathan M. Samet, Professor and Chairman, Department of Epidemiology, School of Hygiene and Public Health, The Johns Hopkins University, 615 North Wolfe Street, Baltimore, MD 21205.

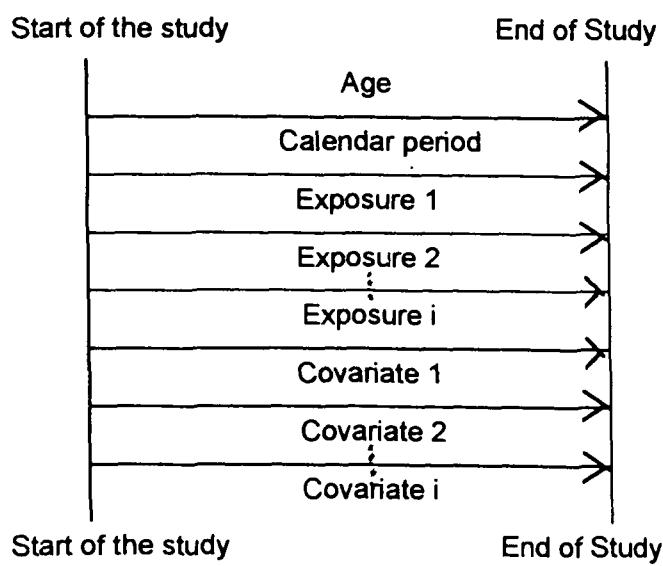


FIGURE 1. The multiple dimensions of time in a cohort study.

Liddell (6) and in papers published from a 1983 American Cancer Society workshop on cohort studies (7). In this presentation, which introduces this special volume of *Epidemiologic Reviews* on cohort studies, we first discuss the terminology, the definitions, and the evolution of the cohort design. We then review the analytical issues in cohort studies, outlining the overall objectives of analytical methods, followed by an historical perspective on their evolution. We close the paper by illustrating the power of the cohort design in epidemiology: the selected study has been key in providing comprehensive data on the epidemiology of the acquired immunodeficiency syndrome (AIDS).

TERMINOLOGY AND DEFINITIONS

While "cohort study" has been variably defined, all definitions incorporate the concept that nonexposed and exposed, or variably exposed, individuals are observed over time for the outcome(s) of interest. The word "cohort" has its origin in the Latin *cohors*, referring to warriors and the notion of a group of persons proceeding together in time. In the past, various terms, including prospective, follow-up, and longitudinal, were sometimes used to reflect the temporal sequence of exposure and disease in study participants, compared with the retrospective sequence of case-control studies (8, 9). These terms have largely been abandoned in the epidemiologic literature, and the term "cohort study" is used most often.

Cohort studies are also designated by the timing of data collection, either prospectively or retrospectively, in the investigator's time. Studies collecting data on events that have already occurred have been labeled as

historical, retrospective, and nonconcurrent. The 1995 edition of *A Dictionary of Epidemiology* (10) offers a definition for "historical cohort study" and the synonyms of "historical prospective study, nonconcurrent prospective study, and prospective study in retrospect." To describe these types of studies, the labels retrospective cohort study and nonconcurrent cohort study are widely used at present. Among epidemiologists, there is now close to uniform use of the term prospective cohort study to refer to studies in which the investigators observe the events as they occur in time. Any lingering debate with regard to terminology should be set aside to avoid unneeded confusion.

EVOLUTION OF THE COHORT DESIGN

The origins of the cohort study can be traced to the need for information on the length of life and the course of disease—such information being central in medical practice and public health and of fundamental interest to the general population. The earliest life tables, developed by Graunt and Halley from cross-sectional mortality data, were intended to project deaths with aging, inherently acknowledging the passage of time. In the 19th century, Farr advanced the use of life tables as an indicator of population health (11). The emergence of the modern insurance industry several centuries ago, created the still-existent profession of the actuarialist whose job is to project risks (12, 13). The experience of policyholders provided data for this purpose, and over a century ago insurance companies pooled data to gain more precise and specific descriptions of mortality among the insured. A report published in 1904, the *Specialized Mortality Investigation*, pooled the experience of policyholders in 34 of the largest companies in the United States and Canada for the years 1870 through 1899. The report gives tables listing 141,977 deaths along with expected numbers by classes of impairment, medical history, and occupation (14, 15). Singer (14) documents other early studies carried out by the insurance industry, directed, for example, at asthma and mortality, hypertension and mortality, and obesity and mortality.

At the turn of the century, tuberculosis was the leading cause of mortality in the United States. Treatment was accomplished in sanatoriums with a variety of therapies; exposure to outdoor air and sunshine was central in the management of the disease. Follow-up studies, using the design now known as the prospective cohort study, were carried out to describe the natural history of the disease in tuberculosis patients and the consequences of the therapies then in use (16, 17). In one of the earliest of these studies of tuberculosis, Brown and Pope (18) traced over 1,000 persons

discharged from the Adirondack Cottage Sanitarium (later to become the famed Trudeau Sanitarium). Brown and Pope used life-table methods and compared patient survival to Farr's English Life Table no. 3. Sartwell (16) noted that this study was the first application of the life-table method in a clinical follow-up study. A similar study was reported in 1910 by Elderton and Perry (19) who described mortality among patients discharged from two sanatoriums in England, calculating observed and expected deaths based on one of Farr's life tables.

Wade Hampton Frost made several methodological advances related to cohort studies in his investigations of tuberculosis. He has been credited with development of cohort analysis of vital statistics data, that is, the separation of age, period, and cohort effects by stratification of vital statistics or other data on these time-varying factors. This development arose from his report on age- and birth-cohort specific patterns of tuberculosis mortality in Massachusetts for the years 1880 through 1930 (20). In 1930, Andvord (21) had published a similar but overlooked analysis.

In describing the risk of tuberculosis in 132 black families in Kingsport, Tennessee, Frost (22) pioneered the use of the retrospective cohort design. He interviewed family members to reconstruct the household composition from the time of establishment and then retrospectively followed the household members to the present. Frost estimated person-years of "life-experience" and calculated the age-specific attack rates for tuberculosis. He also initiated a prospective cohort study, the Williamson County Tuberculosis Study; the study was started in 1931 and follow-up continued through 1955, long after Frost's death (23).

The more contemporary era of cohort research can be traced to the late 1940s and early 1950s when several landmark prospective cohort studies were implemented. Some of these studies continue today: the Framingham Study, the study of the Japanese atomic bomb survivors, the study of British physicians, and the Public Health Service study of Colorado Plateau uranium miners. These studies were distinguished by their size and the richness of the data collected and by the sustained follow-up of participants over decades. All were implemented to address pressing public health concerns: the causes of heart disease, the consequences of smoking, and the risk of radiation.

Dawber (24) has chronicled the origins of the Framingham Study, which was implemented in the late 1940s to address the rising occurrence of cardiovascular disease. The long-term success of the study can be attributed to the selection of a small and cooperative community, sustained support from the National Institutes of Health which maintained the study as an

intramural project, and to the prescience of the original investigators who established rigorous and standardized protocols for data collection. Data were collected relevant to testing the principal extant hypotheses concerning etiology, which were listed at the study's beginning. As a result, much of our initial understanding of risk factors for cardiovascular diseases was based on evidence from this study. Supplementary studies of other diseases capitalized on the opportunity afforded by having the Framingham population under follow-up, and offspring of the original cohort have now been enrolled in a new cohort study that should be informative on familial factors affecting cardiovascular disease risk. The longitudinal data on multiple risk factors necessitated methodological advances, as appropriate multivariate methods had not been available. For example, Truett and colleagues (25) described application of discriminant analysis in a 1967 paper. They predicted 12-year probability of developing coronary heart disease based on levels of seven risk factors.

Another landmark cohort study, the investigation of the atomic bomb survivors in Hiroshima and Nagasaki, addressed the consequences of ionizing radiation exposure. This population, a group having a unique and instantaneous exposure, contrasts with the Framingham Study, a general population study of several diseases that tests multiple hypotheses. The two bombs were dropped in 1945, and investigation of the medical consequences began almost immediately thereafter (26). By 1946, a decision had been made to conduct long-term studies, and the Atomic Bomb Casualty Commission was established in 1947. In 1975, this organization was replaced by the Radiation Effects Research Foundation, still operating, which is funded jointly by the United States and Japan. Acute and chronic effects of radiation exposure were already known but there was little quantitative information available on the risks of radiation. Radiation doses from the blasts were eventually reconstructed and selected survivors were entered into a cohort study that included periodic medical examinations.

This study has become one of the principal sources of evidence on the cancer risks of acute radiation exposure; its findings are the underpinnings of radiation standards throughout the world. Soon after the blast, the occurrence of acute leukemia rose in the survivors, but after peaking around 1952, the excess began to decline. By 1960, excesses of solid tumors were noted and current studies by the Radiation Effects Research Foundation emphasize risks of these cancers. In addition to the study's contributions to the evidence on radiation risks, the challenges of the longitudinal data have prompted substantial methodolog-

ical work directed at such issues as the time- and age-dependence of radiation risk, the joint effect of radiation with other factors, and the consequences of measurement error (27).

In the early 1950s, another key study of a radiation-exposed special population was implemented: the US Public Health Service study of Colorado Plateau uranium miners (28). The participants comprised underground uranium miners exposed to radon progeny released from the uranium ore. Over 3,000 men were enrolled in this study by 1960, and follow-up, as of 1998, is continuing. Unlike the single exposure of the atomic bomb survivors, the uranium miners received continuing exposure throughout employment at a rate that depended on mining conditions. Consequently, both the cumulative exposures and the rates of exposure varied over follow-up. This study has been significant for its definitive demonstration of the cancer risk of radon progeny and for the methodological advances it has fostered on time-varying aspects of exposure and exposure rate (29). The dataset also includes information on smoking, and methods have been developed and applied for characterizing the joint effects of the two causes of lung cancer—radon and smoking (30–32).

A number of now well-known prospective cohort studies were implemented during the 1950s in follow-up of the initial observations from case-control studies of a remarkably strong association between cigarette smoking and lung cancer. In starting the study of British physicians, Doll and Hill (33) commented on the need for more credible, prospectively collected data in follow-up of the mistrusted retrospective data of the case-control studies. The prospective cohort study of British physicians, initiated in 1951, continues as of 1998. A 1994 report (34) provides the findings after 40 years of follow-up. The success of the study reflects the investigators' foresight in selecting a cooperative population that could be readily followed for mortality, and in using a simple, mailed questionnaire to periodically assess smoking by the participants. Other key investigations on smoking included the American Cancer Society's Nine-State Study of approximately 188,000 persons (35) and the study of US veterans (36). By 1964, findings from eight prospective cohort studies on smoking and disease were available for review by the Surgeon General's advisory committee (37).

Case and colleagues (38) reported the prototype retrospective cohort study in 1954. Following up on the hypothesis that aniline-based dyes increased risk for bladder cancer, Case et al. developed a roster of exposed workers in the United Kingdom from 1920 forward and identified bladder cancer cases and

deaths. This study showed the feasibility of retrospective cohort studies when the needed records are available; this design was soon to be widely used for worker groups for whom records documenting employment and exposures were available.

Many cohort studies, both prospective and retrospective in timing, were implemented widely beginning in the 1950s and 1960s. These early studies often proved to be landmarks for particular diseases and exposures: Doll's study (39) of workers in an asbestos textile factory, the study of Selikoff et al. (40) of US insulation workers, the American Cancer Society's Cancer Prevention Study I (41), and the study of Fletcher et al. (42) of lung function in workers in London are examples. In conducting the Cancer Prevention Study I, the American Cancer Society used volunteers to enroll one million participants, establishing feasibility and demonstrating the strength of evidence from large cohort studies. Further general population studies of multiple chronic diseases, following the model of Framingham, were implemented in Tecumseh, Michigan (43); Evans County, Georgia (44); Alameda County, California (45); and Washington County, Maryland (46). In the Washington County study, serum was stored so that serologic markers could be examined as predictors of disease risk; this study model represents an early application of the research approach now referred to as "molecular epidemiology."

The current era of large, focused cohort studies dates to the 1970s. With deepening understanding of risk factors for chronic disease, cohort studies were typically designed to collect extensive information on exposures and to follow participants rigorously for outcomes. In the Tucson, Arizona, study of respiratory diseases, for example, participants visited a central clinic annually for an evaluation that included questionnaires, lung function testing, skin testing, phlebotomy, and other evaluations (47). In the United States, the National Institutes of Health has taken the lead in establishing multicenter prospective cohort studies, particularly in the area of cardiovascular disease—for example, the Atherosclerosis Risk in Communities Study, (48), the Cardiovascular Health Study (49), and the Strong Heart Study (50). These multisite studies gain external validity by drawing participants from communities across the United States. Data collection is standardized and data are accumulated, evaluated, and managed at central coordinating centers.

Opportunities for data linkage have now facilitated the conduct of cohort studies. Using record linkage approaches, researchers can match lists of exposed individuals for outcome against death indexes and disease registries. Pioneering cohort studies based on

this approach were conducted in Canada, where a mortality register of deaths back to 1950 has been available for matching and establishing vital status and cause of death (51). The National Health and Nutrition Examination Survey, conducted by the National Center for Health Statistics, has been given a longitudinal component by linkage against death certificates and additional follow-up data collection (52).

Contemporary reminders of the strength of evidence from cohort studies are abundant. The Nurses' Health Study, started in 1976 to investigate risks of oral contraceptives, has become one of the principal sources of observational data on diet and disease (53). Drawing on the lessons learned in the study of British physicians, the Nurses' Health Study incorporates a cooperative participant group familiar with completing questionnaires, a mailed approach for data collection, and active and passive follow-up for outcome. The investigators have even been able to collect biologic specimens from participants.

Shortly after the recognition of the disease now referred to as AIDS, plans were made for a cohort study to characterize the natural history of the disease and the determinants of prognosis. The Multicenter AIDS Cohort Study, implemented in 1984, comprised a cohort of close to 5,000 homosexual men in four cities (54). The implementation of the study antedated the identification of the causal virus, human immunodeficiency virus 1 (HIV-1), but the study included the collection of complete clinical data and the storage of blood specimens every 6 months. Over time, this repository of information and specimens has been repeatedly used to address determinants of risk for AIDS and prognostic factors. For example, a 1997 publication addresses viral load and the course of disease, an application unanticipated at the study's beginnings (55). In parallel to the evolution of cohort studies for cardiovascular diseases mentioned above, in 1994 a major cohort study on women, the Women's Interagency HIV Study, was assembled to characterize the natural history of HIV so that comparison and gender-specific inferences could be reached.

Methodological advances have been motivated by the complexity of analyzing data from cohort studies. Many of these advances, reviewed below, have come from collaborations between epidemiologists and biostatisticians, as they addressed challenging longitudinal data generated by cohort studies. A key advance in design was the development of sampling methods for efficiently assessing the relation between exposure and outcome. These designs are particularly valuable for relatively infrequent outcomes. The nested case-control study compares exposures of cases of the disease of interest with controls drawn from the remainder of

the cohort at the time the case developed the disease (56, 57). In the case-cohort design, covariate data are fully developed for cases and for a random sample of the full cohort, drawn at the start of the study (58). When properly analyzed, these designs yield unbiased estimates of the relative risk, even though the full suite of covariate data needs to be developed only for a sample of the participants.

Cohort studies, of course, have potential limitations. The use of the retrospective design is possible only if historical data of adequate quality are available. The prospective design is successful only if adequate follow-up of participants can be maintained. Repeated data collection, often warranted for scientific purposes, may be constrained by feasibility concerns, costs, and participant burden. Bias that is differential over time may complicate the interpretation of findings of a cohort study. Information bias may vary in its effect over the cause of data collection due to the sometimes subtle drifting of the quality of data collection. Apparent time-dependent effects may occur as a result. Selection bias may also be differential over time from losses to follow-up (59). To date, there has been little systematic consideration of the time-specific biases that may affect cohort studies.

