



PHW250B Week 4 Reader

Topic 1: Appropriate Measures of Association for Case-Control Studies

| | |
|---|---|
| Lecture: Measures of Association for Case-Control Studies. | 2 |
| Langholz. Case-Control Studies = Odds Ratios Blame the Retrospective Model. Epidemiology. 21(1)10-12. 16 | |

Topic 2: Population Attributable Fractions and Related Measures of Association

| | |
|---|----|
| Lecture: Use and Misuse of Population Attributable Fractions. | 19 |
| Rockhill et al. Use and Misuse of Population Attributable Fractions. Am J Public Health 1998; 88(1). | 31 |

Topic 3: The Hazards of Hazard Ratios

| | |
|--|----|
| Lecture: Hazard Ratios. | 36 |
| Hernan. The Hazards of Hazard Ratios. Epidemiology. 21(1)13-15. | 52 |

Topic 4: Causal Perspective on Measures of Association

| | |
|---|----|
| Lecture: Causal Perspective on Measures of Association..... | 56 |
|---|----|

Journal Club

| | |
|--------------------------|----|
| Arnold et al. 2017..... | 68 |
| Manson et al. 2003. | 78 |

Lecture: Measures of Association for Case-Control Studies



Appropriate measures of association for case-control studies

PHW250 F - Jack Colford

JACK COLFORD: This lesson will focus on the appropriate measures of association to use in case-control studies.

Odds ratios - a mathematical convenience

- For many years, epidemiologists were taught that the only appropriate measure of association in a case-control study is an odds ratio.
 - This is because case-control studies sample from populations based on disease status.
- The mathematical equivalence of the odds of exposure and the odds of disease conveniently allowed epidemiologists to calculate a measure of association that did not require them to incorrectly pool across the diseased and non-diseased in a 2x2 table.

$$\text{Odds}_{\text{exp cases}} = \frac{\frac{a}{a+c}}{1 - \left(\frac{a}{a+c} \right)} = \frac{a}{c}$$
$$\text{OR}_{\text{exp}} = \frac{\frac{a}{c}}{\frac{b}{d}} = \frac{a \times d}{b \times c} = \frac{\frac{a}{b}}{\frac{c}{d}} = \text{OR}_{\text{dis}}$$
$$\text{Odds}_{\text{exp controls}} = \frac{\frac{b}{b+d}}{1 - \left(\frac{b}{b+d} \right)} = \frac{b}{d}$$

Szklo & Nieto, 3rd Ed. Chapter 3



For many years, epidemiologists were taught that the only appropriate measure of association for a case control study was an odds ratio. And that was felt to be because case control studies sample from populations based on disease status rather than exposure status. But the mathematical equivalents of the odds of exposure and the odds of disease conveniently allowed epidemiologists to calculate a measure of association that did not require them to incorrectly pool across the disease and non-disease in a two by two table.

So for example, if we have the odds among the exposed cases, shown in the formula here to simplify to a over and the odds among the exposed controls shown to simplify in the formula here to b over d, then the odds ratio for exposure would be a over c divided by b over d, which could be written as a times d over b times c. All of which would simplify down to the same thing as the odds ratio among the diseased. So the odds ratio among the exposed for disease and the odds ratio among disease for exposure was equivalent.

Odds ratios and the rare disease assumption

- Odds ratios are not naturally intuitive — usually what we're really interested in is the relative risk.
- When the disease of interest is rare, the odds ratio approximates the relative risk.
- $$RR = \frac{OR}{1 - [R_U - (OR \times R_U)]}$$
- R_U = risk or incidence in the unexposed
- As R_U approaches zero, $RR \cong OR$
- Often case-control studies are used to study rare diseases, so this conveniently meant that the OR could approximate the RR in most case-control studies.

What is a "rare" disease?
Typically a disease with incidence < 5% (or 5 per 1000) is considered rare. When incidence is > 10%, the OR and RR differ from each other substantially.

Szklo & Nieto, 3rd Ed, Chapter 3

Berkeley School of Public Health

Odds ratios are not naturally intuitive. Usually, what we're really interested in is the relative risk. When the disease of interest is rare, the odds ratio approximates the relative risk because the relative risk formula as given here is the odds ratio over 1 minus the quantity of the risk of the unexposed minus the odds ratio times the risk in the unexposed, where the risk in the unexposed is represented by the term r_{sub-u} . As r_{sub-u} approaches zero, look at what happens to the overall formula. The relative risk approximates the odds ratio, because the formula reduces to just odds ratio over one. So often case control studies are used to study rare diseases. So this has conveniently meant that the odds ratio could approximate the relative risk in most case control studies. So typically a disease is considered rare when its incidence is less than 5%. When the incidence is greater than 10%, the odds ratio and relative risk differ from each other substantially.

The modern view on odds ratios and case-control studies

- Today, epidemiologists learn that “in theory, every case-control study takes place within a cohort” (Wacholder et al., 1992).
- As a result, to determine the appropriate measure of association in a case-control study, it is necessary to consider the way in which cases and controls were sampled from the target population.
 - (We'll learn more about this in the case-control study unit.)
- In some case-control studies, it is not necessary to make the rare disease assumption when estimating an odds ratio.
 - Example: when the control group is a random sample of the target population



Today, epidemiologists learn that in theory every case control study takes place within a cohort, quote unquote. So as a result, to determine the appropriate measure of association in a case control study, it's necessary to consider the way in which cases and controls were sampled from the target population. We'll learn more about this in a case control study unit. In some case control studies, it's not necessary to make the rare disease assumption at all when estimating the odds ratio. For example, when the control group is a random sample of the target population, it's not necessary to make that rare disease assumption.

Case-control designs in which the OR approximates the RR

- The odds ratio directly estimates the relative risk (no rare disease assumption needed) when the control group is a sample of the total reference population.
 - This is true when the case-control study is defined within a cohort.
 - These are sometimes referred to as "population controls".
- Two designs achieve this:
 - Case-cohort study
 - Nested case-control study

Szklo & Nieto, 3rd Ed, Chapter 3



Let's discuss some situations in which the design of a case control study helps us use the odds ratio to approximate the relative risk. The odds ratio directly estimates the relative risk, without any requirement for a rare disease assumption, when the control group is a sample of the total reference population. This situation is true when the case control study is defined within a cohort. These are sometimes referred to as population controls in this situation, and there are two specific designs that achieve this. The first is called the case cohort study, and the second is called the nested case control study.

Case-control designs in which the OR approximates the RR

- **Case-cohort study**
 - Cases: individuals who develop disease in the cohort
 - Controls: study population (or random sample of study population) at baseline (including future cases)
- The division of the odds of exposure in cases by that in controls yields the relative risk (specifically, the cumulative incidence ratio):

$$OR_{exp} = \frac{Odds_{exp\ cases}}{Odds_{exp\ total\ population}} = \frac{\left(\frac{a}{c}\right)}{\left(\frac{a+b}{c+d}\right)} = \frac{\left(\frac{a}{a+b}\right)}{\left(\frac{c}{c+d}\right)} = RR \quad (\text{Eq. 3.14})$$

TABLE 3-9 Cross-tabulation of a defined population by exposure and disease development.

| Exposure | Cases | Noncases | Total population (cases + noncases) |
|----------|-------|----------|--|
| Present | a | b | a + b |
| Absent | c | d | c + d |

Another advantage of this approach: you can calculate exposure prevalence, which allows you to calculate the **population attributable risk**

Szklo & Nieto, 3rd Ed. Chapter 3



The first of these studies for us to consider is called the case cohort study. And in this name, cases as usual mean individuals who develop disease within the cohort. Controls are the study population or, more commonly, a random sample of the study population from baseline or at baseline. So think about what this means, because this could include future cases. People who are controls at baseline could themselves become cases later. The division of the odds of exposure in cases by that in controls yields the relative risk. Specifically the cumulative incidence ratio.

So let's look at how this division works. So if we're trying to calculate the odds ratio for exposure, we'd want to get the odds of exposure among the cases and divide it by the odds of exposure among the total population. So our numerator here, the odds of exposure among the cases, would be a divided by c. Look at our table below. And the odds of exposure in the total population would be a plus b divided by c plus d. So this simplifies to the final expression here, a over a plus b divided by c over c plus d, and that estimates or gives us the relative risk.

So another advantage of this approach is that you can calculate exposure prevalence, which allows you to calculate the population attributable risk. And that is because you're working here with the total population. So notice in the denominator, how we use the total population as our standard for calculating, because we're taking the odds of exposure in the total population. That's why the denominator is set up as is.

Example: case-cohort study

TABLE 3-12 Case-cohort study of the relationship of previous vaccination to local reaction.²

| Previous vaccination | Cases of local reaction | Cohort sample |
|----------------------|-------------------------|---------------|
| Yes | 260 | 514 |
| No | 68 | 482 |
| Total | 328 | 996 |

$$OR_{cas} = \frac{\frac{260}{68}}{\frac{514}{482}} = 3.59 = RR$$

Szklo & Nieto, 3rd Ed. Chapter 3



So here's an example of some data from a case cohort study. This is looking at the relationship between prior vaccination and a local reaction to the vaccine. So the table is showing us the status of previous or prior vaccination, yes or no, and then how many cases of local reaction occurred, and how big the entire cohort was from which this was sampled. So notice this final column is the cohort sample of the entire population. So the numerator, here, would be 260 over 68. That is the odds of exposure among the cases. So the exposure here is the prior vaccination. So that would be 260 over 68, and the denominator would be 514 over 482. That is the odds of exposure among the entire cohort sample. That's the representation of the cohort. So doing the math, here, we get a relative risk of 3.59.

Case-control designs in which the OR approximates the RR

- **Nested case-control study**
 - Cases: incidence-density / risk set sampling
 - Controls: sampled at approximately the time when a case occurs
- The odds ratio of exposure represents an estimate of the relative rate or incidence density ratio.
- More on this later in the course.

Szklo & Nieto, 3rd Ed, Chapter 3



Another type of study in which the odds ratio approximates the relative risk is a nested case control study. Here, the cases are determined with incidence density or also called risk set sampling, and the controls are those individuals who are sampled at approximately the same time that a case occurs. So the sampling frame here is the time at which a case occurs, sends us out to get a control at about that same time. The odds ratio of exposure here represents an estimate of the relative rate or the incidence density ratio. We'll talk much more about this later in the course, when we discuss case control studies.

Summary of what the OR estimates under different control sampling schemes

TABLE 3-11 Summary of the influence of control selection on the parameter estimated by the odds ratio of exposure in case-control studies within a defined cohort.¹⁰⁰

| Design | Population frame for control selection | Exposure odds ratio estimates |
|---------------------|--|--|
| Nested case-control | Population at approximate times when cases occur during follow-up (Population during follow-up minus cases) | Rate (density) ratio (Density odds ratio) |
| Case-cohort | Total cohort at baseline (Total cohort at baseline minus cases that develop during follow-up) | Cumulative incidence ratio (relative risk) (Probability odds ratio) |

We'll return to this topic in more detail when we learn about case-control study designs.

Szklo & Nieto, 3rd Ed. Chapter 3



Let's put all these together and summarize what the odds ratio estimates are under different control sampling schemes, because it's really about the sampling of the controls that's critical. Let's look at the influence of control selection on the parameter that we're trying to estimate. So in a nested case control study, we're using as a population frame for control selection the population at approximate times when cases occur during follow up. So when a case occurs at about the same time, we take control. And then the exposure odds ratio that's estimated is a rate density ratio. In a case cohort study, we use the entire cohort at the baseline of the study. So note how that's different than nested case control. And here what we're estimating is the cumulative incidence ratio, or relative risks. So think about that, how we're sampling from, in a sense, in a theoretical cohort here. So we'll come back to this topic in more detail when we learn about case control study designs.

Additive scale measures in case-control studies

- We can calculate the risk difference in case-control studies when the OR is a reasonable estimate of the RR.
- This is true when the disease is rare and/or when the control group is representative of non-cases in the study population (e.g., a case-cohort study).
- If either is true, we can substitute the OR for the RR in formulas for the attributable proportion among the exposed (APe%) and attributable proportion in the total population (APt%).

Szklo & Nieto, 3rd Ed, Chapter 3



9

Let's talk, for a moment, about additive scale measures in case control studies. We can calculate the risk difference in case control studies, when the odds ratio is a reasonable estimate of the relative risk. This is true when the disease is rare or when the control group is representative non-cases in the study population. For example, we might see this in a case cohort study. If either is true, we can substitute the odds ratio for the relative risk in formulas for the attributable proportion among the exposed. That's abbreviated as you see here, APe%, and the attributable proportion in the total population abbreviated as APt%.

Estimating the attributable proportion among the exposed (APe%) in case-control studies

- First, let's derive the formula for the APe% that includes the RR. Then we can substitute the OR for the RR under certain conditions.

$$\begin{aligned} \text{APe\%} &= [(R_e - R_u) / R_e] \times 100\% \\ &= [1 - (1 / \text{RR})] \times 100\% \\ &= [(\text{RR} - 1) / \text{RR}] \times 100\% \end{aligned}$$

- If (1) the disease is rare and/or (2) the control group is representative of non-cases in the study population (e.g., a case-cohort study), we can substitute the OR for the RR.

$$\text{APe\%} = [(\text{OR} - 1) / \text{OR}] \times 100\%$$

Szklo & Nieto, 4th Ed, Chapter 3



First let's derive the formula for the APe%, the attributable portion among the exposed, that includes the relative risk. Then we can substitute the odds ratio for the relative risk under certain conditions. So you see the formulas written out here. APe% equals the difference between the risk in the exposed minus the risk in the unexposed divided by the risk in the exposed converted to a percentage. So we can simplify that as 1 minus the quantity 1 divided by the relative risk, and that can be rewritten as the relative risk minus 1 divided by the relative risk. These are just algebraic rearrangements and simplifications. If, number one, the disease is rare and/or, number two, the control group is representative of non cases in the study population, such as in a case cohort study, we can substitute the odds ratio for the relative risk. So then we can rewrite the formula as attributable portion in the exposed equals the odds ratio minus 1 divided by the odds ratio times 100%.