OBJECTIVES OF ANALYTICAL METHODS FOR COHORT STUDIES

The analysis of data collected in cohort studies is determined by 1) the specification of the substantive question to be answered; 2) the consideration of the nature of the outcome or measure of disease occurrence (e.g., time-to-event, change in marker measured repeatedly); and 3) the nature of the exposures and covariates of concern and their relations with the outcome. An analytical model needs to be selected that is appropriate for the substantive question and for the nature of the outcome data, including the specific parameters quantifying the association between exposure and disease. Once a model is selected, the analyst uses statistical procedures for the extraction of information contained in the data at hand. Methods based on the likelihood principle are widely used for this purpose. Likelihood-based procedures estimate the unknown parameters with the values that make the data at hand the most likely to have been observed (i.e., maximum likelihood principle); they also permit the determination of how deviant the likelihood under the estimated parameters is from the likelihood of the data under the assumption that no relation exists between exposure and disease (60). A large deviance between the maximum likelihood based on the data and the likelihood under the null assumption of no exposure/disease association casts doubt on this null hypothesis.³³

p-Values have been widely used in epidemiology to quantify how unlikely the data observed are under the null hypothesis. The more unlikely, then the stronger is the doubt as to the validity of the null hypothesis. Most of the regression methods reviewed here are based on the maximum likelihood principle or an extension, e.g., quasi-likelihood.

Data for a cohort study include, as a minimum, the follow-up experience of each participant and the status of each participant with regard to the occurrence of the event(s) of interest. Even the most basic cohort data include multiple dimensions of time: calendar time, age, and time on study. Exposures may vary over time as may confounding and modifying factors (figure 1). The contemporary study may thus include multiple time-dependent variables, and even the outcome measure may vary over time or occur multiple times, as in the example of some infections. The analytical challenges of such data have been addressed by innovative biostatistical methods developed mostly during the last two decades.

Common to other epidemiologic study designs, the primary objectives of the analysis of data from cohort studies are 1) to summarize and 2) to compare. One fundamental measure of disease occurrence in a cohort study is the incidence rate. We summarize main features of cohort data by graphic and tabular displays of measures of location and scale. Summary measures of location include the mean and median, and summary measures of scale include the standard deviation and interquartile range, which may be summarized within exposure groups and compared across groups. The procedures that are linked to proper summarization are the extensive methods developed under the rubric of estimation in the statistical literature (61). Another useful, although frequently overlooked, procedure for summarizing is the graphic depiction of data, a method now strengthened by current hardware and software (62).

Comparison of the frequency of disease occurrence in exposed and unexposed individuals is a primary objective of cohort studies. An underlying and fundamental assumption for valid epidemiologic inference is that the exposed and unexposed groups are comparable with respect to any other factors that may explain the heterogeneity of disease occurrence and that are related to the exposure and the disease. In clinical trials, randomization leads to comparability, but in cohort (observational) studies, comparability needs to be achieved by design or in analysis by stratification and/or regression. Advances in statistical methods have greatly augmented our capacity to meet this primary analytic goal while fully considering the temporal structure of data.³⁴

EVOLUTION OF METHODS FOR THE ANALYSIS OF COHORT STUDIES

In this section, we review the advances in analytical methods for the basic outcome measures in cohort studies: time-to-event and repeated measures of markers of disease progression. The past 20 years have witnessed the development and wide application of multivariate methods for the elucidation of factors explaining the variability of hazard of disease (i.e., survival analysis), and of trajectories of markers of disease progression measured repeatedly over time (i.e., longitudinal data analysis). For studies using time-to-event as the primary outcome, the longitudinal data on markers are treated as covariates, with a typical substantive question being the distribution of event-free times based on marker values (e.g., AIDS-free times according to amount of HIV in blood and level of immune deficiency (55). Conversely, for studies using change of markers as the primary outcome, the time-to-event is treated as a covariate, with a typical substantive question being the effect that the occurrence of an event (disease) has on the trajectory of a marker (e.g., loss of homeostasis of total T-cell count with the imminent onset of AIDS (63)). The close interrelations between time-to-event and repeated measures of markers in cohort studies have opened an active area of current research to combine the two sets of information into a unified framework.

Prior to the 1970s, analyses of cohort data were based primarily in life-table methods and stratified approaches for handling confounding and evaluating effect modification. Binary variables were the principal outcome measures of concern. The methods first applied to cohort data for multivariate analyses, discriminant analysis, and logistic regression, while appropriate for binary data, did not explicitly incorporate time. These methods, now known to be more appropriate for data that are cross-sectional in time, are not considered further in this review.

In 1972, D. R. Cox published a seminal paper on regression methods for time-to-event data (64), providing the basis for what is now widely known as proportional hazard regression models in survival analysis. This method has the strength of needing no assumption as to the form of the hazard of disease in the unexposed reference group. The hazards of other groups under different exposures are modeled as multiples of the hazard (relative hazards) of the reference group. Measures of relative hazards between groups at all times at which events occur are combined into an overall estimate of the relative hazard (3).

The 1970s also witnessed the full development of Poisson regression methods for the analysis of events-in-person-years data (2). These methods are particu-

larly useful for the analysis of trends and changes in incidence of disease over calendar time; they are of great utility for data in which a specific time origin is not well defined or not of interest. These methods are suitable for data obtained in cohort studies, which provide the number of events and the person-years at risk for the event(s) of interest. Cox and Poisson regression methods are closely linked and, in most cases, they give similar results.

The decade of the 1980s witnessed the development of methods for the analysis of markers of disease progression observed repeatedly for participants in cohort studies (e.g., forced expiratory volume, a lung function measure, in cohort studies of respiratory diseases; blood pressure in cohort studies of cardiovascular disease; and CD4 cell count in cohort studies of infectious diseases). The methods, developed for the analysis of levels of markers over time and of trajectories of change, are now widely used and defined as the methods for the analysis of longitudinal data (5).

Methods for the analysis of longitudinal data can be broadly classified into three groups: marginal, transition, and random effects models which address distinct epidemiologic questions. The marginal approach combines the multiple cross-sections corresponding to data collected at cohort study visits to provide the most efficient summary of the relations between prevalence of disease (binary outcome) or other mean response and the prevalence of exposure. In this approach, the longitudinal element is typically incorporated by including age or time since baseline as a covariate in a regression model. Approaches for the incorporation of the correlation between repeated measurements within individuals include parametric (65) and non-parametric methods, the latter handling the correlation as a nuisance (66).

Transition models regress current outcome on past values of the outcome, and on current and previous exposures. Classic Markovian models for binary outcome data were introduced to epidemiologists in 1979 (67), and applied, for example, in 1980, to a study of air pollution and asthma (68). Extensions for the continuous outcome were used for the study of the effect of cigarette smoking on respiratory function (69).

Random effects models allow each individual to have unique regression parameters (e.g., intercept and slope) according to components of variance, and provide direct averages of rates of change across individuals. Methods for random effects models have been provided for Gaussian outcomes (70), binary outcomes (71), and for event-in-person-years outcomes (72).

Another advance in methods for cohort studies during the 1980s was the extension of regression trees methodology to survival data (73, 74). Regression

trees are extremely flexible for handling interactions and are very effective for communicating epidemiologic inferences to wide audiences. The primary concern is that their flexibility may result in inferences that are too specific to the data at hand and, therefore, are of limited generalizability.

The first quinquennium of the 1990s witnessed the dissemination of a unified framework for linear models (75) under which linear, logistic, Poisson, and many survival regression models could be viewed as specific cases of generalized linear models. Extension of this framework also allowed for relaxation of assumptions and triggered the development of quasi-likelihood methods. These new methods can account for different variance structures and can handle nuisance correlations using robust methods for appropriately estimating standard errors. In parallel to these advances, graphic procedures have been substantially improved with the availability of smoothing algorithms and the development of additive models that free regression models from the usual linear assumptions (76). These generalized additive models are especially useful for summarizing data but are somewhat limited for comparison and determination of measures of differences in a probabilistic framework.

During the early 1990s, methods were also developed that allowed for late or staggered entries into observation in a cohort study. Using these methods, the analyst can select the most appropriate time scale from a biomedical perspective, not only time-on-study, and can describe the occurrence of disease in person-years and thereby address incidence—the fundamental measure of disease—directly. The unit of analysis becomes individual-periods-at-risk, as opposed to the individual.

These new methods for late entries have been used for juxtaposition of incident and prevalent cohorts (77), and analysis of time-varying exposures whereby follow-up time is partitioned into as many individual periods as changes in exposure are recorded (78). They control for the survival bias that can be introduced by classifying persons as never or ever exposed. They also facilitate proper inferences regarding intermediate events that actually increase the hazard of the event but that may appear protective under improper analysis because the intermediate events occur only after some period of time. In another application, the methods for late entries can be used in evaluating effectiveness of therapies over time by considering calendar time as an external time-dependent exposure. The hazards can be compared for individuals who reach the same duration of time at risk in different calendar periods (79).

Beginning in the 1980s, Cox regression has been

widely used to estimate the relative hazard. It was initially applied to cancer clinical trials which have the objective of assessing efficacy of therapies in the setting of a high underlying hazard and, consequently, there was little interest in describing the underlying hazard. In this context, proportional hazard methods were ideal because the underlying hazard is allowed to be arbitrary. However, in cohort studies the underlying hazard itself is generally of interest, particularly in studies describing the natural history of diseases; for example, the hazard of AIDS at different intervals since infection with HIV. Regression under parametric models (e.g., lognormal) provides direct measures of the underlying hazard and, more importantly, of relative percentiles (78, 80, 81). Relative percentiles or relative times compare exposure groups according to the ratio of the times over which a given percent of individuals in the groups under different exposures develop the disease. These methods are also consonant with the renewed interest in quantifying the disease-free years at the population level that an intervention may produce (79).

The use of nested designs and the development of related analytic techniques is another major advance. The primary objective of nesting substudies within a cohort is to use all the cases of interest but only a subsample of the noncases so that validity is not compromised and precision is adequate. The two most widely used approaches are the nested case-control and case-cohort substudies; the primary distinction between the two is the timing of the selection of a comparison group from the noncases. Analytical methods for the analysis of nested studies are readily available; namely, conditional logistic regression for nested case-control studies and Cox regression with staggered entries and robust methods for calculation of standard errors for nested case-cohort studies (82). An

alternative method for nesting studies is based on trajectories of markers of disease progression (i.e., stable versus fast progressors) (83).

Table 1 provides a summary of the analytical methods for cohort studies reviewed here. Software is widely available for the implementation of different methods, and several statistical packages, including SAS (SAS Institute, Cary, North Carolina), Splus (Statistical Sciences, Inc., Seattle, Washington), STATA (Stata Corporation, College Station, Texas), and EGRET (Cytel Software Corporation, Cambridge, Massachusetts), provide procedures, functions, and commands to carry out analysis of data from cohort studies. For example, analysis of cohort studies with staggered entries can be equally accomplished by the PROC PHREG of SAS; the survfit, coxph and Surv functions of Splus; the stset with the t0 option and stcox functions of STATA; and the menu-driven options of the Kaplan-Meier and Cox regression modules of EGRET (84).

An important advance in the last 10 years has been the development of methods to incorporate measurement error into the analysis of cohort data (85, 86). The application of these methods requires the appropriate design of validation/reproducibility substudies within cohort studies. Unfortunately, in many cases these studies are not properly designed or their results are not properly incorporated in the analysis of the core questions in cohort studies. These methods hold promise as a partial solution to the persistent problem of measurement error.

In this section we have reviewed the advances in analytical methods for time-to-event and repeated measures of markers as two separate fields. An area of active methodological research since the mid 1990s has been the unification of time-to-event (survival analysis) and repeated measurement methods (longi-

TABLE 1. Overview of analytical methods for cohort studies

Outcome	Summary measure	Comparison		Measure of association
		Exposed/unexposed (2-sample)	Multiple (regression)	
Events in person-years	Incidence rate	(O-E) ² /var	Poisson	Relative incidence
Time to event	Kaplan-Meier/maximum likelihood estimates	Logrank or Mantel-Haenszel/likelihood ratio test	Proportional hazards/parametric	Relative hazard/relative percentile or time
Time to event; exposures changing	Extended Kaplan-Meier	Extended logrank	Proportional hazards, staggered entries	Relative hazard
Case in nested case-control	Proportion exposed	Paired chi-square or McNemar	Conditional logistic	Odds ratio
Case in nested case-cohort	Proportion exposed	(Robust) logrank	Proportional hazards, staggered entries	Relative hazard
Intermediate outcome repeatedly measured	Change		Regression for correlated data; marginal, conditional, random effects	Differences in change over time

tudinal data analysis). Both types of data have incompleteness or missing data, and the nature of incompleteness is often informative (e.g., individuals with low values in markers cease to provide longitudinal data due to imminence of disease onset). The developments for the handling of missing data have generated promising approaches for a unified framework (87–92). These methods are certain to advance.

AN EXAMPLE: NATURAL HISTORY OF HIV INFECTION AND THE MULTICENTER AIDS COHORT STUDY

In this section, we illustrate the application of analytical methods for cohort studies in the context of the natural history of HIV infection, drawing on the Multicenter Aids Cohort Study. It is included in this review as an example of the great utility of cohort studies contributed by epidemiologic research to the overall goals of science and public health. There are numerous comparable examples in cohort studies of cancer, cardiovascular disease, and occupational and environmental agents.

Figure 2 depicts the key events in the natural history of HIV infection that take place during a cohort study of an at-risk population. On enrollment, some individuals would enter with antibodies to HIV (i.e., seroprevalent), and of those who enter seronegative, some would become infected during follow-up (i.e., sero-

converters). Both seroprevalent individuals and seroconverters are subject to progressive immune suppression and opportunistic infections as a consequence of infection with HIV. These key natural history events shown can be linked to specific epidemiologic aims for which different analytical methods are pertinent (figure 2, table 2). Although the table is specific to the natural history of HIV infection as elucidated by a particular cohort study, it illustrates the scope of analytic endpoints in a cohort study and gives examples of the methods used for the different types of data pertinent to specific scientific aims.

In 1984–1985, a cohort of 4,954 men was recruited into the Multicenter AIDS Cohort Study in Baltimore, Maryland, Chicago, Illinois, Los Angeles, California, and Pittsburgh, Pennsylvania. To increase minority enrollment, an additional 625 men were recruited from 1987–1991, of whom 433 (69.3 percent) were non-Caucasian, and an additional 43 seroconverters from Pittsburgh were also recruited at the same time. The entire Multicenter AIDS Cohort Study cohort, therefore, consists of 5,622 men, of whom 2,195 (39 percent) were seroprevalent for HIV at entry. All men were followed up every 6 months, and serologic tests for HIV antibody were routinely done at each visit. Up to July 1, 1997, 551 men had known dates of last negative and first positive visits for HIV (i.e., seroconverters). Through July 1997, 1,400 and 244 AIDS

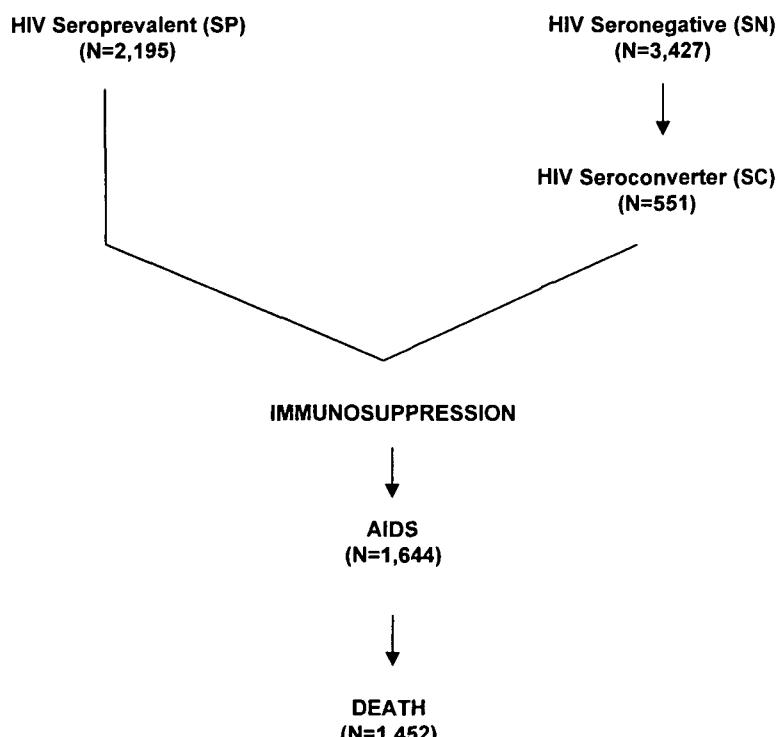


FIGURE 2. Descriptive statistics of participants in the Multicenter AIDS Cohort Study, 1984–1997.

TABLE 2. Analytical methods for cohort studies; epidemiology of human immunodeficiency virus (HIV) infection in gay men, Multicenter AIDS* Cohort Study, 1984–1997

Aim	Method	Reference(s)
Prevalence of infection (SP*)	Logistic regression	Chmiel et al. (93)
Incidence of infection (SC*)	Poisson regression	Kingsley et al. (94)
Immunosuppression (CD4+)	Regression for continuous correlated data	Margolick et al. (95) Mellors et al. (55)
Incubation (SC to AIDS)	Lognormal regression Juxta-analysis: incident plus prevalent subcohorts	Muñoz and Xu (78) Muñoz et al. (77, 97, 98) Taylor et al. (99)
Markers and disease progression (Viral load, CD4 to AIDS and death)	Cox regression Regression trees	Fahey et al. (100) Enger et al. (101) Mellors et al. (55)
Survival (AIDS to death)	Cox regression	Jacobson et al. (102)
Incidence of specific AIDS diagnosis by period	Poisson regression	Muñoz et al. (103)
Risk factors for fast progression to AIDS	Conditional logistic regression	Phair et al. (104)
High risk behavior	Regression for categorical correlated data	Gange et al. (105)
Effectiveness of AIDS therapies	Cox regression with staggered entries	Muñoz and Hoover (106) Detels et al. (79)
Long-term survivors and pathogenesis of HIV	Nested studies based on markers	Muñoz et al. (107) Gange et al. (108)

* SP, seroprevalent; SC, seroconverters; AIDS, acquired immunodeficiency syndrome.

cases had been observed among the seroprevalent individuals and seroconverters, respectively. Among the 1,644 AIDS cases in HIV-positive individuals, 1,452 have died. Table 2 provides references based on the Multicenter AIDS Cohort Study data using specific analytical methods for corresponding scientific aims.

In an initial cross-sectional analysis, Chmiel et al. (93) used logistic regression to relate the odds of being seroprevalent for HIV to a constellation of putative risk factors, including sexual behavior, demographic characteristics, and history of other infectious diseases. It is noteworthy to point out that the serostatus at baseline in 1983–1984 could only be assessed *retrospectively*, since the test for HIV first became available in late 1985 and testing was performed on blood samples kept in a national repository. This example shows the great utility of storing samples collected *prospectively* in cohort studies.

To describe the incidence of infection, Kingsley et al. (94) modeled the number of seroconversions as a Poisson variable in strata defined by calendar, age, and ethnicity. At the time of publication in 1991, there was a suggestion of rising incidence over time, but subsequent follow-up did not confirm a trend and the rates of seroconversion remained low up to 1995, when follow-up of seronegative participants was confined to

a small portion of the full cohort of seronegative participants.

The longitudinal data collected on CD4-cell count, a marker of immunodeficiency caused by infection with HIV, were used to characterize CD4 trajectories of individual participants. Margolick et al. (95), using the simplest case of random effects models (i.e., random intercept), documented that the total T-cell count remains fairly constant during the course of HIV infection up to approximately 1–5 years prior to the occurrence of AIDS, when it declines precipitously. These observations led to the postulate that homeostasis of the total T-cell count fails prior to the onset of AIDS. Mellors et al. (55) used random regression (i.e., intercept and slope follow a bivariate normal distribution) to show the very close relation between viral load at one time point and subsequent decline of CD4 cell count. These data were used to define the principles of HIV therapy now endorsed by the United States Public Health Service (96).

As for all infectious diseases, the incubation period of AIDS is of central interest. To describe the incubation period of AIDS and the corresponding hazard of AIDS at different durations of infection, Muñoz and Xu (78) showed that the lognormal model and regression were appropriate. This and several other reports

(97–99) on the AIDS incubation period in Multicenter AIDS Cohort Study participants combine data from incident and prevalent cases. These data can be combined by using an extension of Kaplan-Meier and Cox regression methods so as to consider at risk for AIDS only those individuals who have already entered into observation. Since this actually corresponds to a juxtaposition of the two subcohorts, the authors have suggested the name *juxta-analysis* for this type of approach (77).

Cox regression methods have been widely used to describe predictors of disease progression, generally measured by the length of AIDS-free survival time. In a seminal report, Fahey et al. (100) evaluated the prognostic value of three cellular and five serologic markers that are affected by infection with human immunodeficiency virus. As new methods were developed in 1995 to reproducibly quantify plasma viral load, Multicenter AIDS Cohort Study investigators used samples stored in the third semiannual visit (around September 1985) to assess viral load and natural history; Mellors et al. (55) showed that plasma viral load was the single best predictor of progression to AIDS and death; and using regression trees methodology (73, 74), the investigators showed that the prognosis of HIV-infected persons is more accurately defined by combined measurements of plasma HIV RNA and of CD4 lymphocytes. Enger et al. (101) estimated the expected survival time by calendar period before (1985–1988) and after (1989–1993) the widespread availability of AIDS treatments, and by stage of HIV disease quantified by the CD4 cell count at the beginning of each of the periods. In addition, Jacobson et al. (102) documented the changes in survival after AIDS in periods covering the years between 1984 and 1991.

Cohort studies have the substantial advantage of describing the incidence of different outcomes of interest. Muñoz et al. (103) used Poisson regression methods to describe the incidence of six groupings of the conditions that define AIDS. The investigators documented the effectiveness of *Pneumocystis carinii* pneumonia prophylaxis, showing a significant decline of the incidence of *P. carinii* pneumonia during follow-up of Multicenter AIDS Cohort Study participants. This decline was concomitant with upward trends of other opportunistic infections.

Phair et al. (104) nested a case-control study within the Multicenter AIDS Cohort Study to explore factors that may identify seroconverters who rapidly progress to AIDS. Consonant with the matched design of the study, they used conditional logistic regression to analyze the data, finding that high-risk behavior prior to seroconversion was related not only to the risk of

infection but also to the risk of fast progression after infection. High-risk behavior is of interest as an outcome itself. Analytical challenges posed by behavioral data over time have led to methodological developments for regression methods for categorical correlated data (105).

Methods for cohort studies with staggered entries are also useful for the evaluation of effectiveness of AIDS therapies. Muñoz and Hoover (106) and Detels et al. (79) have analyzed the Multicenter AIDS Cohort Study data to determine if therapies as used by participants increase disease-free periods and/or survival. These analyses use calendar period as an external time-dependent covariate and as a proxy measure of relative intensity of exposure to antiretroviral therapies.

Based on the trajectories of markers of disease progression, cohort studies offer the possibility of comparing subgroups of individuals who exhibit different trajectories of markers in spite of starting at the same level (107). Furthermore, individuals who exhibit a stable profile during the first part of a cohort study could subsequently exhibit heterogeneous progression toward disease. These data offer the possibility of comparing late progressors with consistent nonprogressors, using a case-control study, in which cases and controls are matched longitudinally, thus presenting a hybrid of the case-control and cohort designs (108).

EPILOGUE

We have reviewed the evolution of the cohort study from the substantive and methodological perspectives, and have illustrated how different methods have been useful for the key epidemiologic aims in cohort studies, using studies of HIV infection as our example. Because of their longitudinal nature, cohort studies offer an invaluable resource for the elucidation of disease pathogenesis. Cohort studies have provided fundamental knowledge for prevention strategies and have been a cornerstone of public health and policy. Methodological advances continue to strengthen this design and facilitate our understanding of how multiple factors acting over time can determine the etiology and natural history of disease.

ACKNOWLEDGMENTS

This research was supported in part by grants UO1-AI-35043 and UO1-AI-42590 from the National Institutes of Health. The authors dedicate this review to B. Frank Polk, a mutual friend and colleague, whose contributions were seminal to the creation of the Multicenter AIDS Cohort Study.³⁹

REFERENCES

- Armenian HK (ed). Applications of the case-control method. *Epidemiol Rev* 1994;16:1-164.
- Breslow NE, Day NE. Statistical methods in cancer research. Vol II—The design and analysis of cohort studies. Lyon, France: International Agency for Research on Cancer, 1987. (IARC scientific publications no. 82).
- Cox DR, Oakes D. Analysis of survival data. New York, NY: Chapman and Hall, 1984.
- Dwyer JH, Feinleib M, Lippert P, et al. Statistical models for longitudinal studies of health. New York, NY: Oxford University Press, 1992.
- Diggle PJ, Liang KY, Zeger SL. Analysis of longitudinal data. New York, NY: Oxford University Press, 1994.
- Liddell FDK. The development of cohort studies in epidemiology: a review. *J Clin Epidemiol* 1988;41:1217-37.
- Changes in cigarette-related disease risks and their implication for prevention and control. Bethesda, MD: US Department of Health and Human Services, National Institutes of Health, National Cancer Institute, 1997. (NIH publication no. 97-4213).
- Bailar JC III, Mosteller F. Medical uses of statistics. 2nd ed. Waltham, MA: NEJM Books, 1992.
- MacMahon B, Trichopoulos D. Epidemiology: principles and methods. 2nd ed. Boston, MA: Little, Brown, 1996.
- Last JM. A dictionary of epidemiology. 3rd ed. New York, NY: Oxford University Press, 1995.
- Eyler JM. The conceptual origins of William Farr's epidemiology: numerical methods and social thought in the 1830s. In: Lilienfeld AM, ed. Times, places, and persons: aspects of the history of epidemiology. Baltimore, MD: Johns Hopkins University Press, 1980:1-21.
- Lew EA. Actuarial contributions to life table analysis. *Natl Cancer Inst Monogr* 1985;67:29-36.
- Bernstein PL. Against the gods: the remarkable story of risk. New York, NY: Wiley, 1996.
- Singer RB. Examples of early mortality follow-up studies. *Natl Cancer Inst Monogr* 1985;67:15-21.
- Specialized mortality investigation. New York, NY: Actuarial Society of America, 1903.
- Sartwell PE. Editorial: The time factor in studies of the outcome of chronic disease. *Am Rev Tuberc* 1951;63:608-12.
- Comstock GW. Early studies of tuberculosis. *Natl Cancer Inst Monogr* 1985;67:23-7.
- Brown L, Pope EG. The post-discharge mortality among the patients of the Adirondack Cottage Sanitarium. *Am Med* 1904;8:879-82.
- Elderton WP, Perry SJ. Drapers' company (London, England). Research memoirs. Studies in national deterioration VI. A third study of the statistics of pulmonary tuberculosis: the mortality of the tuberculous and sanatorium treatment. London, England: Dalau, 1910.
- Frost WH. The age selection of mortality from tuberculosis in successive decades. *Am J Hyg* 1939;30A:91-6.
- Andvord KF. What can we learn by studying tuberculosis generations? (English translation by Gerard Wijsmuller). *Norsk mag Lægevidensk* 1930;91:642-60.
- Frost WH. Risk of persons in familial contact with pulmonary tuberculosis. *Am J Public Health* 1933;23:426-32.
- Zeidberg LD, Gass RS, Dillon A, et al. The Williamson County Tuberculosis Study: a twenty-four-year epidemiologic study. *Am Rev Respir Dis* 1963;87(no.3, part 2):1-88.
- Dawber TR. The Framingham study: the epidemiology of atherosclerotic disease. Cambridge, MA: Harvard University Press, 1980.
- Truett J, Cornfield J, Kannel W. A multivariate analysis of the risk of coronary heart disease in Framingham. *J Chronic Dis* 1967;20:511-24.
- Schull WJ. Effects of atomic radiation: a half-century of studies from Hiroshima and Nagasaki. New York, NY: Wiley-Liss, 1995.
- Samet JM. Epidemiologic studies of ionizing radiation and cancer: past successes and future challenges. *Environ Health Perspect* 1997;105(Suppl 4):883-9.
- Lundin FE, Wagoner JK, Archer VE. Radon daughter exposure and respiratory cancer, quantitative and temporal aspects. National Institute for Occupational Safety and Health-National Institute of Environmental Health Sciences joint monograph no. 1. Washington, DC: US Department of Health, Education, and Welfare, Public Health Service, 1971.
- National Research Council Committee on Health Risks of Exposure to Radon. Report of the committee on health risks of exposure to radon (BEIR VI). Washington, DC: National Academy Press, 1997.
- Whittemore AS, McMillan A. Analyzing cohort data: application to US uranium miners. In: Prentice RL, Whittemore AS, eds. Environmental epidemiology: risk assessment. Philadelphia, PA: SIAM, 1982:65-81.
- Lubin JH, Gail MH. On power and sample size for studying features of the relative odds of disease. *Am J Epidemiol* 1990;131:552-66.
- Thomas DC, McNeill KG, Dougherty C. Estimates of lifetime lung cancer risks resulting from Rn progeny exposure. *Health Phys* 1985;49:825-46.
- Doll R, Hill AB. The mortality of doctors in relation to their smoking habits: a preliminary report. *Br Med J* 1954;1:1451-5.
- Doll R, Peto R, Wheatley K, et al. Mortality in relation to smoking: 40 years' observations on male British doctors. *BMJ* 1994;309:901-11.
- Hammond EC, Horn D. Smoking and death rates report on forty-four months of follow-up of 187,783 men. I. Total mortality. II. Death rates by cause. *JAMA* 1958;166:1159-72, 1294-308.
- Dorn HF. Tobacco consumption and mortality from cancer and other diseases. *Public Health Rep* 1959;74:981-93.
- US Department of Health Education and Welfare, Public Health Service. Smoking and health; report of the advisory committee to the Surgeon General of the Public Health Service. Washington, DC: US GPO, 1964.
- Case RAM, Hosker ME, McDonald DB, et al. Tumors of the urinary bladder in workmen engaged in the manufacture and use of certain dyestuff intermediates in the British chemical industry. Part I: The role of aniline, benzidine, alpha-naphthylamine, and beta-naphthylamine. *Br J Ind Med* 1954;11:75-104.
- Doll R. Mortality from lung cancer in asbestos workers. *Br J Ind Med* 1955;12:81-6.
- Selikoff LJ, Hammond EC, Seidman H. Mortality experience of insulation workers in the United States and Canada, 1943-1976. *Ann NY Acad Sci* 1979;330:91-116.
- Hammond EC. Smoking in relation to the death rates of one million men and women. *Natl Cancer Inst Monogr* 1966;19:127-204.
- Fletcher CM, Peto R, Tinker C, et al. The natural history of chronic bronchitis and emphysema: an eight year study of early chronic obstructive lung disease in working men in London. New York, NY: Oxford University Press, 1976.
- Higgins MW, Keller JB. Predictors of mortality in the adult population of Tecumseh. *Arch Environ Health* 1970;21:418-24.
- Heyden S, Cassel JC, Bartel A, et al. Body weight and cigarette smoking as risk factors. *Arch Intern Med* 1971;128:915-19.
- Berkman LF, Syme SL. Social networks, host resistance, and mortality: a nine-year follow-up study of Alameda County residents. *Am J Epidemiol* 1979;109:186-204.
- Menkes MS, Comstock GW, Vuilleumier JP, et al. Serum beta-carotene, vitamins A and E, selenium, and the risk of lung cancer. *N Engl J Med* 1986;315:1250-4.
- Lebowitz MD, Burrows B. Tucson epidemiologic study of