Estimating the attributable proportion in the total population (APt%) in case-control studies

$$APt\% = [(R_t - R_u) / R_t] \times 100\%$$

P_e = prevalence of exposure

$$R_t = (R_e \times P_e) + R_u(1 - P_e)$$

Now we substitute this formula in for R_t and rearrange terms to obtain a version of the formula with the RR.

$$APt\% = \frac{(R_e \times P_e) + (R_u \times (1 - P_e)) - R_u}{(R_e \times P_e) + (R_u \times (1 - P_e))} \times 100$$

$$APt\% = \frac{(R_e \times P_e) + (R_u \times P_e)}{(R_e \times P_e) - (R_u \times P_e) + R_u} \times 100$$

Szklo & Nieto, 4th Ed. Chapter 3

We can also estimate the attributable proportion in the total population in case control studies. So, again, the abbreviation here is APt% equals the difference between the risk in the total population minus the risk in the unexposed population divided by the risk in the total population. And we're going to write the prevalence of exposure as P_{sub-e} . So we can write the risk in the total population as the sum of two quantities. It's the sum of the risk in the exposed times the prevalence of exposure, plus the risk in the unexposed times the quantity 1 minus the prevalence of exposure. If we substitute this formula in for the risk in the total population and rearrange the terms to obtain a version of the formula with the relative risk, you see the formula as written out here. First, the attributable portion in the total population as a percentage, given by this first formula, and then the simplification of it, a little bit, in the second formula.

Estimating the attributable proportion in the total population (APt%) in case-control studies (continued)

$$AP_t\% = \frac{(R_e \times P_e) + (R_u \times P_e)}{(R_e \times P_e) - (R_u \times P_e) + R_u} \times 100$$

Next, divide all terms by R_u

$$AP_t\% = \frac{\frac{R_e \times P_e}{R_u} + \frac{R_u P_e}{R_u}}{\frac{R_e \times P_e}{R_u} - \frac{R_u P_e}{R_u} + \frac{R_u}{R_u}} \times 100$$

Rearrange including RR in the formula

$$AP_t\% = \frac{(RR \times P_e) + P_e}{(RR \times P_e) - P_e + 1} \times 100 = \frac{(OR \times P_e) + P_e}{(OR \times P_e) - P_e + 1} \times 100$$

If (1) the disease is rare and/or
(2) the control group is
representative of non-cases in
the study population (e.g., a
case-cohort study), we can
substitute the OR for the RR.



Our next step is to divide all the terms by the risk in the unexposed. So the attribute of proportion in the total population equals the formula that you see here. And when we rearrange to include the relative risk in the formula, it now equals the formula that you see here. So note that if the disease is rare and/or the control group is representative of non cases in the study population, like in the case cohort study, we can substitute the odds ratio for the relative risk, as we discussed earlier.

Summary of key points

- Epidemiologists were previously taught that ORs were the only appropriate measure of association in case-control studies.
- The modern view of case-control studies is that they occur within a cohort.
- To assess what measure of association is estimated by the OR and whether the rare disease assumption is needed for the OR to approximate the RR, one must consider the method used to sample controls from the underlying cohort.

In summary, epidemiologists were previously taught that odds ratios were the only appropriate measure of association in case control studies, but the modern view of case control studies is that they actually occur within a cohort. Even if you haven't defined a cohort, the case control studies being sampled from that theoretical cohort. So to assess what measure of association is estimated by the odds ratio and whether the rare disease assumption is needed for the odds ratio to approximate the relative risk, one must consider the method used to sample the controls from the underlying cohort.

Editors' note: This series addresses topics that affect epidemiologists across a range of specialties. Commentaries start as invited talks at symposia organized by the Editors. This paper was presented at the 2009 Society for Epidemiologic Research Annual Meeting in Anaheim, CA.

Case-Control Studies = Odds Ratios

Blame the Retrospective Model

Bryan Langholz

Many epidemiologists and statisticians believe that the odds ratio is the only measure that can be reliably estimated from case-control studies. I have received many reviews of case-control study papers instructing us to change “rate ratio” to “odds ratio” when, in fact, it was the rate ratio that we had estimated. Further, in discussions about methods we have developed to estimate absolute risk measures from nested case-control studies, I have found that many are surprised to learn that absolute risk can be estimated systematically and reliably in the case-control setting. While case-control studies are best suited for estimation of relative measures, and I do not wish to minimize the challenges of estimation on other scales from case-control data generally, there are reliable methods for doing so. Reporting of absolute risk estimates seems desirable to supplement the usual case-control relative measure analyses,¹ but this is only rarely done even when it is feasible.

So why is there an odds-ratio fixation? I believe the core problem is that epidemiologists often think of case-control studies as a “retrospective model.” According to this view, we start with a set of cases and controls. Then the covariates (exposures and other factors) occur as independent realizations with distribution dependent on disease status. This is in contrast to the “prospective model” of cohort data, in which we start with a group of subjects with given covariates, and disease status is the result of independent realizations with probability dependent on the covariate values. Students learn that the odds ratio parameters in the retrospective logistic model are same as the odds ratio parameters in the corresponding prospective logistic model, but that other measures do not translate. The matter is further confused because valid estimation of odds ratio parameters from retrospective model case-control data may be obtained using the corresponding prospective cohort data logistic regression, with the estimated “baseline odds” a nuisance parameter to be ignored.^{2,3}

But case-control studies are not backwards cohort (“trohoc”) studies.⁴ A more realistic way to represent case-control designs is as sampling from a prospective cohort that depends on disease outcomes and other information available on cohort subjects—what we have called the nested case-control model.⁵ While this representation is certainly not new, and is often used in epidemiology textbooks as a “conceptual framework” to think about basic sampling and bias issues, almost invariably the retrospective model approach is used to develop the analysis methods. The alternative approach my colleagues and I have taken is to develop case-control study methods based completely on the nested case-control model, and to provide a unifying framework across cohort and case-control analysis methods, as well as across individually matched and unmatched case-control study designs. While there are still some important gaps to be filled, we have made progress. After

From the Department of Preventive Medicine, University of Southern California, Los Angeles, CA.

Editors' note: Related articles appear on pages 13 and 3.