- obstructive lung diseases. II. Effects of in-migration factors on the prevalence of obstructive lung diseases. *Am J Epidemiol* 1975;102:153–63.
48. The ARIC Investigators. The Atherosclerosis Risk in Communities (ARIC) Study: design and objectives. *Am J Epidemiol* 1989;129:687–702.
 49. Fried LP, Borhani NO, Enright P, et al. The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* 1991;1:263–76.
 50. Lee ET, Welty TK, Fabsitz R, et al. The Strong Heart Study: a study of cardiovascular disease in American Indians: design and methods. *Am J Epidemiol* 1990;132:1141–55.
 51. Newcombe HB. Handbook of record linkage: methods for health and statistical studies, administration and business. New York, NY: Oxford University Press, 1988.
 52. Cox CS, Rothwell ST, Madans JH, et al. Plan and operation of the NHANES I Epidemiologic Followup Study, 1987. National Center for Health Statistics. *Vital Health Stat* [1] 1992;(27):1–190.
 53. Colditz GA, Willett WC, Hunter DJ, et al. Family history, age, and risk of breast cancer: prospective data from the Nurses' Health Study. *JAMA* 1993;270:338–43.
 54. Polk BF, Fox R, Brookmeyer R, et al. Predictors of the acquired immunodeficiency syndrome developing in a cohort of seropositive homosexual men. *N Engl J Med* 1987;316:61–6.
 55. Mellors JW, Muñoz A, Giorgi JV, et al. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. *Ann Intern Med* 1997;126:946–54.
 56. Mantel N. Synthetic retrospective studies and related topics. *Biometrics* 1973;29:479–86.
 57. Liddell FDK, McDonald JC, Thomas DC. Methods of cohort analysis: appraisal by application to asbestos mining. *J R Stat Soc* 1977;140:469–91.
 58. Prentice RL. A case-cohort design for epidemiologic cohort studies and disease prevention trials. *Biometrika* 1986;73:1–11.
 59. Greenland S. Response and follow-up bias in cohort studies. *Am J Epidemiol* 1977;106:184–7.
 60. Royall RM. Statistical evidence: a likelihood paradigm. New York, NY: Chapman and Hall, 1997.
 61. Lehmann EL. Theory of point estimation. New York, NY: Springer-Verlag, 1997.
 62. Tufte ER. The visual display of quantitative information. Cheshire, CT: Graphics Press, 1983.
 63. Gange SJ, Muñoz A, Chmiel JS, et al. Identification of inflections in T-cell counts among HIV-1 infected individuals and relationship with progression of clinical AIDS. *Proc Natl Acad Sci* 1998 (in press).
 64. Cox DR. Regression models and life tables. *J R Stat Soc* 1972;B34:187–202.
 65. Jennrich RI, Schluchter MD. Unbalanced repeated-measures models with structured covariance matrices. *Biometrics* 1986;42:805–20.
 66. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics* 1986;42:121–30.
 67. Korn EL, Whittemore AS. Methods for analyzing panel studies of acute health effects of air pollution. *Biometrics* 1979;35:795–802.
 68. Whittemore AS, Korn EL. Asthma and air pollution in the Los Angeles area. *Am J Public Health* 1980;70:687–96.
 69. Rosner B, Muñoz A. Conditional linear models for longitudinal data. In: Dwyer JH, Feinleib M, Lippert P, et al., eds. Statistical models for longitudinal studies of health. New York, NY: Oxford University Press, 1992:115–37.
 70. Laird NM, Ware JH. Random-effects models for longitudinal data. *Biometrics* 1982;38:963–74.
 71. Kupper LL, Haseman JK. The use of a correlated binomial model for the analysis of certain toxicological experiments. *Biometrics* 1978;34:69–76.
 72. Breslow N. Extra-Poisson variation in log-linear models. *Appl Stat* 1984;33:38–44.
 73. Segal MR. Regression trees for censored data. *Biometrics* 1988;44:35–48.
 74. Ciampi A, Lawless JF, McKinney SM, et al. Regression and recursive partition strategies in the analysis of medical survival data. *J Clin Epidemiol* 1988;41:737–48.
 75. McCullagh P, Nelder JA. Generalized linear models. New York, NY: Chapman and Hall, 1989.
 76. Hastie TJ, Tibshirani RJ. Generalized additive models. New York, NY: Chapman and Hall, 1990.
 77. Muñoz A, Sabin CA, Phillips AN. The incubation period of AIDS. *AIDS* 1997;11(Suppl A):S69–76.
 78. Muñoz A, Xu J. Models for the incubation of AIDS and variations according to age and period. *Stat Med* 1996;15:2459–73.
 79. Detels R, Muñoz A, McFarlane G. Effectiveness of highly active antiretroviral therapy on the natural history of HIV/AIDS in a large cohort of seroconverters. *JAMA* 1998 (in press).
 80. Wei LJ. The accelerated failure time model: a useful alternative to the Cox regression model in survival analysis. *Stat Med* 1992;11:1871–9.
 81. Muñoz A, Sunyer J. Comparison of semiparametric and parametric survival models for the analysis of bronchial responsiveness. *Am J Respir Crit Care Med* 1996;154(Suppl):S234–9.
 82. Wacholder S, Gail MH, Pee D, et al. Alternative variance and efficiency calculations for the case-cohort design. *Biometrika* 1989;76:117–23.
 83. Muñoz A, Gange SJ. Methodological issues for biomarkers and intermediate outcomes in cohort studies. *Epidemiol Rev* 1998;20:15–28.
 84. Lamarca R, Alonso J, Gomez G, et al. Left-truncated data with age as time scale: an alternative for survival in the elderly. *J Gerontol A Biol Sci Med Sci* (in press).
 85. Armstrong BK, White E, Saracci R. Principles of exposure measurement in epidemiology. New York, NY: Oxford University Press, 1992.
 86. Carroll RJ, Ruppert D, Stefanski LA. Measurement error in nonlinear models. London, England: Chapman and Hall, 1995.
 87. Tsodis AA, De Gruttola V, Wulfsohn MS. Modeling the relationship of survival to longitudinal data measured with error: applications to survival and CD4 counts in patients with AIDS. *J Am Stat Assoc* 1995;90:27–37.
 88. De Gruttola V, Tu XM. Modelling progression of CD4-lymphocyte count and its relationship to survival time. *Biometrics* 1994;50:1003–14.
 89. Diggle P, Kenward MG. Informative drop-out in longitudinal data analysis. *Appl Stat* 1994;43:49–93.
 90. Little RJA. Modeling the drop-out mechanism in repeated-measures studies. *J Am Stat Assoc* 1995;90:1112–21.
 91. Hogan JW, Laird NM. Mixture models for the joint distribution of repeated measures and event times. *Stat Med* 1997;16:239–57.
 92. Hogan JW, Laird NM. Model-based approaches to analysing incomplete longitudinal and failure time data. *Stat Med* 1997;16:259–72.
 93. Chmiel JS, Detels R, Kaslow RA, et al. Factors associated with prevalent human immunodeficiency virus (HIV) infection in the Multicenter AIDS Cohort Study. *Am J Epidemiol* 1987;126:568–77.
 94. Kingsley LA, Zhou SY, Bacellar H, et al. Temporal trends in human immunodeficiency virus type 1 seroconversion 1984–1989: a report from the Multicenter AIDS Cohort Study (MACS). *Am J Epidemiol* 1991;134:331–9.
 95. Margolick JB, Muñoz A, Donnenberg AD, et al. Failure of T-cell homeostasis preceding AIDS in HIV-1 infection: the Multicenter AIDS Cohort Study. *Nat Med* 1995;1:674–80.
 96. Centers for Disease Control and Prevention. Report of the NIH panel to define principles of therapy of HIV infection and guidelines for the use of antiretroviral agents in HIV-

- infected adults and adolescents. *MMWR* 1998;47(RR-5):1–82.
97. Muñoz A, Wang MC, Bass S, et al. Acquired immunodeficiency syndrome (AIDS)-free time after human immunodeficiency virus type 1 (HIV-1) seroconversion in homosexual men: Multicenter AIDS Cohort Study Group. *Am J Epidemiol* 1989;130:530–9.
 98. Muñoz A, Carey V, Taylor JM, et al. Estimation of time since exposure for a prevalent cohort. *Stat Med* 1992;11:939–52.
 99. Taylor JM, Muñoz A, Bass SM, et al. Estimating the distribution of times from HIV seroconversion to AIDS using multiple imputation: Multicenter AIDS Cohort Study. *Stat Med* 1990;9:505–14.
 100. Fahey JL, Taylor JM, Detels R, et al. The prognostic value of cellular and serologic markers in infection with human immunodeficiency virus type 1. *N Engl J Med* 1990;322:166–72.
 101. Enger C, Graham N, Peng Y, et al. Survival from early, intermediate, and late stages of HIV infection. *JAMA* 1996;275:1329–34.
 102. Jacobson LP, Kirby AJ, Polk S, et al. Changes in survival after acquired immunodeficiency syndrome (AIDS): 1984–1991. *Am J Epidemiol* 1993;138:952–64.
 103. Muñoz A, Schrager LK, Bacellar H, et al. Trends in the incidence of outcomes defining acquired immunodeficiency syndrome (AIDS) in the Multicenter AIDS Cohort Study: 1985–1991. *Am J Epidemiol* 1993;137:423–38.
 104. Phair J, Jacobson L, Detels R, et al. Acquired immune deficiency syndrome occurring within 5 years of infection with human immunodeficiency virus type-1: the Multicenter AIDS Cohort Study. *J Acquir Immune Defic Syndr* 1992;5:490–6.
 105. Gange SJ, Jacobson LP, Muñoz A. Parametric models to assess tracking of categorical high-risk sexual behaviour for HIV infection. *Stat Med* 1996;15:2323–36.
 106. Muñoz A, Hoover DR. Use of cohort studies for evaluating AIDS therapies. New York, NY: Wiley, 1995.
 107. Muñoz A, Kirby AJ, He YD, et al. Long-term survivors with HIV-1 infection: incubation period and longitudinal patterns of CD4+ lymphocytes. *J Acquir Immune Defic Syndr Hum Retrovirol* 1995;8:496–505.
 108. Gange SJ, Muñoz A, Schrager LK, et al. Design of nested studies to identify factors related to late progression of HIV infection. *J Acquir Immune Defic Syndr Hum Retrovirol* 1997;15(Suppl 1):S5–9.

Lecture: Evaluating and Reporting Cohort Studies



Evaluating and reporting cohort studies

PHW250 F - Jack Colford

Let's discuss best practices for evaluating and reporting cohort studies.

Assessment of cohort studies

How much selection bias was present?

- Were only people at risk of the outcome included?
- Was the exposure clear, specific, and measurable?
- Were the exposed and unexposed groups similar in all important respects except for the exposure?

What steps were taken to minimise information bias?

- Was the outcome clear, specific, and measurable?
- Was the outcome identified in the same way for both groups?
- Was determination of outcome made by an observer blinded as to treatment?

How complete was the follow-up of both groups?

- What efforts were made to limit loss to follow-up?
- Was loss to follow-up similar in both groups?

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

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

Berkeley School of Public Health
Grimes et al. 2002

When we think about cohort studies, we want to consider several different features.

First, we want to talk about whether there was any selection bias that was present.

Now, you'll recall when we studied selection bias, the kinds of things we talked about were questions such as, were only people who were at risk of the outcome included? Obviously, if we're doing a study of a particular condition that only involves women, we wouldn't want to include men, for example.

So that's an extreme example, but we want to make sure that the population we pick were at risk for what we were studying. We want to ask whether the authors made clear what the exposure was. Was the exposure clearly stated? Was it specific? Was it measurable in a way that we can understand?

Next, were the exposed and unexposed groups similar in all important respects, except for the exposure. Because what we're trying to determine is, if they differ only by exposure, then if we see a difference in an outcome, then we can attribute the difference we saw to the difference in exposure. If they differ by too many other features, then we wouldn't be sure that any difference in outcome was due to the difference in exposure.

Next, we want to know whether the authors took steps to minimize information bias. In order to do that, we're really thinking about the outcome that was measured. Was the outcome very clear, specific, and measurable? Was it identified in the same way for

both groups? We wouldn't want to use one tool to measure lung cancer in the exposed group-- say, a high-tech MRI-- and a different tool to measure the outcome in the unexposed group-- say a low-tech X-ray.

And finally, was determination of the outcome made by an observer blinded as to treatment? Because we don't want the observer to themselves have introduced bias by hoping that a certain outcome was seen in one group or the other-- the exposed group or the unexposed group.

Next question we ask is, how complete was the follow up in both groups? Were there specifically any efforts made to limit loss to follow-up. If we lose too many people from either of our groups, we no longer have a representative sample. Was the loss to follow-up similar in both groups? Even if we did lose people in the study, was the loss to follow-up balanced in the two groups? Asymmetric loss is something we hope to not see in studies that we review or conduct.

And finally, did the authors control potential confounding factors in their analysis? Did they anticipate potential confounding factors, and gather information about them? And what methods did they use to assess and control the confounding?

STROBE Checklist for reporting cohort studies

The screenshot shows the first page of a journal article. At the top left is the PLoS Medicine logo. The title "Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and Elaboration" is centered. Below the title is a section titled "ABSTRACT". The abstract text discusses the importance of reporting observational studies and the development of the STROBE statement to improve reporting quality. It highlights the 22-item checklist, common items for cohort, case-control, and cross-sectional studies, and specific items for each study design. The text is dense and technical, providing detailed guidance for researchers.

- 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.



To help us evaluate cohort studies, there are different checklists that can be used. And these are worth following. The one I'd like to recommend that you think about for reporting out cohort studies, is called the STROBE checklist. That stands for Strengthening The Reporting of Observational Studies in Epidemiology. And this checklist grew out of concerns that observational studies were being poorly and inconsistently reported.

And of course, if the study's poorly reported, it makes it difficult to assess the strengths and weakness of a study. The STROBE checklist has 22 items on it-- 18 of these items are common to cohort studies, case control studies, and cross-sectional studies. And then four items-- each are specific to the three different study design. Depending on whether you do in a cohort, case control or cross-sectional study, there are four items that are specific to each of those three types.

STROBE Statement—Checklist of items that should be included in reports of *cohort studies*

Item No	Recommendation
Title and abstract	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
Introduction	
Background/rationale	2 Explain the scientific background and rationale for the investigation being reported
Objectives	3 State specific objectives, including any prespecified hypotheses
Methods	
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 selection of participants. Describe methods of follow-up (b) For matched studies, give matching criteria and number of exposed and unexposed
Variables	7 Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable

Just to quickly give a high-level overview of major categories you see here in the STROBE checklist. Every aspect of cohort study report is covered here this. Is a good tool to use when helping to design a cohort study. Thinking about the eventual reporting of the cohort study is a really great way to think ahead about building your study.

First of all, the title and abstract. And you can see the recommendations here for making the title very specific to conform to what the design of the study was, so readers can tell that quickly. And then there are checklist items to cover with respect to the introduction and the methods.

For example, with the participants, what were the specific eligibility criteria? You'd like other authors to be able to repeat your study, and enroll the similar type of participant in the study. And then for the variables used in the study. Did the authors clearly define all the outcomes, the exposures, the predictors, the potential confounders, and the effect modifiers? As you're beginning in epidemiology, it's good for you to go through each of these, and make a list of what variables the authors use for each of these different categories of variable.

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 loss to follow-up was addressed (e) Describe any sensitivity analyses



What were the different data sources and measurement? How was bias addressed in the study? This is something, of course, we really want you to focus heavily on in this course is, what are the different types of bias that might be present in a given study that you're reviewing? And how did the authors address each of those types of bias? How was the study sample size calculated and arrived at? What were the quantitative variables measured in the analyses, and how were they analyzed? And finally, what were the details of the statistical methods? All of these specific details of the cohort study should be reported.

Results		
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 (c) Summarise follow-up time (eg, average and total amount)
Outcome data	15*	Report numbers of outcome events or summary measures over time
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

With respect to the results of the study, you can see here a list of the checklist items. What are the participants? How many were there at each stage of the study? Why were non-participants excluded? Why did non-participants not take part in the study? And a flow diagram can be helpful here. For descriptive data, we want to talk about the characteristics of the study participants, the number of participants who had missing data, and then summarize what the follow-up time was for each of these.

How about the outcome data in the study? Do the authors give us clearly the numbers of outcome events, or the summary measures over time? And finally, what were the main results? Usually, it's appropriate to give both unadjusted estimates and adjusted estimates. And by adjusted estimates I mean estimates that have been adjusted for potential confounders.

And then when continuous variables are used, we're going to give category boundaries. And if it's relevant for the particular study, we might want to consider translating estimates of relative risk into absolute risk for some 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



With respect to the discussion that the authors provide, did they summarize the key results? And did they reference that back to the study objectives? What were the limitations of the study? How were the authors interpreting the results of the study? Hopefully, in a cautious way, that takes into account the objectives, limitations, the multiplicity of different analyzes they did. How were the results generalized by the authors? By generalizability here, we're usually referring to external generalizability, or external validity.

We did the study in one particular population. How do these results likely apply to other populations where the study wasn't conducted? And very importantly, somewhere the authors should fully disclose all the related links to funding for the study. Who funded the study? Are there any potential conflicts of interest?

Summary of key points

- You can use the list of questions from the Grimes et al. 2002 article when evaluating the quality of a cohort study.
- We recommend that you use the STROBE reporting checklist when publishing results of a cohort study.
- The article by Vandenbrouke et al. 2007 article provides detailed examples of how to use the STROBE checklist with an example paper.



To summarize, we can take the list of questions that come from the Grimes, 2002 article we've provided for you, when you're evaluating the quality of a cohort study. I recommend that you use the STROBE reporting checklist when you're publishing the results of a cohort study, and when you're reviewing-- perhaps in class, or outside of class-- a cohort study. And the article we provided by Vandenbrouke et al, from 2007, gives detailed examples of how to use the STROBE checklist with an example paper.

STROBE Checklist for Cohort Studies

STROBE Statement—Checklist of items that should be included in reports of *cohort studies*

Item No	Recommendation	
Title and abstract	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
Introduction		
Background/rationale	2	Explain the scientific background and rationale for the investigation being reported
Objectives	3	State specific objectives, including any prespecified hypotheses
Methods		
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 selection of participants. Describe methods of follow-up (b) For matched studies, give matching criteria and number of exposed and unexposed
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 loss to follow-up was addressed (e) Describe any sensitivity analyses
Results		
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 (c) Summarise follow-up time (eg, average and total amount)
Outcome data	15*	Report numbers of outcome events or summary measures over time
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 exposed and unexposed groups.

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>.

[MUSIC PLAYING]

JADE BENJAMIN- Hi, everyone. Today, we're going to talk to Dr. Benjamin Arnold about the surfer health study,
CHUNG: and this is actually a study that I was a co-investigator on. But for the purpose of this interview,
I'm going to pretend like I don't know anything about it and just ask Ben a lot of questions. So
thanks so much for joining me today, Ben.

BENJAMIN Oh, my pleasure.

ARNOLD:

JADE BENJAMIN- Really excited to hear more about the study, and I'm hoping we can start by having you tell us
CHUNG: what the main hypothesis of the study was.

BENJAMIN Right. So we set out to study whether surfers who entered the sea-- so who entered the
ARNOLD: ocean-- in the three days following rainstorms were at higher risk of acute infectious symptoms
and illness compared to periods when they entered the sea and it was dry weather. We
hypothesized that exposure to seawater following wet weather would increase illness rates
more than exposure during dry weather periods because there is an immense amount of
bacteria flushed into the ocean from large storms in California.

JADE BENJAMIN- And how did you become interested in this research question?

CHUNG:

BENJAMIN Well, there had been a lot of research in California on health risks associated with seawater
ARNOLD: exposure, but those were predominantly done among swimmers in the summer because the
vast majority of people who enter the ocean in California do so between April and November.

There's been increasing pressure on the urban coastal communities to improve water quality
during the winter months as well because there actually are a lot of people who go into the sea
during the winter. And in Southern California, surfing is incredibly popular. There are
thousands of surfers in the water every day, and so getting a better understanding of whether
the risk was even higher falling rain-- it was really important to help determine whether the risk
standards need to be different during the wet season or not.

monitoring studies have documented that there was a whole lot higher levels of fecal indicator bacteria in the water during those periods right after rainstorms.

Now, fecal indicator bacteria are types of bacteria that tend to indicate there's some feces in the water. Now, that doesn't mean that they necessarily cause illness. It's not necessarily human pathogens, but they tend to move with pathogens, and they're easier to measure.

JADE BENJAMIN- So some of our viewers may be wondering, when you talk about feces in the water, are these **CHUNG:** feces coming from people who enter the water or from other sources?

BENJAMIN Yeah, that's a great question. So it could be human feces, but it could also be lots of animals.

ARNOLD: So even in urbanized coastlines in California like San Diego, they have large catchments that have agricultural regions as well. So you can have cattle. You can have birds, dogs, cats, all different types of animals.

And these indicator bacteria can actually reflect other sources of feces besides human feces.

Now, sometimes there are zoonotic pathogens in bird feces, for example, but by and large, human feces are much more dangerous in terms of human health risk than some of the other types.

JADE BENJAMIN- And you alluded to the sort of fleshing during rainstorms. Are there other sources of human **CHUNG:** feces outside of the water that you would expect to get flushed into the ocean?

BENJAMIN Yeah, there are a few different sources. So in more rural areas that eventually flow to the sea, **ARNOLD:** you have a lot of septic systems. And with higher rainfall, those can sometimes leak a whole lot more untreated sewage into freshwater drainages.

But there's also leaky pipes. And so urban infrastructure is always degrading, and pipes inevitably leak. Now, if sewer pipes leak into right around the area around the pipes, that's OK. But oftentimes, the pathogens are just sitting there, and then when there's a lot of freshwater flowing through the soil, it can actually seep all the way through and get into a sea.

That, and then there are, unfortunately, many homeless people living in coastal California, especially in Southern California, and they often live in freshwater catchments areas. So there is potential for just feces on the surface getting literally flushed into the streams and rivers from the large storms.

- CHUNG:** when you were preparing the study design and thinking about the results? How would it inform, for example, beach closure or other kinds of local policies?
- BENJAMIN** Yeah, so currently in Southern California and many parts of California, beach managers will post warnings for people not to go in the water for three days after big rainstorms. And a lot of that policy has been guided by these environmental monitoring studies that I mentioned before where the indicator bacteria levels tend to be high. By three days, they've gone back down to background levels, and so that's widely been considered the window of risk for beach-goers. Now, there weren't any actual health studies to determine whether that was the appropriate window. It might be shorter. It might actually be longer. And so that motivated the study.
- JADE BENJAMIN-** For our viewers who are not surfers, can you give a little bit of a-- paint a picture for us about
- CHUNG:** the exposure surfers face when they enter the water and how that informs their disease risk?
- BENJAMIN** Yeah. So compared to swimmers, as one example, surfers tend to have much higher levels of exposure in a few different dimensions. So first of all, they tend to go in the ocean very often. At least in San Diego, just within this surfer health study, most participants went surfing two to three times each week. So they're regularly exposed.
- They're also in the water for quite a while. So they'll typically surf for an hour, sometimes up to two hours at a time. And then, once in the water, the nature of their exposure is really intense. So there's a lot of head immersion. With wipe-outs, people often swallow water, and that type of exposure you don't see as often among swimmers.
- Sometimes as young children, they will get bowled over by a wave and end up swallowing a lot of water, but by and large, for you and I, we don't tend to drink a lot of water when we go swimming in the sea. Surfers, though, have high levels of exposure, and for that reason, they're almost a canary in the coal mine, to use a bit of a metaphor there, for this type of exposure risk.
- JADE BENJAMIN-** Can you tell me how you came to choose a cohort design for this study?
- CHUNG:**
- BENJAMIN** Yeah, so we were very interested in being careful about temporality between exposure and outcome, so we wanted to establish very clearly that seawater exposure preceded illness. And the reason that we'd be concerned about that is because when somebody is sick, they may be less likely to go into the ocean, right? Because if you're not feeling well, you're probably not

going to go surfing. So a cross-sectional design can't really disentangle that temporality, and that was part of the reason.

The other part was that to help make the design more robust, we followed surfers longitudinally. And what that meant is they contributed exposure to both unexposed surf sessions during dry periods as well as during wet periods. And so they actually served as their own control, in some sense, and so it was a more robust design.

JADE BENJAMIN- Do people ever do randomized trials of this type of recreational water research question?

CHUNG:

BENJAMIN Yeah, it's a good question. So there have been randomized trials among swimmers in the UK

ARNOLD: and in Florida. And in those studies, they enroll people who come to the beach and then randomly allocate whether or not they're allowed to go in the water or not. And among those who are assigned to go in the water, they then often will have a very prescribed number of head dunks per time and things like that.

And it turns out, though, that the randomized trials and observational cohort studies of these questions have by and large come up with very similar estimates. There doesn't seem to be a whole lot of unmeasured confounding associated with the observational studies. And one benefit of the observational cohorts is that you tend to enroll-- in more general populations, sometimes the people who consent to trials are different from the general beach-goer population. And in this case, I doubt we would have been able to recruit any surfers if they had been subject to not being able to go surfing when they wanted to.

JADE BENJAMIN- Right. And the nature of surfing probably makes it really difficult to control the type of

CHUNG: exposure. You can't say dunk your head or swallow this much water this many times when you're surfing.

BENJAMIN That's right. Yeah, that kind of artificial exposure wouldn't mimic the real surf experience, I

ARNOLD: don't think.

JADE BENJAMIN- We've talked a bit about exposures. What about outcomes? So which outcomes did you

CHUNG: measure, and how did you choose them in the study?

BENJAMIN Right. So unlike some cohort studies where we would pick one primary outcome and design

ARNOLD: everything around that, we didn't do that in this study, and the reason was that when we talked

to other members of our study team at Surfrider Foundation and many stakeholders in the surfer community, it really came down to the fact that there were a range of illnesses that they were very concerned about.

So from a regulatory perspective, a lot of recreational water studies have focused on gastrointestinal illness, which includes diarrhea, vomiting, nausea and stomach cramps and a composite of those. That's one type of illness, but surfers have many other types. So we were also interested, for example, in sinus infections, which are quite common among surfers, ear ache and ear infections, infected open wounds. So if you go into the sea with a cut that's open, there's a very high chance that it could become infected.

And a lot of anecdotal reports among surfers-- again, not in the epidemiologic studies, but many case studies of MRSA infections and things like that among people in the community. So we tried to capture as many infectious symptoms as we could that we felt there was strong biologic possibility as well as were of interest to surfers in general.

JADE BENJAMIN- And then for the exposure, can you tell me in a little more detail how you classified exposed
CHUNG: and unexposed over different periods of time, and also for the weather exposure, water quality, and the type of recreation that they did?

BENJAMIN Sure. So I'll break it down into three parts. So first, at the most simple level, we followed
ARNOLD: individuals over time. And so we would have daily information about whether they went in the water or they didn't. So if a surfer went in the ocean, for the three days after that, we would classify that as an exposed period because we'd be interested to know whether they developed illness within that three-day window following entry in the sea. Now, everything else, all the other time that they contributed would be unexposed. So they weren't exposed to the ocean.

Now, within those ocean exposures, we were also interested in classifying whether it was dry weather or wet weather exposure, and to do that, we used a definition of wet weather that was commonly used in Southern California to post beach warnings, which is-- if it's rained at least a quarter centimeter-- that's about 0.1 inches-- that's considered a wet weather event. So no just foggy drizzle. It has to actually rain. So if it's rained 0.1 inches in 24 hours, that's a wet weather day.

we had this three tiers of exposure-- unexposed, dry weather exposure, and wet weather exposure.

Now, the third dimension of exposure was actually looking at the fecal indicator bacteria that we talked about earlier. So there are a range of different indicator bacteria, but one that we focused on most is a genus called Enterococcus.

And so what we did was within the study, which captured surfing events all over San Diego County, we had two sentinel beaches where we did much more detailed water quality measurement, and those included Ocean Beach, which is where the San Diego River flows into the sea. So that's a big catchment. It's a huge river. There's a lot that goes into it and a lot that goes out.

So we had four monitoring sites on Ocean Beach where our field staff, comprised of students, mostly from UCSD, would actually collect water samples every morning. They would then test them for quantitative measures of how much this Enterococcus was in the water. So we had four sites at Ocean Beach.

We also had a second sentinel beach called Tourmaline Surf Park and Tourmaline Surf Beach, where we had two more sites. And at those beaches, again, we would measure Enterococcus every morning of the study, and we would link those quantitative measures of Enterococcus to any surf sessions that happened on that day at that beach. So it was a subset of all the surfing that happened during the study, but it actually comprised of almost half of the surf sessions at those two beaches because we heavily recruited from there. And so we could link up the fecal indicator bacteria with actual surfers' exposure.

JADE BENJAMIN- Can you briefly tell us why it's necessary to measure fecal indicator bacteria?

CHUNG:

BENJAMIN Yeah. So historically, it's been very hard to measure actual pathogens in the water. So in
ARNOLD: recreational water exposure, we think that a lot of the pathogens are virus-- so norovirus is one example for gastrointestinal illness.

And detecting virus in environmental samples has historically been very hard, so there's a lot of inhibition with molecular methods. That's changing now, but even up and through this study, it was very hard to get pathogen-specific measurements. So the indicator bacteria tend to move with pathogens and are much easier to measure, and so in a lot of cases, they do

provide a reliable signal of actual presence of pathogens, and therefore health risk.

JADE BENJAMIN- You've referred a few times to daily level information on surf activity, and can you tell us a bit
CHUNG: about how you collected that information, and also how you recruited the surfers for the study?

BENJAMIN Yeah. So first I'll talk about recruitment, then how we collected information from them. So we
ARNOLD: did a two-pronged approach to recruitment. The study covered two surf seasons over two winters. And in the first winter, we considered enrollment through in-person-- so on the beach-- and also via the web, because we partnered with Surfrider Foundation, which has a huge chapter in San Diego-- I think over 10,000 members. But we weren't sure that if we advertised through Surfrider Foundation if we would actually get a representative sample of surfers compared to-- and/or surfers who would reliably go surfing in the winter versus if we actually went out on the beach in the winter and saw the people who were going in the water and tried to recruit them into the study.

So in the first winter, we did have a field team of local surfers who we hired, and they went out on the beach with tablets, actually would try to recruit people into the study, who would then later contribute surveys that I'll talk about in a sec. Turned out that surfers recruited on the beach and those recruited online were almost identical in almost every respect. So there were a few small differences, but they were so similar that in the second year, we decided to just go for online recruitment only because it's so much easier to scale, and it was a lot less expensive.

Now, once somebody enrolled into the study, they actually downloaded an app to their smartphone. And so our colleagues at the Southern California Coastal Water Research Project, who really led a lot of the microbiology, also led a lot of the IT and data collection logistics. And they developed a custom smartphone app for the study that-- we had a pretty simple questionnaire. We always try to keep them minimal, so it'd take about 10 minutes.

And each week, participants would get a text or an email, or both, whichever they preferred, just as a reminder. we'd send it on a Tuesday, because it was more likely to get good responses. And then they'd go and fill out a 10-minute questionnaire that would detail over the past seven days if they'd been surfing and which days, and then where they went surfing, and then separately, a range of illness symptoms that we talked about and when those occurred-- so on which days they were and how long they lasted. And some people get reminders every week, and for every four surveys they completed, they'd get a gift card to a local surf retailer to

be able to reimburse them for their time.

JADE BENJAMIN- And their follow-up was pretty good?

CHUNG:

BENJAMIN Yeah. So the median follow-up was eight weeks in the study, and we had-- so about two

ARNOLD: months in each winter. And it ranged from one week to 20 weeks. So some people stayed in for a full four months, roughly.

JADE BENJAMIN- Great. It sounds like a great system.

CHUNG:

BENJAMIN Yeah. Yeah, no, it worked quite well. We were very pleased with it, also because when you

ARNOLD: collect data electronically, there's a lot of opportunity to do logic checks in the questionnaire and things and ensure your range value, checking and-- on the fly to ensure that somebody accidentally mistypes or does something that doesn't make sense, it'll send a little warning to them and correct it.

JADE BENJAMIN- So the web app collected information about symptoms, and you used that information to

CHUNG: calculate incidence rates. Can you talk a bit about your choice to use that measure of disease for the study?

BENJAMIN Yeah. Yeah, no, it's a good question because in epidemiology, we love the incidence rate on

ARNOLD: one level, but we hate it in others. So the reason I love it is because it's the fundamental measure of disease, and when we're looking at etiology and risk factors like this study, it's a great outcome measure. But the reason I don't like it is that its units are sometimes hard to interpret for people who aren't epidemiologists.

So yeah, we measured incidence rates. They're sometimes called the incidence density, which is the number of episodes per unit of person time. And the reason that we measured incidence rates in this study is twofold. First, it was an open cohort design, which meant that we didn't enroll surfers at one point in time and follow them for a very fixed window and then measure outcomes over that period. A lot of past recreational water studies would do that. They'd enroll people at the beach and then follow them for two weeks, and that was a very well-defined risk window. Here, we had surfers entering and exiting the study almost on a revolving basis, and so they contributed a lot of varying exposure time, and we wanted to use all of that. We didn't want to just arbitrarily shorten it to a 10-day window or something.

The second factor, though, is that at least among our study participants-- I think this is true in general in San Diego-- surfers enter the ocean all the time. So I mentioned they go two to three times a week. That means that for each individual exposure, you only have a couple of days with which to look at-- tie incident outcomes to that particular exposure before they're exposed again. And so we had a very-- this isn't a technical term, but we would then slice their person time into very small units of exposure and outcome.

So we were interested in incident episodes happening right after that. This wasn't just an artifact of them going off. It turns out the incubation period for a lot of the pathogens that cause gastrointestinal illness and sinus infections-- they have incubation periods around 24 to 30 hours. So a two- to three-day window still captured the vast majority of excess incidents that we would expect. But because of those two considerations, we really went with incidence rates as the main measure in the study.