Correspondence: Bryan Langholz, Keck School of Medicine, University of Southern California, 1540 Alcazar St, CHP-220, Los Angeles, CA 90033. E-mail: langholz@usc.edu.

Copyright © 2009 by Lippincott Williams & Wilkins

ISSN: 1044-3983/10/2101-0010

DOI: 10.1097/EDE.0b013e3181c308f5

working within the nested case-control model over a number of years, I have come to find the retrospective model both unnatural and limiting. There are 3 commonly held misconceptions as a result of the retrospective model mindset.

The first is that cases and controls need to be “random samples,” meaning having equal probability of being sampled, from their respective populations (eg, selection cannot depend on covariates). This misconception is not directly relevant to the topic of estimation of measures other than odds ratios, so I will not discuss it at length here except to reiterate our opinion that “The ‘random sampling of controls’ principle needs to be replaced by the principle that ‘the method of control selection must be incorporated into the analysis.’”^{1,6}

The second misconception is that only odds ratio parameters can be estimated from case-control studies. As a quick counter-example, rate-ratio parameters can be estimated when controls are sampled from the risk sets in disease incidence cohort data.^{7–9} With additional information related to the overall cohort rates or size, absolute risk or rates can be estimated. In the retrospective model, there is no element of an “underlying cohort,” so that the connection between a case-control sample and the cohort is not integrated into the conceptual framework. I refer to our work in this area, but there are other approaches.^{10–12} Our methods for case-control data are a natural extension of well established methods for disease incidence cohort studies, such as Kaplan-Meier estimates of survival and risk^{13,14} and are easily implemented using Cox regression software that computes baseline survival or cumulative hazard functions such as SAS, Stata, S-Plus, or R.⁵ As originally presented, these methods require the number in each risk set, and are thus appropriate when the underlying cohort can be fully enumerated. However, as I discuss below, the methods also apply in more general situations. Estimation of other measures is also possible, including excess risk estimation via nonparametric^{15,16} and Poisson regression^{17,18} methods.

The third misconception is that exact specification of the sampling design in the analysis does not matter as long as the controls are randomly sampled. The crux of the argument can be discussed in the context of estimation of the baseline odds. As mentioned earlier, analysis of case-control data using cohort (unconditional) logistic regression arose based on the retrospective view. From this standpoint, the natural and common approach to estimation of baseline odds is to “fix up” the logistic regression intercept parameter using the case and control sampling fractions.¹¹ While this is certainly valid for point estimation of the baseline odds, variance estimation requires more precise specification of the sampling¹⁹—a complication not addressed in most basic textbooks advocating the approach. The prospective view of case-control studies naturally suggests an approach to analysis of case-control data based on a conditional logistic likelihood. This likelihood depends on the specific sampling method, and the baseline odds may (or may not) be estimable from the case-control data, without supplemental in-

formation, depending on the sampling design. So, for instance, the baseline odds is estimable if controls are drawn as Bernoulli trials or as case-base samples, but not as frequency matched samples.⁶

The “retrospective” model has led to the mindset of “case-control study = odds ratio.” This equation is a fallacy. In many situations, absolute risk and other measures can be reliably estimated from case-control data and, in fact, from a much wider range of case-control studies than the nested case-control study setting I have discussed thus far. In none of the nested case-control studies in which our methods were applied were the cohorts fully enumerated.^{20,21} In each, the target cohort was a subset of the assembled cohort, with eligibility determined as part of the process of identifying cases and controls. Our methods can accommodate this situation, incorporating the numbers of potential cases and controls that were sampled and disqualified from the analysis, as well as the number of controls in the assembled cohort risk sets.^{5,22} Moreover, I conjecture that the methods can be extended to situations where controls are matched to the case within some group characteristic (unit) that can be enumerated, such as school or defined geographic areas, and where the numbers of potential controls in the units in which cases occurred can be ascertained.

In conclusion, reliable estimation of risk and other nonodds-ratio measures from case-control studies is certainly possible as long as the ancillary information about the underlying cohort is obtained and the details of the case and control selection process recorded. Such analysis requires a little extra planning and the use of appropriate design and analysis techniques—but little added expense.

ACKNOWLEDGMENTS

I would like to thank Drs. David Richardson and Duncan Thomas for their excellent feedback on the manuscript.

REFERENCES

- Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. *Am J Epidemiol.* 2004;160:301–305.
- Prentice R, Pyke R. Logistic disease incidence models and case-control studies. *Biometrika.* 1979;66:403–411.
- Carroll RJ, Wang S, Wang CY. Prospective analysis of logistic case-control studies. *J Am Stat Assoc.* 1995;90:157–169.
- Feinstein AR. Clinical Biostatistics XX. The epidemiologic trohoc, the ablative risk ratio, and “retrospective” research. *Clin Pharmacol Ther.* 1973;14:291–307.
- Langholz B. Use of cohort information in the design and analysis of case-control studies. *Scand J Stat.* 2007;34:120–136.
- Langholz B, Goldstein L. Conditional logistic analysis of case-control studies with complex sampling. *Biostatistics.* 2001;2:63–84.
- Greenland S, Thomas D. On the need for the rare disease assumption in case-control studies. *Am J Epidemiol.* 1982;116:547–553.
- Oakes D. Survival times: aspects of partial likelihood (with discussion). *Int Stat Rev.* 1981;49:235–264.
- Borgan Ø, Goldstein L, Langholz B. Methods for the analysis of sampled cohort data in the Cox proportional hazards model. *Ann Stat.* 1995;23:1749–1778.
- Wacholder S. The case-control study as data missing by design: estimating risk differences. *Epidemiology.* 1996;7:144–150.

11. Rothman KJ, Greenland S. *Modern Epidemiology*. Philadelphia: Lippincott-Raven Publishers; 1998:chap 21.
12. Benichou J, Gail M. Methods of inference for estimates of absolute risk derived from population-based case-control studies. *Biometrics*. 1990; 51:182–194.
13. Langholz B, Borgan O. Estimation of absolute risk from nested case-control data. *Biometrics*. 1997;53:767–774.
14. Borgan Ø, Langholz B. Risk set sampling designs for proportional hazards models. In: Everitt BS, Dunn G, eds. *Statistical Analysis of Medical Data: New Developments*. London: Arnold; 1998:75–100.
15. Aalen OO. Nonparametric inference for a family of counting processes. *Ann Stat*. 1978;6:701–726.
16. Borgan Ø, Langholz B. Estimation of excess risk from case-control data using Aalen's linear regression model. *Biometrics*. 1997;53:690–697.
17. Samuelsen SO. A pseudolikelihood approach to analysis of nested case-control data. *Biometrika*. 1997;84:379–394.
18. Samuelsen SO, Ånestad H, Skrondal A. Stratified case-cohort analysis of general cohort sampling designs. *Scand J Stat*. 2007;34:103–119.
19. Arratia R, Goldstein L, Langholz B. Local central limit theorems, the high order correlations of rejective sampling, and logistic likelihood asymptotics. *Ann Stat*. 2005;33:871–914.
20. Habel LA, Shak S, Jacobs MK, et al. A population-based study of tumor gene expression and risk of breast cancer death among lymph node-negative patients. *Breast Cancer Res*. 2006;8:R25.
21. Lacey JV Jr, Sherman ME, Ronnett ME, et al. Absolute risk of endometrial carcinoma over 20-year follow-up among women diagnosed with endometrial hyperplasia. *J Clin Oncol*. In press.
22. Langholz B. *Estimation of Absolute Risk When Controls are Sequentially Sampled Until a Specified Number of Valid Controls are Found*. Los Angeles: USC Department of Preventive Medicine Biostatistics Division; 2005. Technical Report 172.