JADE BENJAMIN- Well, it's been really interesting to hear about the study design, so thank you. And when we **CHUNG:** come back in the second half of this video, we'll talk about any concerns he had or thoughts he had about bias and confounding, and the results of the study. Thanks.

BENJAMIN Thank you.

ARNOLD:

Original Contribution

Acute Illness Among Surfers After Exposure to Seawater in Dry- and Wet-Weather Conditions

Benjamin F. Arnold*, Kenneth C. Schiff, Ayse Ercumen, Jade Benjamin-Chung, Joshua A. Steele, John F. Griffith, Steven J. Steinberg, Paul Smith, Charles D. McGee, Richard Wilson, Chad Nelsen, Stephen B. Weisberg, and John M. Colford, Jr.

* Correspondence to Dr. Benjamin F. Arnold, Division of Epidemiology, School of Public Health, University of California, Berkeley, 101 Haviland Hall, MC #7358, Berkeley, CA 94720-7358 (e-mail: benarnold@berkeley.edu).

Initially submitted September 8, 2016; accepted for publication January 23, 2017.

Rainstorms increase levels of fecal indicator bacteria in urban coastal waters, but it is unknown whether exposure to seawater after rainstorms increases rates of acute illness. Our objective was to provide the first estimates of rates of acute illness after seawater exposure during both dry- and wet-weather periods and to determine the relationship between levels of indicator bacteria and illness among surfers, a population with a high potential for exposure after rain. We enrolled 654 surfers in San Diego, California, and followed them longitudinally during the 2013–2014 and 2014–2015 winters (33,377 days of observation, 10,081 surf sessions). We measured daily surf activities and illness symptoms (gastrointestinal illness, sinus infections, ear infections, infected wounds). Compared with no exposure, exposure to seawater during dry weather increased incidence rates of all outcomes (e.g., for earache or infection, adjusted incidence rate ratio (IRR) = 1.86, 95% confidence interval (CI): 1.27, 2.71; for infected wounds, IRR = 3.04, 95% CI: 1.54, 5.98); exposure during wet weather further increased rates (e.g., for earache or infection, IRR = 3.28, 95% CI: 1.95, 5.51; for infected wounds, IRR = 4.96, 95% CI: 2.18, 11.29). Fecal indicator bacteria measured in seawater (*Enterococcus* species, fecal coliforms, total coliforms) were strongly associated with incident illness only during wet weather. Urban coastal seawater exposure increases the incidence rates of many acute illnesses among surfers, with higher incidence rates after rainstorms.

diarrhea; *Enterococcus*; rain; seawater; waterborne diseases; wound infection

Abbreviations: CI, confidence interval; IRR, incidence rate ratio.

Freshwater runoff after rainstorms increases levels of fecal indicator bacteria measured in seawater (1), but little is known about whether persons who participate in ocean recreation have a higher risk of acute illness after rainstorms. Absent epidemiologic studies to inform beach management guidelines after rainstorms, California beach managers post advisories at beaches that discourage contact with seawater for 72 hours after rainfall—a practice that is based on fecal indicator bacteria profiles in storm water outflows, which typically decline to prerainstorm levels within 3–5 days (2, 3).

In prospective cohorts in California, investigators have found increased incidence of gastrointestinal illness and other acute symptoms (e.g., eye and ear infections) associated with seawater exposure during dry summer months (4–8). In the

same studies, researchers found that levels of fecal indicator bacteria in seawater were positively associated with incident gastrointestinal illness if there was a well-defined source of human fecal contamination impacting the seawater (4–8). Individual cases of acute infections and deaths associated with waterborne pathogens have been reported among surfers in southern California who surfed during or after rainstorms (9), and 2 cross-sectional studies of surfers found that seawater exposure after heavy rainfall increased reported illness (10, 11). To our knowledge, there have been no prospective studies to determine whether rainstorms increase illness among persons who participate in ocean recreation and no studies that have evaluated whether levels of fecal indicator bacteria are associated with incident illness during wet weather periods.

We conducted a longitudinal cohort study among surfers in San Diego, California. We focused on surfers because they are a well-defined population that regularly enters the ocean year-round, even during and immediately after rainstorms, given that surfing conditions often improve during storms (12). Our objectives were to determine whether exposure to seawater increased rates of incident illness among surfers compared with periods when they did not surf in order to determine whether exposure during or immediately after rainstorms increased rates more than did exposure during dry weather. We also sought to evaluate the relationship between levels of fecal indicator bacteria in seawater and incident illness rates during dry and wet weather.

METHODS

Setting

Southern California has one of the most urbanized coastlines in the world, and it receives nearly all of its annual rainfall during the winter months (November–April). San Diego County beaches have some of the best water quality in California based on levels of fecal indicator bacteria, but water quality deteriorates after rainstorms (13). The most heavily used beaches in the region are affected by urban runoff after storms, and local beach managers post advisories that discourage water contact within 72 hours of rainfall. In the present study, we focused enrollment and conducted extensive water quality measurement at 2 monitored beaches within San Diego city

limits—Ocean Beach and Tourmaline Surfing Park. Both monitored beaches have storm-impacted drainage, attract surfers year-round, and have water quality levels similar to those of other beaches in the county (13). Ocean Beach is adjacent to the San Diego river, which drains a 1,088-km² varied land-use watershed with many flow-control structures; Tourmaline Surfing Park is adjacent to Tourmaline Creek and a storm drain, which together drain an urban, largely impervious, 6-km² watershed (Figure 1). The study's technical report includes additional details (14).

Study design and enrollment

We conducted a longitudinal cohort study of surfers recruited in San Diego over 2 winters, with enrollment and follow-up periods chosen to capture most rainfall events in the region. During the first winter (open enrollment from January 14, 2014, to March 18, 2014; end of follow-up on June 4, 2014), we enrolled surfers through in-person interviews at the 2 monitored beaches and through targeted online advertising on [Surfline.com](#), a popular website on which surf conditions are reported. We enrolled participants at monitored beaches and online to assess whether individuals enrolled through these 2 modes were similar in their exposures and other characteristics. Participants enrolled on the beach were very similar to those enrolled online (Table 1), so we exclusively enrolled participants through the study's website during the second winter (open enrollment from December 1, 2014,

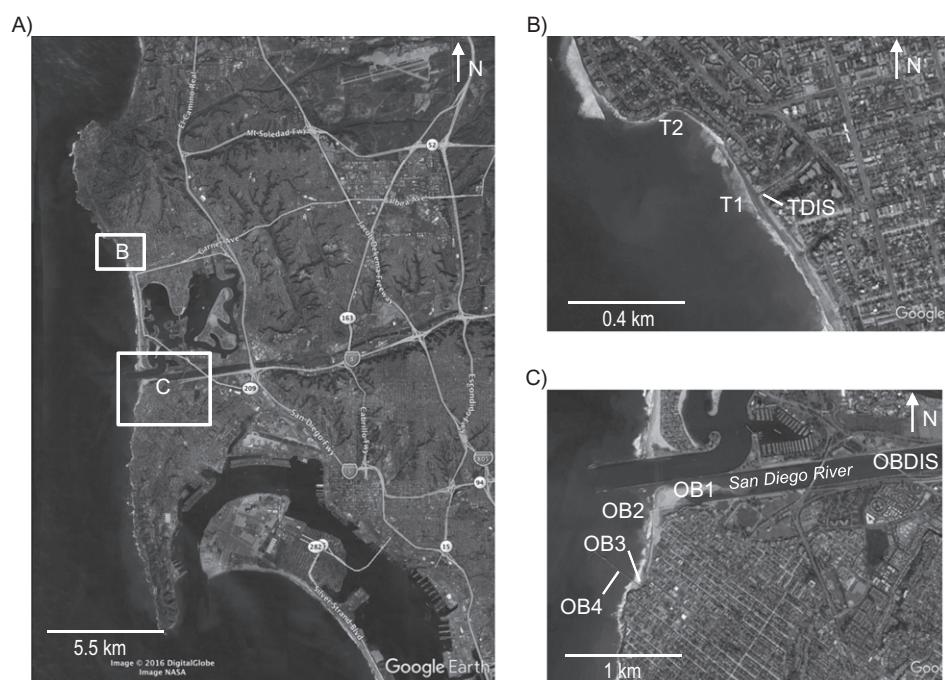


Figure 1. Monitoring beach water quality sampling locations in San Diego, California, winters of 2013–2014 and 2014–2015. Shown are the locations of the 2 monitored beaches along the San Diego coastline (A) and the water quality sampling sites at Tourmaline Surfing Park (B) and Ocean Beach (C). Samples were only collected at Ocean Beach and Tourmaline Surfing Park discharge locations (OBDIS and TDIS, respectively) during wet weather. Wet weather was defined as 0.25 cm or more of rain in 24 hours. T1 and T2, Tourmaline Surfing Park sampling sites 1 and 2; OB1–OB4, Ocean Beach sampling sites 1–4. Map Data: Google, DigitalGlobe, NASA.

Table 1. Characteristics of the Study Population by Mode of Enrollment, San Diego, California, 2013–2015

Characteristic	Beach ^a		Online ^a		Total	
	No.	%	No.	%	No.	%
No. of participants	89		565		654	
Participants with background survey	72	100	535	100	607	100
Age, years ^b						
18–30	35		35		35	
31–40	22		26		26	
41–50	11		16		16	
≥51	29		13		15	
Unreported	3		9		8	
Female sex	19		21		21	
College educated	68		63		63	
Currently employed	74		76		75	
Household income ^b						
<\$15,000	11		6		7	
\$15,000–\$35,000	15		10		11	
\$35,001–\$50,000	11		7		7	
\$50,001–\$75,000	8		13		12	
\$75,001–\$100,000	17		14		14	
\$100,001–\$150,000	17		14		14	
>\$150,000	7		13		12	
Unreported	14		23		22	
Days of surfing per week ^b						
≤1	11		15		14	
2	12		18		17	
3	26		26		26	
4	26		20		21	
≥5	24		18		19	
Unreported	1		3		3	
Chronic health conditions						
Ear problems	12		14		14	
Sinus problems	7		8		8	
Gastrointestinal condition	0		3		2	
Respiratory condition	4		3		3	
Skin condition	1		6		5	
Allergies	10		16		15	
Total days of observation	2,623	100	30,754	100	33,377	100
Days of observation by exposure						
Unexposed	46		47		47	
Dry-weather exposure	48		43		43	
Wet-weather exposure	6		10		10	

^a Beach enrollment only took place during the first winter (2013–2014); online enrollment spanned both winters (2013–2014 and 2014–2015). The study enrolled 73 individuals online during the first winter.

^b Percentages within categories might not sum to 100 because of rounding.

to March 22, 2015; end of follow-up on April 16, 2015). We recruited surfers through postcards distributed at the monitored beaches and through an electronic newsletter distributed

by the Surfrider Foundation's San Diego County chapter. Surfers were eligible if they were 18 years of age or older, could speak and read English, planned to surf in southern California

during the study period, had a valid e-mail address or mobile telephone number, and could access the internet with a computer or smartphone.

Participants completed a brief enrollment questionnaire, and each Tuesday they received a text message or e-mail reminder to complete a short weekly survey. Participants reported daily surf activity (location, date, and times of entry and exit) and illness symptoms (details below) for the previous 7 days using the study's web or smartphone (iOS or Android) application. We used an open cohort design in which participants were allowed to enter and exit the cohort over the follow-up period. We excluded follow-up time during which participants reported surfing outside of southern California. The study protocol was reviewed and approved by the institutional review board at the University of California, Berkeley, and all participants provided informed consent. Participants received a modest incentive for participation (\$20 gift certificate per 4 weekly surveys completed). Web Table 1 (available at <https://academic.oup.com/aje>) includes a Strengthening the Reporting of Observational Studies in Epidemiology checklist.

Outcome definition and measurement

In weekly surveys, participants reported daily records of the following symptoms: diarrhea (defined as ≥ 3 loose/watery stools in 24 hours), sinus pain or infection, earache or infection, infection of an open wound, eye infection, skin rash, and fever. During the second winter, we added sore throat, cough, and runny nose. We created composite outcomes from the symptoms, including: gastrointestinal illness, which was defined as 1) diarrhea, 2) vomiting, 3) nausea and stomach cramps, 4) nausea and missed daily activities due to gastrointestinal illness, or 5) stomach cramps and missed daily activities due to gastrointestinal illness (15); and upper respiratory illness, which was defined as any 2 of the following: 1) sore throat, 2) cough, 3) runny nose, and 4) fever (16). We created a composite outcome of "any infectious symptom" defined as having any 1 of the following: gastrointestinal illness, diarrhea, vomiting, eye infection, infection of open wounds or fever. Our rationale was that it would exclude outcomes that could potentially have noninfectious causes (earache or infection, sinus pain or infection, skin rash, upper respiratory illness) and would capture a broad spectrum of sequelae associated with waterborne pathogens. We defined incident episodes as the onset of symptoms preceded by 6 or more symptom-free days to increase the likelihood that separate episodes represented distinct infections (17, 18).

Exposure definition and measurement

We classified the 3 days after each seawater exposure as exposed periods and all other days of observation as unexposed periods. We defined wet-weather exposure as exposure to seawater within 3 days of 0.25 cm or more of rainfall in a 24-hour period, which is the rainfall criterion used by San Diego County for posting wet-weather beach advisories; we classified all other seawater exposure as dry-weather exposure. We used rainfall measurements from the National Oceanic and Atmospheric Administration Lindbergh Field

Station. Among surfers, most exposure took place during the morning hours, so if a storm's precipitation started after 12:00 PM, we did not classify that day as wet weather (only the following day) to reduce exposure misclassification.

Staff collected daily water samples from January 15, 2014, to March 5, 2014, and from December 2, 2014, to March 31, 2015, at 6 sites across the 2 monitored beaches (Figure 1). Staff collected 1-liter water samples in the morning (08:30 AM \pm 2 hours) just below the water surface (0.5–1.0 meters) in sterilized, sample-rinsed bottles. We sampled discharges during 6 rainstorms immediately upstream from where Tourmaline Creek and the San Diego River discharge to the sea (Figure 1). We tested samples for culturable *Enterococcus* (US Environmental Protection Agency method 1600), fecal coliforms (standard method 9222D), and total coliforms (standard method 9222B). All laboratory analyses met quality-control objectives for absence of background contamination (blanks) and precision (duplicates).

Statistical analysis

We prespecified all analyses (19). Web Appendices 1 and 2 contain statistical details and sample size calculations. In the seawater exposure analysis, we calculated incidence rates by dividing incident episodes by person-days in unexposed and exposed periods during follow-up. If participants missed weekly surveys during follow-up, we did not include those periods in the analysis. We measured the association between seawater exposure and subsequent illness using an incidence rate ratio, which we estimated using a log-linear rate model with robust standard errors to account for repeated observations within individuals (20, 21). To examine illness rates separately for dry- and wet-weather exposures, we created a 3-level categorical exposure that classified each participant's follow-up time into unexposed, dry-weather exposure, and wet-weather exposure periods. We calculated a log-linear test of trend in the incidence rate ratios for dry- and wet-weather exposures (22).

In the fecal indicator association analysis, we estimated the association between levels of fecal indicator bacteria and illness using the subset of surf sessions matched to water-quality indicator measurements at the monitored beaches. We matched daily geometric mean indicator levels to surfers by beach and date (weighted by time in water if recent exposure included multiple days). We modeled the relationship between indicator levels and illness using a log-linear model and estimated the incidence rate ratio associated with a 1– \log_{10} increase in indicator level. We also estimated the incidence rate ratio associated with exposures to water above versus below US Environmental Protection Agency regulatory guidelines (geometric mean *Enterococcus* > 35 colony-forming units per 100 mL) (23) or, in a second definition, if any single sample on the exposure day exceeded 104 colony-forming units per 100 mL. We hypothesized that the relationship between fecal indicator bacteria and illness could be modified by dry- or wet-weather exposure and allowed the exposure-response relationship to vary during dry and wet weather by including an indicator for wet-weather periods and a term for the interaction between indicator bacteria levels and the indicator of wet weather. We controlled for potential confounding (24) from demographic,

exposure-related, and baseline health characteristics (Web Appendix 1). In Web Appendices 3–6 we describe additional analyses, including conversion of estimates to the absolute risk scale, sensitivity analyses, and negative control exposure analyses (25, 26).