Lecture: Use and Misuse of Population Attributable Fractions



Use and misuse of population attributable fractions

PHW250 F - Jack Colford

Based on the article by Rockhill et al. *Am J Public Health* 1998; 88(1).

JACK COLFORD: In this video, we're going to discuss the use and misuse of population attributable fractions, often abbreviated as PAF.

Quick review: population attributable fractions (PAFs)

- **Most common definitions:**

- Proportional reduction in average disease risk over a specific time interval that would be achieved by eliminating the exposure(s) of interest from the population while distributions of other risk factors in the population remain unchanged
- Proportion of disease cases over a specified time that would be prevented following elimination of the exposures, assuming the exposures are causal
- **Example:** a cohort study found that air pollution was associated with the cumulative incidence of asthma (CIR=2.0). The cumulative incidence was 0.4 among the exposed and 0.2 among the unexposed. The total risk in the population was 0.9.
 - $\text{PAF} = (\text{Risk}_{\text{total}} - \text{Risk}_{\text{unexposed}}) / \text{Risk}_{\text{total}} \times 100\%$
 - $\text{PAF} = (0.9 - 0.2) / 0.9 = 0.78$
 - **Interpretation:** 78% of incident asthma cases in the study population would be prevented if no one in the study population was exposed to air pollution.



Quickly, to review population attributable fractions, let's talk about their most common definitions.

One way to think of these is a proportional reduction in average disease risk over a specific time interval that would be achieved by eliminating the exposure or exposures of interest from the population while keeping all the other risk factors in the population unchanged, or the proportion of disease cases over a specified time that would be prevented following elimination of the exposures, assuming the exposures are actually causal.

Here's an example. Those definitions, I know, can sometimes be confusing in words, but let's look at some examples. There was a cohort study that found that air pollution was associated with the cumulative incidence of asthma-- cumulative incidence ratio of 2.0.

The cumulative incidence was 0.4 among the exposed and 0.2 among the unexposed. So dividing 0.4 by 0.2, you come up with the CIR of 2.0. The total risk in the population was 0.9. And of course, the total population is a mixture of exposed and unexposed.

The population attributable fraction is the total risk minus the unexposed risk and dividing that by that total risk and expressing it as a percentage, basically saying, what proportion of the total risk is represented by the difference between the total and the unexposed risk? Just plugging in the numbers, you can come up with the calculation.

Different terms are often used interchangeably for PAFs

| Measure of association | Formula | Interpretation |
|--------------------------------------|---|---|
| Population attributable fraction | $(R_{total} - R_{unexposed}) / R_{total} \times 100\%$ | Proportion of disease cases over a specified time that would be prevented following elimination of the exposures, assuming the exposures are causal |
| Population attributable risk | $R_{total} - R_{unexposed}$ | Difference in risk over a specified time that would be prevented following elimination of the exposures, assuming the exposures are causal |
| Excess fraction / Etiologic fraction | $(R_{exposed} - R_{unexposed}) / R_{exposed}$ <small>(definition in Rothman)</small> | Proportion of exposed risk that is attributable to exposure Note: to the etiologic fraction is not estimable without strong biologic assumptions |

Berkeley School of Public Health

Let's look at these different measures of association and the formulas and interpretations that go with them. First, the population attributable fraction, which we just did an example of in the last slide, is the difference between the total and unexposed risk divided by the total risk expressed as a population. And this gives us the proportion of disease cases over a specified time that would be prevented following elimination of the exposures, assuming that the exposures are causal.

Next measure is the population attributable risk, or PAR. And this is simply the difference between the total risk and the unexposed risk in the population. And the interpretation of this is a difference in risk over a specified time that would be prevented following elimination of the exposures, assuming the exposures, again, are causal. This assumption runs through all of these.

And finally, the last measure is excess fraction or the etiologic fraction, which is now comparing the risk in the exposed to the risk in the unexposed, taking that difference and dividing by the risk in the exposed. So notice that the denominator here is different. And also we're looking at the exposed group, not the total population.

The proportion of exposed risk that is attributable to the exposure is the interpretation here. And again, this is in the exposed population, not in the total population. That's the key difference I'd like you to see from the two measures above, the exposed population rather than the total population.

Distributive property of PAFs

- The PAF can be partitioned into exposure category-specific attributable fractions, which then sum to the population attributable fraction.
- Category-specific PAF = fraction of total risk in the population that would be eliminated if persons in only that specific exposure category were shifted to the unexposed group

$$= p_i \frac{(RR_i - 1)}{RR_i}$$

→ Where p_i is the proportion of total cases in the population arising from the i th exposure category.

One of the properties of population attributable fractions is that it's called the distributive property and that we can partition it into exposure category-specific attributable fractions which then sum to the total population attributable fraction. Category-specific population attributable fractions are the fraction of total risk in the population that would be eliminated if persons in only that specific exposure category were shifted to the unexposed group.

That would be expressed as the proportion in the particular group we're studying, where we take the relative risk for that group minus 1 and divide by the relative risk. Here, p_i is the proportion of total cases in the population arising from the i th exposure category.

Distributive property of PAFs

Example: calculate category-specific PAFs for both smoking categories

| | Lung cancer | No lung cancer | Total |
|---------------------------------|-------------|----------------|-------|
| E2: Smoked 2+ packs per day | 50 | 50 | 100 |
| E1: Smoked up to 1 pack per day | 30 | 120 | 150 |
| E0: Non smokers | 20 | 180 | 200 |
| Total | 100 | 325 | 450 |

$$RR_{E2} = 0.5 / 0.1 = 5$$

$$RR_{E1} = 0.2 / 0.1 = 2$$

$$p_{E2} = 50/100 = 0.5$$

$$p_{E1} = 30/100 = 0.3$$

$$PAF_{E2} = 0.5 * (5-1) / 5 = 0.4$$

$$PAF_{E1} = 0.3 * (2-1) / 2 = 0.15$$

$$PAF = 0.4 + 0.15 = 0.55$$

$$= (R_{total} - R_{unexp}) / R_{total} =$$

$$= (0.22 - 0.1) / 0.22 = 0.55$$

Here's an example of this distributive property. We have these three different levels of smoking, exposure 2, exposure 1, and exposure 0. And you can see the different level of exposures there. And in each of these groups, we have some people who developed lung cancer and some people didn't develop lung cancer.

We could calculate the relative risk of exposure level 2 by taking 0.5, which is the probability of disease in that group, divided by 0.1, that is the probability of disease in the nonsmokers, E2 over E0, which would give us a risk ratio of 5.

Similarly, for E1 we would calculate 0.2 over 0.1, with the 30 divided by 150 over 20 divided by 200 and so forth. In the exposure level 2, the P sub E2 would be 50 over 100 or 0.5. Half the cases are in exposure category E2. And in a level E1, it's 30 over 100.

The population attributable fraction calculation in E2 would be 0.5 times 5 minus 1. That's the relative risk in that level minus the relative risk in the control group, which is 1, divided by 5, which is the relative risk in the exposure 2 level, just plugging into the formulas we used before. That gives us a population attributable fraction of 0.4.

Similar calculations for exposure level 1. And what you notice here is that if you

calculate population attributable fraction for E2, you get 0.4. The population attributable fraction for E1, you get 0.15. I'm assuming you can just do those calculations from the table.

There's this property, distributive property, where we can add those together. And that's the same that we got from calculating the total risk minus the unexposed divided by the total risk. 0.22 minus 0.1 divided by 0.22 again gives us 0.5, just to double check to show that this property holds.

Caution: PAFs for different risk factors **cannot be summed to derive the total fraction of disease risk attributable to all risk factors**

- Example:
 - PAF for smoking = 0.55
 - PAF for air pollution = 0.56
- Almost always total PAF for smoking and air pollution $\neq 0.55 + 0.56$
- This will almost always yield a value larger than correctly calculated summary population attributable fraction for all of the factors considered simultaneously.
- This is because a set of individual PAFs from a study will often sum to more than 1.0.
- Most epidemiologists assume that a given case of disease can be prevented by eliminating any of the necessary causal factors present.

Population attributable fractions for different risk factors cannot be summed to derive the total fraction of disease attributable to all risk factors. For example, if I had a PAF for smoking of 0.55 and a PAF for air pollution of 0.56, almost always the total population attributable fraction for combining smoking and air pollution will not be the sum of those because often this will yield a value larger than the correctly calculated summary population attributable fraction for all of the factors considered simultaneously.

They'll often sum in a study to more than 1. So most epidemiologists assume that a given case of disease can be prevented by eliminating any of the necessary causal factors that are present.

Defining exposure / intervention levels in the PAF

- From a public health perspective, the PAF is most useful when the exposure of interest is:
 - Clearly causally related to the outcome
 - Amenable to intervention
- It is still common practice to estimate PAFs for **attributes that are not modifiable**, such as ethnicity or family history of cancer.
 - Example: A PAF for ethnicity and breast cancer does not provide the public health community with actionable information. Relative risks stratified by ethnicity may be more appropriate.
- Using a **surrogate for more proximate exposure** may also be less helpful than estimating the PAF for the exposure itself.
 -  Example: Marital status might be a surrogate for certain health behaviors. Estimating the PAF for breast cancer and marital status does not provide information about how to reduce breast cancer because ensuring all women marry is not an appropriate or feasible public health intervention.

For defining exposure or intervention levels in the population attributable fraction, when you're working from a public health perspective, the PAF is most useful when the exposure of interest is clearly causally related to the outcome and it's amenable to intervention, it's something we can intervene on.

It's still common practice to estimate PAFs for attributes that are not modifiable, such as ethnicity or family history of cancer. We can't change those, but epidemiology will still make some estimation of the PAF. For example, one might estimate a PAF for ethnicity in breast cancer, but it doesn't provide the public health community with actionable information because we can't change the ethnicity. Relative risk stratified by ethnicity may be more appropriate.

Also, we might use a surrogate from a more proximate exposure, and that might be less helpful than estimating the PAF for the exposure itself. So for example, marital status might be a surrogate for certain behaviors. Estimating the PAF for breast cancer and marital status doesn't provide information about how to reduce breast cancer because ensuring all women is not an appropriate or feasible public health intervention.

Defining exposure / intervention levels in the PAF

- Ideally the unexposed category in a PAF is defined in a way that makes it realistic for the exposed people to become unexposed.
- This is another reason why non-modifiable exposures are not ideal choices for the PAF.
- Even for modifiable exposures, in some instances it is difficult for the exposed to become unexposed:
 - PAF for lung cancer and never smokers (*not realistic*) **can't undo!**
 - • PAF for lung cancer and current non smoker (*realistic*)

Ideally, the unexposed category in a PAF is defined in a way that makes it realistic for the exposed people to become unexposed. That is, we can do an intervention that moves them from exposed status to unexposed status. This is another reason why non-modifiable exposures are not ideal choices for the PAF.

Even for modifiable exposures, in some instances it's difficult for the exposed to become unexposed. For example, the PAF for lung cancer in never smokers is not realistic, but the PAF for lung cancer in current non-smokers is realistic. Why is the first one not realistic? Because you can't go back and undo the smoking that already happened. But you can in a realistic way turn current smokers into current nonsmokers.

Common misinterpretations of the PAF

- A study estimated a PAF of 15% for 10 breast cancer risk factors.
- **Quiz:** Here are several interpretations of this finding. Read these and indicate whether each was interpreted correctly or incorrectly.
 - "Although various risk factors have been identified as causes of breast cancer, the fact remains that in 75% of all breast cancer no identifiable risk factor can be found."
 - "Only 25% of cancers could be attributed to one or more risk factors, meaning that the majority of cancers occur in women with no risk factors."
 - "25% of the population risk of breast cancer would be eliminated if all 10 risk factors were to be eliminated from the population."

Berkeley School of Public Health

There are common misinterpretations of the population attributable fraction. One study estimated a PAF of 15% for 10 breast cancer risk factors. Here are several interpretations of this finding. Let's read these together and then indicate whether each was interpreted correctly or incorrectly.

First interpretation. Although various factors have been identified as causes of breast cancer, the fact remains that in 75% of all breast cancer, no identifiable risk factor can be found. Is that correct? Next. Only 25% of cancers could be attributed to one or more risk factors, meaning that the majority of cancers occur in women with no risk factors. Is that correct? Or finally, 25% of the population risk of breast cancer would be eliminated if all 10 risk factors were to be eliminated from the population.

Common misinterpretations of the PAF

- "Although various risk factors have been identified as causes of breast cancer, the fact remains that in 75% of all breast cancer no identifiable risk factor can be found." **Incorrect**
 - This is a very frequent misinterpretation of the PAF.
 - This interpretation equates the PAF with the proportion of cases having any risk factors. The PAF measures the proportion of disease that would be prevented following elimination of the exposures, not the proportion of the population with the risk factors.
- The study evaluated 10 risk factors, but there are many other potential risk factors that were not studied or measured.