RESULTS

Study population

We enrolled 654 individuals who contributed on average 51 days of follow-up (range, 6–139 days). The study population's median age was 34 years (interquartile range, 27–45), and the majority of participants were male (73%), college-educated (63%), and employed (75%) (Table 1). Follow-up included 33,377 person-days of observation after excluding time spent outside of southern California (623 person-days). We excluded from adjusted analyses 47 individuals (1,599 person-days of observation) who provided outcome and exposure information but failed to complete a background questionnaire and thus had missing covariate information.

Water quality and surfer exposure

There were 10 rainstorms with 0.25 cm or more of rain during the study. Field staff collected 1,073 beach water samples and 92 wet-weather discharge samples for fecal indicator bacteria analysis. Median *Enterococcus* levels were higher during wet weather than during dry weather (Figure 2). During follow-up, surfers entered the ocean twice per week on average and experienced 10,081 total days of seawater exposure, including 1,327 days of wet-weather exposure. Surfers were less likely to enter the ocean during or within 1 day of rain. The median ocean entry time was 08:00 AM (interquartile range, 06:45–10:30 AM), and the median time spent in the water was 2 hours (interquartile range, 1–2 hours) (Web Figure 1). Of the 10,081 exposure days, surfers reported wearing a wetsuit during 95%, immersing their head during 96%, and swallowing water during 38%. The most frequented surf locations were the 2 monitored beaches: Tourmaline Surfing Park (25% of surf days) and Ocean Beach (16% of surf days), which reflected targeted enrollment at those beaches (Web Figure 2). There were 5,819 days of observation matched to water-quality measurements at monitored beaches, including 1,358 days during wet weather.

Illness associated with seawater exposure

Seawater exposure in the past 3 days was associated with increased incidence rates of all outcomes except for upper respiratory illness (Web Table 2). Unadjusted and adjusted incidence rate ratio estimates were similar, and for most outcomes, adjusted incidence rate ratios were slightly attenuated toward the null (Web Table 2). With the exception of fever and skin rash, incidence rates increased from unexposed to dry-weather exposure to wet-weather exposure periods (Table 2), a pattern also present on the risk scale (Web Figure 3). Compared with unexposed periods, wet-weather exposure led to the largest relative increase in earaches/infec-

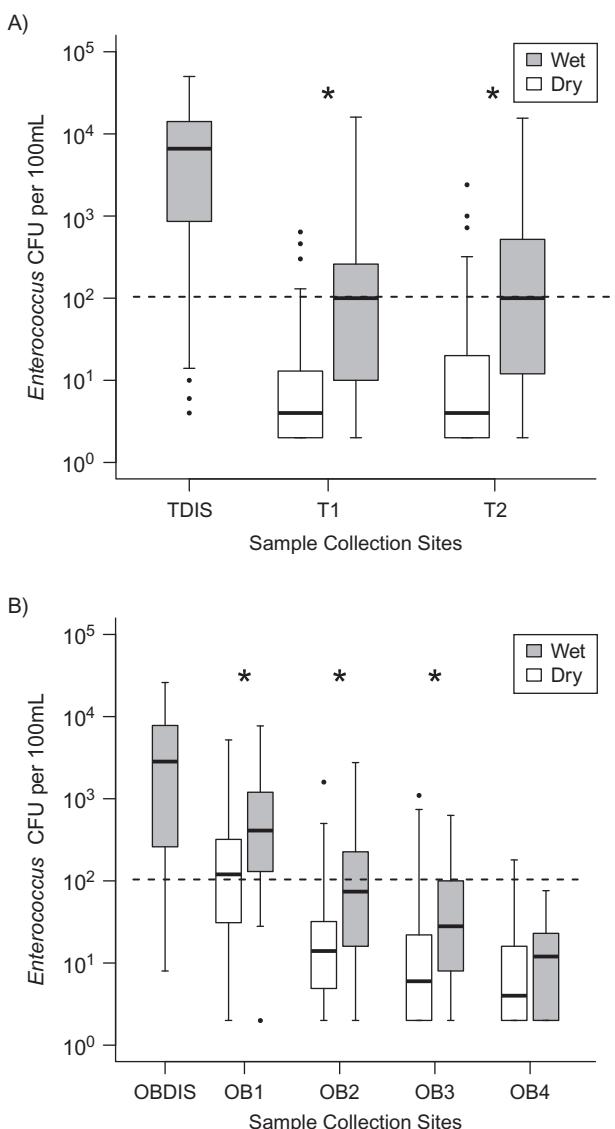


Figure 2. *Enterococcus* levels during dry and wet weather at the sampling locations at Tourmaline Surfing Park (A) and Ocean Beach (B) mapped in Figure 1. Boxes mark interquartile ranges, vertical lines mark 1.5 times the interquartile range, and points mark outliers. Horizontal dashed lines mark the single-sample California recreational water quality guideline (104 CFU/100 mL). Asterisks (*) identify sampling locations with levels that differ between wet and dry periods based on a 2-sample, 2-sided t-test ($P < 0.05$) assuming unequal variances. Samples were only collected at Ocean Beach and Tourmaline Surfing Park discharge locations (OBDIS and TDIS, respectively) during wet weather. Wet weather was defined as 0.25 cm or more of rain in 24 hours. CFU, colony-forming units; T1 and T2, Tourmaline Surfing Park sampling sites 1 and 2; OB1–OB4, Ocean Beach sampling sites 1–4.

tions (Table 3; adjusted incidence rate ratio (IRR) = 3.28, 95% confidence interval (CI): 1.95, 5.51) and infection of open wounds (Table 3; adjusted IRR: 4.96, 95% CI: 2.18, 11.29). Sensitivity analyses that shortened the wet-weather window increased the difference between dry- and wet-weather incidence rates for most outcomes (Web Figure 4).

Table 2. Incidence Rates Among Surfers by Type of Seawater Exposure, San Diego, California, 2013–2015

Outcome	Unexposed Periods			Dry-Weather Exposure			Wet-Weather Exposure ^a		
	No. of Episodes	No. of Days at Risk	Rate per 1,000	No. of Episodes	No. of Days at Risk	Rate per 1,000	No. of Episodes	No. of Days at Risk	Rate per 1,000
Gastrointestinal illness	90	14,884	6.0	116	13,769	8.4	31	3,037	10.2
Diarrhea	75	15,086	5.0	88	13,909	6.3	27	3,061	8.8
Sinus pain or infection	109	14,475	7.5	139	13,391	10.4	37	2,998	12.3
Earache or infection	59	14,931	4.0	111	13,618	8.2	37	3,008	12.3
Infection of open wound	14	15,456	0.9	30	14,080	2.1	11	3,119	3.5
Skin rash	42	15,024	2.8	66	13,750	4.8	15	3,007	5.0
Fever	51	15,156	3.4	69	14,138	4.9	6	3,152	1.9
Upper respiratory illness ^b	117	12,001	9.7	111	11,025	10.1	31	2,543	12.2
Any infectious symptom ^c	138	14,445	9.6	181	13,176	13.7	47	2,926	16.1

^a Defined as entering the sea within 3 days of 0.25 cm or more of rain in 24 hours.^b Only measured in year 2 of the study.^c Includes gastrointestinal illness, eye infections, infected wounds, and fever.

Illness associated with fecal indicator bacteria levels

Enterococcus, total coliform, and fecal coliform levels were positively associated with increased incidence of almost all outcomes during the study (Web Table 3). Rainfall was a strong effect modifier of the association (Table 4). During dry weather, there was no association between *Enterococcus* levels and illness except for infected wounds, but *Enterococcus* was strongly associated with illness after wet-weather exposure (e.g., for each \log_{10} increase, gastrointestinal illness IRR = 2.17, 95% CI: 1.16, 4.03; Table 4, Web Figure 5, and Web Table 4). Associations were attenuated in adjusted analyses, but relationships were similar (e.g., for gastrointestinal illness, wet-weather IRR = 1.75, 95% CI: 0.80, 3.84; Table 4). There was evidence for excess risk of gastrointestinal illness at higher *Enterococcus* levels only during wet-weather periods (Web Figure 6): The predicted excess risk that corresponded to the current US Environmental Protection Agency regulatory guideline of 35 colony-forming units per 100 mL was 16 episodes per 1,000 (95% CI: 5, 27). Negative control analyses showed no consistent association between fecal indicator bacteria and illness among participants during periods in which they had no recent seawater contact (Web Table 5).

DISCUSSION

Key results

To our knowledge, this is the first prospective cohort study in which the association between incident illness and exposure to seawater in wet weather has been measured, and the findings represent novel empirical measures of incident illness associated with storm water discharges. There was a consistent increase in acute illness incidence rates between unexposed, dry-weather, and wet-weather exposure periods (Tables 2 and 3). Rainstorms led to higher levels of fecal indicator bacteria (Figure 2), and a sensitivity analysis illustrated that a 2–3 day window after rainstorms captured the majority of excess incidence associated with wet-weather ex-

posure (Web Figure 4). Fecal indicator bacteria matched to individual surf sessions were strongly associated with illness only during wet weather periods (Table 4, Web Figure 5).

Interpretation

Swimmers are more rare during the winter months, and surfers' frequent and intense exposure made them an ideal population in which to study the relationship between illness and exposure to seawater in wet weather (27). The associations estimated in this study may not reflect those of the general population, but among a highly exposed subgroup of athletes, our results measure the illness associated with seawater exposure after rainstorms in southern California. Enrolling surfers led to some important differences between the present study population and most swimmer cohorts. We enrolled adults because we could not guarantee adequate consent for minors through online enrollment, whereas swimmer cohorts have historically enrolled predominantly families with children (28); children are more susceptible and have greater risk than do adult swimmers (15). Participants surfed twice per week for 2 hours each session, with nearly universal head immersion (96% of exposures) and frequent water ingestion (38% of exposures). This far exceeds exposure levels recorded in swimmer cohorts. Likely because of surfers' repeated exposures to pathogens in seawater, studies have found higher levels of immunity to hepatitis A and more frequent gut colonization by antibiotic-resistant *Escherichia coli* among surfers than among the general population (29, 30).

Despite surfers' intense and frequent exposures, gastrointestinal illness rates observed in the present study were similar to those measured among beachgoers California cohorts in the summer (Web Appendix 6, Web Figure 7), and the increase in gastrointestinal illness rates associated with seawater exposure (adjusted IRR = 1.33, 95% CI: 0.99, 1.78; Web Table 2) was similar to estimates measured in marine swimmer cohorts in California and elsewhere in the United States (15, 31). However, the 3-fold increase in rates of

Table 3. Incidence Rate Ratios for Surfer Illnesses Within 3 Days of Dry- and Wet-Weather Seawater Exposure Compared With Unexposed Periods, San Diego, California, 2013–2015

Outcome	Unadjusted ^a				Adjusted ^{a,b}			
	Dry Weather		Wet Weather ^c		Dry Weather		Wet Weather ^c	
	IRR	95% CI	IRR	95% CI	IRR	95% CI	IRR	95% CI
Gastrointestinal illness	1.39	1.05, 1.86	1.69	1.10, 2.59	1.30	0.95, 1.76	1.41	0.92, 2.17
Diarrhea	1.27	0.92, 1.76	1.77	1.11, 2.83	1.22	0.86, 1.73	1.51	0.95, 2.41
Sinus pain or infection	1.38	1.05, 1.80	1.64	1.12, 2.40	1.23	0.93, 1.64	1.51	1.01, 2.26
Earache or infection	2.06	1.47, 2.90	3.11	1.94, 4.98	1.86	1.27, 2.71	3.28	1.95, 5.51
Infection of open wound	2.35	1.27, 4.36	3.89	1.83, 8.30	3.04	1.54, 5.98	4.96	2.18, 11.29
Skin rash	1.72	1.16, 2.54	1.78	0.98, 3.24	1.64	1.11, 2.41	1.80	0.97, 3.35
Fever	1.45	0.99, 2.12	0.57	0.24, 1.31	1.56	1.04, 2.34	0.64	0.27, 1.52
Upper respiratory illness ^d	1.03	0.79, 1.35	1.25	0.84, 1.86	1.04	0.79, 1.36	1.17	0.79, 1.74
Any infectious symptom ^e	1.44	1.14, 1.82	1.68	1.19, 2.38	1.50	1.17, 1.92	1.62	1.14, 2.30

Abbreviations: CI, confidence interval; IRR, incidence rate ratio.

^a Unadjusted and adjusted incidence rate ratios compare incidence rates in the 3 days after seawater exposure during dry or wet weather with incidence rates during unexposed periods. Table 2 includes the underlying data. Tests of trend in the IRR between exposure categories are significant ($P < 0.05$) if the confidence interval for wet-weather exposure excludes 1.0 (22).

^b We controlled for the following time-invariant potential confounders: age, sex, educational level, employment status, household income, years the individual had surfed, reported behavior of typically avoiding the ocean after wet weather, surfboard length, mode of enrollment (beach vs. Internet). We controlled for chronic health conditions only for the corresponding outcomes: ear problems, sinus problems, gastrointestinal conditions, respiratory conditions, skin conditions. We also controlled for the following time-varying potential confounders: entered the ocean for an activity other than surfing, any illness symptoms in the week preceding the risk window, day of recall, day of the week, and rainfall total during the past 3 days.

^c Defined as entering the sea within 3 days of 0.25 cm or more of rain in 24 hours.

^d Only measured in year 2 of the study.

^e Includes gastrointestinal illness, eye infections, infected wounds, and fever.

earache/infection and 5-fold increase in infected open wounds associated with exposure after rainstorms (Table 3) are stronger associations than have been reported in previous studies, and they provide evidence for increased incidence of a broad set of infectious symptoms after seawater exposure within 3 days of rain.

Fecal indicator bacteria were a reliable marker of human illness risk in this setting only within 3 days of rainfall (Table 4). Our results are consistent with summer studies in California in which investigators found associations between *Enterococcus* levels and illness only if there was a well-defined source of human fecal contamination (4–8). Our findings are also consistent with model predictions of higher gastrointestinal illness risk among southern California surfers after storms (32). Molecular testing for pathogens in storm water discharge to study monitored beaches identified near-ubiquitous presence of norovirus and *Campylobacter* species, and models parameterized with pathogen measurements predicted higher illness risk after rainstorms (14). The association between fecal indicator bacteria measured during wet weather and a range of nonenteric illnesses, such as sinus pain or infection and fever (Table 4), suggests that fecal indicator bacteria may mark broader bacterial or viral pathogen contamination in seawater after rainstorms.

Some study outcomes could have noninfectious causes associated with surfing. Earache and sinus pain can result

from physical incursion of saltwater through surfing's high-intensity exposure, ingestion of saltwater can cause gastrointestinal symptoms, and wetsuit use could cause skin rashes. If the association between surf exposure and symptoms resulted from noninfectious causes, we would expect similar incidence rates after wet- and dry-weather exposures. This was observed for skin rash, but incidence rates for sinus, ear, and gastrointestinal illnesses were higher after wet-weather exposure (Table 2), and the strong association between fecal indicator bacteria and fever during wet-weather conditions was consistent with an infectious etiology (Table 4).

It is also possible that some infections acquired during surfing could result from nonanthropogenic sources. The ocean was warmer than usual during the second winter because of a weak El Niño, which caused conditions favorable to naturally occurring *Vibrio parahaemolyticus* and toxin-producing marine algae that can cause human illness (33). Wound infection was the single outcome strongly associated with fecal indicator bacteria measured during dry weather (Table 4), an observation consistent with a pathogen source like *V. parahaemolyticus* that covaries with fecal indicator bacteria even in nonstorm conditions. Yet, the consistently higher rates of infected wounds and other symptoms after wet-weather exposure compared with dry-weather exposure (Tables 2 and 3) suggests that storm water runoff impacted by anthropogenic sources constitutes an important pathogen source in this setting.