Although those various risk factors have been identified as causes of breast cancer, the fact remains that in 75% of all breast cancer, no identifiable risk factor can be found. This is incorrect. It's a very frequent misinterpretation of the PAF. This interpretation equates the PAF with the proportion of cases having any risk factor.

The PAF measures the proportion of disease that would be prevented following elimination of exposures, not the proportion of the population with the risk factors. The study evaluated 10 risk factors, but there are many other potential factors that were not studied or measured.

Common misinterpretations of the PAF

- "Only 25% of cancers could be attributed to one or more risk factors, meaning that the majority of cancers occur in women with no risk factors." **Incorrect**
- The proportion of the population exposed to the considered risk factors is different from the PAF.
- The PAF itself does not provide information about the proportion of the population that was exposed.

The next statement. Only 25% of cancers could be attributed to one or more risk factors, meaning that the majority of cancers occur in women with no risk factors. This is also incorrect. The proportion of the population exposed to the considered risk factors is different from the PAF. And the PAF itself does not provide information about the proportion of the population that was exposed.

Common misinterpretations of the PAF

- "25% of the population risk of breast cancer would be eliminated if all 10 risk factors were to be eliminated from the population" **Correct**
- This interpretation is consistent with the definition of the PAF.

Finally, the third statement is the correct one. 25% of the population risk of breast cancer would be eliminated if all 10 risk factors were to be eliminated from the population. This is correct. And this interpretation is consistent with the definition of the PAF.

Other sources of confusion about PAFs

- In interpreting PAFs, authors often use the words “cause”, “explain”, and “attribute”
 - Example: Study by Madigan et al., PAF = 0.40
 - “A substantial proportion of breast cancer cases in the US are explained by well-established risk factors”.
 - Their data suggest that nearly all women in the US have at least one of the risk factors included in the PAF
 - Yet, the vast majority of “exposed” women will never experience breast cancer.
- Extreme example: PAF for age > 15 years and breast cancer. To imply that being over age 15 years causes breast cancer is not valuable.

Berkeley School of Public Health

12

There's other sources of confusion about PAFs. In interpreting PAFs, authors often use the words cause, explain, attribute. For example, in a study by Madigan et al., the PAF was measured as 0.40. And a sentence in this study said, "A substantial proportion of breast cancer cases in the US are explained by well-established risk factors."

Their data suggests that nearly all women in the US have at least one of the risk factors included in the PAF. Yet the majority of exposed women will never experience breast cancer. So in an extreme example, the PAF for age is greater than 15 years in breast cancer. To imply that being over age 15 years causes breast cancer is not a valuable observation.

A summary of key points. The population attributable fraction provides information about how the population level risk would change under different hypothetical interventions. We might remove the harmful exposure or introduce beneficial interventions. The PAF can be partitioned into exposure category-specific attributable fractions which then sum up to the population attributable fraction.

PAFs for different risk factors cannot be summed to derive the total fraction of disease risk attributable to all risk factors. And from a public health perspective, the PAF is most useful when the exposure of interest is clearly causally related to the outcome and, number two, amenable to intervention.

Commentary

Use and Misuse of Population Attributable Fractions

Beverly Rockhill, PhD, Beth Newman, PhD, and Clarice Weinberg, PhD

Introduction

How much of the disease burden in a population could be eliminated if the effects of certain causal factors were eliminated from the population? To address this question, epidemiologists calculate the population attributable fraction. As noted in a recent editorial in the Journal, population attributable fraction estimates can help guide policymakers in planning public health interventions.¹ Despite numerous articles on population attributable fraction estimation,²⁻⁷ errors in computation and interpretation persist. In addition, in certain settings, the value of a population attributable fraction estimate may be questionable. This commentary considers computational and conceptual issues relevant to population attributable fraction estimation that are infrequently discussed elsewhere, with illustrations from the breast cancer literature.

Background

In 1953, Levin⁸ first proposed the concept of population attributable fraction. Since then, the phrases "population attributable risk," "population attributable risk proportion," "excess fraction," and "etiological fraction" have been used interchangeably to refer to the proportion of disease risk in a population that can be attributed to the causal effects of a risk factor or set of factors. Greenland and Robins⁴ distinguish between excess fraction (what epidemiologists usually estimate when they compute "population attributable risk" or "population attributable fraction") and etiologic fraction, which is not estimable without strong biologic assumptions. Our use of the term "population attributable fraction" corresponds to Greenland and Robins' (population) excess fraction.

The population attributable fraction is most commonly defined as the proportional reduction in average disease risk over a specified time interval that would be achieved by eliminating the exposure(s) of interest from the population while distributions of other risk factors in the population remain unchanged. This also can be interpreted as the proportion of disease cases over a specified time that would be prevented following elimination of the exposures, assuming the exposures are causal.

While population attributable fractions usually are estimated for single risk factors, they also can be estimated for groups of factors considered simultaneously. In this situation, a population attributable fraction estimates the proportional amount by which disease risk would be reduced if all of the factors were to be simultaneously eliminated from the population. The exposed group consists of those exposed to at least one of the factors. A population attributable fraction for a set of risk factors considered simultaneously is sometimes termed a summary population attributable fraction.

The preceding definitions show that the word "risk" in attributable risk is technically incorrect; it is more correct to speak of proportion or fraction of risk. For this reason, although the term "population attributable risk" is most commonly used, terms such as "population attributable risk propor-

Beverly Rockhill is with Channing Laboratory and Harvard School of Public Health, Boston, Mass. Beth Newman is with the Department of Epidemiology, University of North Carolina, Chapel Hill. Clarice Weinberg is with the National Institute of Environmental Health Sciences, Research Triangle Park, NC.

Requests for reprints should be sent to Beverly Rockhill, PhD, Department of Epidemiology, CB #7400, University of North Carolina, Chapel Hill, NC 27599-7400.

tion" and "population attributable fraction" are more accurate.

Basic Computational Issues

The expression corresponding to the preceding definition of population attributable fraction can be written as

$$\frac{P(D) - \sum_C P(D|\bar{C}, \bar{E}) P(C)}{P(D)}$$

where $P(D)$ is the average probability of disease in the population (containing both exposed and unexposed individuals) over a specified time interval and $\sum_C P(D|\bar{C}, \bar{E}) P(C)$ represents the marginal conditional probability of disease given no exposure, averaged over strata of other risk factors or confounders (C). Several formulas more commonly seen than the preceding one are used to estimate population attributable fractions. Some of these formulas are valid only under the assumption of no confounding of the exposure-disease association. Table 1 presents the most commonly seen computational formulas and discusses the limitations, if any, on the use of each formula. Several authors have provided derivations and detailed discussions of the various formulas.^{5-7,9}

The Distributive Property of the Population Attributable Fraction

A property of the population attributable fraction that is not always appreciated by epidemiologists is what Wacholder et al. term the distributive property.^{9,10} A population attributable fraction can be quantitatively partitioned, or distributed, into exposure-category-specific attributable fractions, which then sum to the population attributable fraction. A category-specific attributable fraction is the fraction of total disease risk in the population that would be eliminated if persons in only that specific exposure category were to be shifted to the unexposed group. When several risk factors are being considered simultaneously, the exposure categories arise from a complete cross classification of the risk factors under consideration.