Table 4. Surfer Illness Associated With a log₁₀ Increase in Fecal Indicator Bacteria Levels, Stratified by Exposure During Dry and Wet Weather, Tourmaline Surfing Park and Ocean Beach, San Diego, California, 2013–2015

Fecal Indicator Bacteria and Illness Symptom	Dry Weather		Wet Weather		Dry Weather		Wet Weather		Dry Weather		Wet Weather		P Value ^b	
	Episodes	Days at Risk	Episodes	Days at Risk	Dry Weather		Wet Weather		Dry Weather		Wet Weather			
					IRR	95% CI	IRR	95% CI	IRR	95% CI	IRR	95% CI		
Enterococcus														
Gastrointestinal illness	30	4,251	10	1,297	0.86	0.47, 1.58	2.17	1.16, 4.03	0.04	0.85	0.46, 1.56	1.75	0.80, 3.84	
Diarrhea	24	4,285	9	1,305	1.13	0.62, 2.07	2.38	1.27, 4.46	0.11	1.16	0.63, 2.14	2.00	0.92, 4.32	
Sinus pain or infection	44	4,130	19	1,262	1.34	0.79, 2.26	1.93	1.17, 3.19	0.33	0.96	0.53, 1.76	1.61	0.96, 2.69	
Earache or infection	38	4,233	14	1,274	0.74	0.37, 1.47	1.23	0.50, 3.02	0.38	0.70	0.35, 1.40	1.32	0.51, 3.41	
Infection of open wound	19	4,360	6	1,332	2.69	1.05, 6.90	2.24	0.65, 7.69	0.83	2.79	1.12, 6.95	2.94	0.79, 10.97	
Skin rash	19	4,230	5	1,267	1.46	0.68, 3.14	0.89	0.21, 3.82	0.56	1.09	0.42, 2.80	0.51	0.06, 4.04	
Fever	22	4,366	2	1,342	1.33	0.69, 2.56	3.29	2.35, 4.59	0.01	1.29	0.66, 2.52	3.53	2.37, 5.24	
Upper respiratory illness ^c	37	3,679	15	1,090	0.89	0.55, 1.45	1.94	0.85, 4.42	0.10	0.74	0.44, 1.25	1.89	0.87, 4.11	
Any infectious symptom ^d	50	4,080	17	1,264	1.12	0.69, 1.83	2.51	1.49, 4.24	0.04	1.06	0.64, 1.76	2.52	1.41, 4.50	
Total coliforms														
Gastrointestinal illness	30	4,251	10	1,297	0.82	0.42, 1.61	2.96	1.50, 5.83	0.01	0.76	0.38, 1.54	2.59	1.02, 6.56	
Diarrhea	24	4,285	9	1,305	1.04	0.53, 2.04	3.34	1.72, 6.47	0.02	1.05	0.51, 2.16	3.20	1.31, 7.85	
Sinus pain or infection	44	4,130	19	1,262	1.57	0.87, 2.84	2.18	1.11, 4.26	0.48	0.75	0.35, 1.58	1.52	0.62, 3.73	
Earache or infection	38	4,233	14	1,274	0.83	0.39, 1.76	1.46	0.63, 3.39	0.29	0.99	0.51, 1.92	1.59	0.84, 3.01	
Infection of open wound	19	4,360	6	1,332	2.76	0.91, 8.36	2.67	0.85, 8.41	0.97	3.21	1.03, 10.03	4.12	0.95, 17.91	
Skin rash	19	4,230	5	1,267	1.69	0.72, 3.99	1.03	0.24, 4.43	0.56	1.18	0.39, 3.56	0.54	0.09, 3.06	
Fever	22	4,366	2	1,342	1.15	0.49, 2.70	4.99	3.19, 7.79	0.00	1.16	0.49, 2.73	6.22	3.88, 9.96	
Upper respiratory illness ^c	37	3,679	15	1,090	0.97	0.50, 1.89	2.33	0.75, 7.23	0.19	0.73	0.38, 1.40	2.03	0.70, 5.89	
Any infectious symptom ^d	50	4,080	17	1,264	1.17	0.69, 1.97	3.21	1.84, 5.58	0.01	1.11	0.65, 1.91	3.42	1.76, 6.66	
Total coliforms														
Gastrointestinal illness	30	4,251	10	1,297	0.77	0.40, 1.47	2.62	1.63, 4.24	0.01	0.83	0.42, 1.63	1.96	1.22, 3.15	
Diarrhea	24	4,285	9	1,305	0.66	0.29, 1.51	2.59	1.53, 4.38	0.02	0.78	0.35, 1.70	1.99	1.19, 3.35	
Sinus pain or infection	44	4,130	19	1,262	1.52	0.84, 2.77	2.02	1.04, 3.93	0.55	1.08	0.54, 2.19	1.79	0.93, 3.44	
Earache or infection	38	4,233	14	1,274	1.03	0.54, 1.96	1.67	0.63, 4.41	0.40	0.92	0.46, 1.82	1.72	0.64, 4.61	
Infection of open wound	19	4,360	6	1,332	3.46	0.79, 15.20	2.16	0.46, 10.16	0.69	4.02	0.91, 17.67	2.38	0.60, 9.43	
Skin rash	19	4,230	5	1,267	1.58	0.73, 3.40	1.14	0.34, 3.81	0.65	1.30	0.48, 3.53	1.11	0.28, 4.41	
Fever	22	4,366	2	1,342	1.59	0.78, 3.22	7.48	4.28, 13.08	0.00	1.62	0.77, 3.37	9.24	4.64, 18.41	
Upper respiratory illness ^c	37	3,679	15	1,090	0.87	0.49, 1.52	2.04	0.84, 4.96	0.12	0.72	0.40, 1.30	1.87	0.84, 4.19	
Any infectious symptom ^d	50	4,080	17	1,264	1.35	0.78, 2.34	3.26	1.76, 6.01	0.06	0.69	0.23, 2.07	3.02	1.56, 5.38	

Abbreviations: CI, confidence interval; IRR, incidence rate ratio.

^a We controlled for the following time-invariant potential confounders: age, sex, educational level, employment status, household income, years the individual had surfed, reported behavior of typically avoiding the ocean after wet weather, surfboard length, mode of enrollment (beach vs. Internet). We controlled for chronic health conditions only for the corresponding outcomes: ear problems, sinus problems, gastrointestinal conditions, respiratory conditions, skin conditions. We also controlled for the following time-varying potential confounders: entered the ocean for any activity other than surfing, any illness symptoms in the week preceding the risk window, day of recall, day of the week, and rainfall total during the past 3 days.

^b P value for multiplicative effect modification of dry versus wet weather.

^c Only measured in year 2 of the study.

^d Includes gastrointestinal illness, eye infections, infected wounds, and fever.

Limitations

The use of self-reported symptoms could bias the association between seawater exposure and illness away from the null if surfers overreported illness after exposure; conversely, random (nondifferential) errors in exposures or outcomes could bias associations toward the null (34). The survey measured daily exposure and outcomes in separate modules—an intentional decision to separate the measurements and inhibit systematic reporting bias. Adjusted analyses controlled for day of recall and day of the week to reduce nondifferential bias from recall errors but would not control for systematic bias. Negative control exposure analyses found no association between *Enterococcus* levels and illness on days with no recent water exposure (Web Table 5), which suggests that unmeasured confounding or reporting bias is unlikely to explain the association between *Enterococcus* levels and illness. Moreover, the use of daily average levels of fecal indicator bacteria could bias the association between water quality and illness toward the null if the averaging resulted in nondifferential misclassification error (35).

We measured incident outcomes within 3 days of seawater exposure because the population regularly entered the ocean, a 3-day period captures the incubation period for the most common waterborne pathogens (e.g., norovirus, *Campylobacter* species, *Salmonella* species) (36), and past studies found that most excess episodes of gastrointestinal illness associated with seawater exposure occurred in the first 1–2 days (15). Illness caused by waterborne pathogens with longer incubation periods (e.g., *Cryptosporidium* species) (37) could have been misclassified in this study, which could bias results toward the null by artificially increasing incidence rates in unexposed periods and decreasing rates in exposed periods.

Conclusions

Surfing was associated with increased incidence of several categories of symptoms, and associations were stronger if surfing took place shortly after rainstorms. Higher levels of fecal indicator bacteria were strongly associated with fever, sinus pain/infection, wound infection, and gastrointestinal symptoms within 3 days of rainstorms. The internal consistency between water-quality measurements, patterns of illness after dry- and wet-weather exposures, and incidence profiles with time since rainstorms lead us to conclude that seawater exposure during or close to rainstorms at beaches impacted by urban runoff in southern California increases the incidence rates of a broad set of acute illnesses among surfers. These findings provide strong evidence to support the posting of beach warnings after rainstorms and initiatives that would reduce pathogen sources in urban runoff that flows to coastal waters.

ACKNOWLEDGMENTS

Author affiliations: Division of Epidemiology, School of Public Health, University of California, Berkeley, California (Benjamin F. Arnold, Ayse Ercumen, Jade Benjamin-

Chung, John M. Colford, Jr.); Southern California Coastal Water Research Project, Costa Mesa, California (Kenneth C. Schiff, Joshua A. Steele, John F. Griffith, Steven J. Steinberg, Paul Smith, Stephen B. Weisberg); Orange County Sanitation District, Fountain Valley, California (Charles D. McGee; retired); and Surfrider Foundation, San Clemente, California (Richard Wilson, Chad Nelsen).

The study was funded by the city and county of San Diego, California.

We thank the field team members who enrolled participants at the beach and collected water samples throughout the study. We also thank Laila Othman, Sonji Romero, Aaron Russell, Joseph Toctocan, Laralyn Asato, Zaira Valdez, and the staff at City of San Diego Marine Microbiology Laboratory who generously provided laboratory space to test water specimens, and Jeffrey Soller, Mary Schoen, and members of the study's external advisory committee for earlier comments on the results.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interest: none declared.

REFERENCES

- Noble RT, Weisberg SB, Leecaster MK, et al. Storm effects on regional beach water quality along the southern California shoreline. *J Water Health*. 2003;1(1):23–31.
- Leecaster MK, Weisberg SB. Effect of sampling frequency on shoreline microbiology assessments. *Mar Pollut Bull*. 2001; 42(11):1150–1154.
- Ackerman D, Weisberg SB. Relationship between rainfall and beach bacterial concentrations on Santa Monica bay beaches. *J Water Health*. 2003;1(2):85–89.
- Haile RW, Witte JS, Gold M, et al. The health effects of swimming in ocean water contaminated by storm drain runoff. *Epidemiology*. 1999;10(4):355–363.
- Colford JM, Wade TJ, Schiff KC, et al. Water quality indicators and the risk of illness at beaches with nonpoint sources of fecal contamination. *Epidemiology*. 2007;18(1): 27–35.
- Colford JM, Schiff KC, Griffith JF, et al. Using rapid indicators for *Enterococcus* to assess the risk of illness after exposure to urban runoff contaminated marine water. *Water Res*. 2012; 46(7):2176–2186.
- Arnold BF, Schiff KC, Griffith JF, et al. Swimmer illness associated with marine water exposure and water quality indicators: impact of widely used assumptions. *Epidemiology*. 2013;24(6):845–853.
- Yau VM, Schiff KC, Arnold BF, et al. Effect of submarine groundwater discharge on bacterial indicators and swimmer health at Avalon Beach, CA, USA. *Water Res*. 2014;59:23–36.
- Taylor K. Contagion Present. Surfer Magazine. <http://www.surfermag.com/features/contagion-present>. Published July 20, 2016. Accessed August 17, 2016.
- Dwight RH, Baker DB, Semenza JC, et al. Health effects associated with recreational coastal water use: urban versus rural California. *Am J Public Health*. 2004;94(4):565–567.
- Harding AK, Stone DL, Cardenas A, et al. Risk behaviors and self-reported illnesses among Pacific Northwest surfers. *J Water Health*. 2015;13(1):230–242.

12. Stormsurf. Weather basics. <http://www.stormsurf.com/page2/tutorials/weatherbasics.shtml>. Published September 26, 2003. Accessed October 27, 2016.
13. Heal the Bay. Heal the Bay's 2014-2015 Annual Beach Report Card. Santa Monica, CA: Heal the Bay; 2015. http://www.healthebay.org/sites/default/files/BRC_2015_final.pdf. Accessed December 5, 2016.
14. Schiff K, Griffith J, Steele J, et al. The Surfer Health Study: A Three-Year Study Examining Illness Rates Associated With Surfing During Wet Weather. Costa Mesa, CA: Southern California Coastal Water Research Project; 2016. http://ftp.sccwrp.org/pub/download/DOCUMENTS/TechnicalReports/943_SurferHealthStudy.pdf. Published September 20, 2016. Accessed December 5, 2016.
15. Arnold BF, Wade TJ, Benjamin-Chung J, et al. Acute gastroenteritis and recreational water: highest burden among young US children. *Am J Public Health*. 2016;106(9):1690–1697.
16. Wade TJ, Sams E, Brenner KP, et al. Rapidly measured indicators of recreational water quality and swimming-associated illness at marine beaches: a prospective cohort study. *Environ Health*. 2010;9:66.
17. Colford JM, Wade TJ, Sandhu SK, et al. A randomized, controlled trial of in-home drinking water intervention to reduce gastrointestinal illness. *Am J Epidemiol*. 2005;161(5):472–482.
18. Colford JM, Hilton JF, Wright CC, et al. The Sonoma Water Evaluation Trial: a randomized drinking water intervention trial to reduce gastrointestinal illness in older adults. *Am J Public Health*. 2009;99(11):1988–1995.
19. Arnold B, Ercumen A. The Surfer Health Study. Open Science Framework. <https://osf.io/hvn78>. Published July 29, 2015. Updated July 29, 2016. Accessed December 5, 2016.
20. Rothman KJ, Greenland S, Lash TL. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams and Wilkins; 2008.
21. Zou G. A modified poisson regression approach to prospective studies with binary data. *Am J Epidemiol*. 2004;159(7):702–706.
22. Vittinghoff E, Glidden DV, Shiboski SC, et al. *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. 2nd ed. New York, NY: Springer Science & Business Media; 2012.
23. United States Environmental Protection Agency. *Recreational Water Quality Criteria*. Washington, DC: United States Environmental Protection Agency; 2012. (Office of Water publication no. 820-F-12-058). <https://www.epa.gov/sites/production/files/2015-10/documents/rwqc2012.pdf>. Accessed January 24, 2017.
24. VanderWeele TJ, Shpitser I. A new criterion for confounder selection. *Biometrics*. 2011;67(4):1406–1413.
25. Lipsitch M, Tchetgen Tchetgen E, Cohen T. Negative controls: a tool for detecting confounding and bias in observational studies. *Epidemiology*. 2010;21(3):383–388.
26. Arnold BF, Ercumen A, Benjamin-Chung J, et al. Brief report: negative controls to detect selection bias and measurement bias in epidemiologic studies. *Epidemiology*. 2016;27(5):637–641.
27. Rothman KJ, Gallacher JE, Hatch EE. Why representativeness should be avoided. *Int J Epidemiol*. 2013;42(4):1012–1014.
28. Wade TJ, Pai N, Eisenberg JN, et al. Do US Environmental Protection Agency water quality guidelines for recreational waters prevent gastrointestinal illness? A systematic review and meta-analysis. *Environ Health Perspect*. 2003;111(8):1102–1109.
29. Gammie A, Morris R, Wyn-Jones AP. Antibodies in crevicular fluid: an epidemiological tool for investigation of waterborne disease. *Epidemiol Infect*. 2002;128(2):245–249.
30. Leonard A. *Are Bacteria in the Coastal Zone a Threat to Human Health?* [dissertation]. Exeter, UK: University of Exeter; 2016. <https://ore.exeter.ac.uk/repository/handle/10871/22805>. Accessed October 14, 2016.
31. Fleisher JM, Fleming LE, Solo-Gabriele HM, et al. The BEACHES Study: health effects and exposures from non-point source microbial contaminants in subtropical recreational marine waters. *Int J Epidemiol*. 2010;39(5):1291–1298.
32. Tseng LY, Jiang SC. Comparison of recreational health risks associated with surfing and swimming in dry weather and post-storm conditions at Southern California beaches using quantitative microbial risk assessment (QMRA). *Mar Pollut Bull*. 2012;64(5):912–918.
33. Van Dolah FM. Marine algal toxins: origins, health effects, and their increased occurrence. *Environ Health Perspect*. 2000;108(suppl 1):133–141.
34. Copeland KT, Checkoway H, McMichael AJ, et al. Bias due to misclassification in the estimation of relative risk. *Am J Epidemiol*. 1977;105(5):488–495.
35. Fleisher JM. The effects of measurement error on previously reported mathematical relationships between indicator organism density and swimming-associated illness: a quantitative estimate of the resulting bias. *Int J Epidemiol*. 1990;19(4):1100–1106.
36. Widdowson MA, Sulka A, Bulens SN, et al. Norovirus and foodborne disease, United States, 1991-2000. *Emerg Infect Dis*. 2005;11(1):95.
37. Jokipii L, Jokipii AM. Timing of symptoms and oocyst excretion in human cryptosporidiosis. *N Engl J Med*. 1986;315(26):1643–1647.