A category-specific attributable fraction is estimated as

$$pd_i \left(\frac{RR_i - 1}{RR_i} \right)$$

where RR_i is the (adjusted) relative risk for the i th exposure category (relative to the

unexposed stratum) and pd_i represents the proportion of total cases in the population arising from the i th exposure category. The category-specific attributable fraction for the unexposed group ($i = 0$) is 0, since the RR_i is 1.0 by definition. The sum of the category-specific attributable fractions is thus

$$\sum_i pd_i \left(\frac{RR_i - 1}{RR_i} \right),$$

which can be simplified to

$$1 - \sum_i \frac{pd_i}{RR_i}$$

(formula 5 in Table 1).

Important implications of the distributive property have been previously noted.^{7,10,11} The population attributable fraction will increase with an increasingly inclusive definition of exposure, provided that each group added to the "exposed" segment has a relative risk greater than 1.0 (in comparison with the remaining unexposed group). However, there may be a loss of precision with a broad exposure definition, since the standard error of the population attributable fraction increases as the proportion of exposed cases and controls increases above 0.50.^{12,13}

Errors in Computation

Perhaps as a result of the proliferation of computational formulas for population attributable fractions, errors in estimation are common. Probably the most common error is the use of adjusted relative risks in formula 3 (see Table 1).¹⁴⁻¹⁹ The magnitude of bias resulting from this error will depend on the degree of confounding.

Another type of computational error that is likely to involve more substantial bias is illustrated in an article on poverty and mortality in the United States.²⁰ The authors inappropriately used formula 3 (see Table 1) to estimate a "weighted" population attributable fraction across strata of a third (nonexposure) variable; that is, they misapplied the stratum reference in formula 3. As a result, the published estimates of the fraction of all-cause US mortality attributable to poverty were overestimated by approximately a factor of three.²¹

Conceptual Issues

Summing Population Attributable Fraction Estimates

Some epidemiologists inappropriately sum single risk factor population attribut-

able fraction estimates in an attempt to derive the total fraction of disease risk attributable to all of the factors. This strategy is rarely appropriate and will almost always yield a value larger than the correctly calculated summary population attributable fraction for all of the factors considered simultaneously. Walter¹² discusses the limited conditions under which individual population attributable fractions may be validly summed.

One corollary of the preceding discussion is that it is possible, albeit counterintuitive, that a set of individual population attributable fractions will sum to more than 1.0. An implication of the "multicausal" model under which most epidemiologists work is that a given case of disease can be prevented by eliminating any one of the necessary causal factors present. For a specific disease, therefore, population attributable fractions computed separately for different risk factors are not constrained to sum to 1.0 or less. This issue has sometimes led to inappropriate analyses. At least two papers have attempted to attribute cancer risk to a variety of life-style and environmental factors considered singly, and the authors forced the population attributable fractions for the single factors (including a catchall factor of "unknown cause") to sum to 1.0.^{22,23}

Interpretation and Communication

Perhaps the most important aspect of population attributable fraction estimation is correct interpretation and communication. Consider an extensively cited paper devoted to population attributable fraction estimation for "established" breast cancer risk factors.¹⁹ Seidman et al. estimated population attributable fractions of 0.21 in the 30 to 54-year age group and 0.29 in the 55 to 84-year age group for 10 breast cancer risk factors. Misinterpretations of the population attributable fractions presented in this paper have been common, both in the scientific and lay literatures.

The most frequent error involves equating the population attributable fraction with the proportion of cases having any risk factors: "Although various risk factors have been identified as causes of breast cancer, the fact remains that in 75% of all breast cancer no identifiable risk factor can be found."²⁴ This error was made in an article advising clinicians on patient education: "Only 21 per cent of the cancers occurring in women from 30 to 54 years of age and 29 per cent in the women over 50 could be attributed to one or more risk factors, meaning that the majority of cancers occur in women with no risk fac-

TABLE 1—Commonly Seen Formulas for Attributable Fraction Estimation

| | |
|---|---|
| <p>1. $\frac{IP_t - IP_0}{IP_t}$</p> | <p>Empirical approximation of $P(D) - \sum_C P(D \bar{C}, \bar{E}) P(C)$ $P(D)$</p> <p>IP_t = cumulative proportion of total population developing disease over specified interval; IP_0 = cumulative proportion of unexposed persons who develop disease over interval. Valid only when no confounding of exposure(s) of interest exists. If disease is rare over time interval, ratio of average incidence rates I_t/I_0 approximates ratio of cumulative incidence proportions, and thus formula can be written as $(I_t - I_0)/I_t$. Both formulations found in many widely used epidemiology textbooks.</p> |
| <p>2. $\frac{p_e(RR - 1)}{p_e(RR - 1) + 1}$</p> | <p>Transformation of formula 1. Not valid when there is confounding of exposure–disease association. p_e = proportion of source population exposed to the factor of interest. RR may be ratio of two cumulative incidence proportions (risk ratio), two (average) incidence rates (rate ratio), or an approximation of one of these ratios. Found in many widely used epidemiology texts, but often with no warning about invalidness when confounding exists.</p> |
| <p>3. $\frac{\sum_{i=0}^k (p_i)(RR_i - 1)}{1 + \sum_{i=0}^k (p_i)(RR_i - 1)} = \\ 1 - \frac{1}{\sum_{i=0}^k p_i (RR_i)}$</p> | <p>Extension of formula 2 for use with multicategory exposures. Not valid when confounding exists. Subscript i refers to the ith exposure level. p_i = proportion of source population in ith exposure level, RR_i = relative risk comparing ith exposure level with unexposed group ($i = 0$). Derived by Walter¹²; given in Kleinbaum et al.²⁹ but not in other widely used epidemiology texts.</p> |
| <p>4. $pd \left(\frac{RR - 1}{RR} \right)$</p> | <p>Alternative expression. Produces internally valid estimate when confounding exists and when, as a result, adjusted relative risks must be used.⁹ pd = proportion of cases exposed to risk factor. In Kleinbaum et al.²⁹ and Schlesselman.³⁰</p> |
| <p>5. $\sum_{i=0}^k pd_i \left(\frac{RR_i - 1}{RR_i} \right) = 1 - \sum_{i=0}^k \frac{pd_i}{RR_i}$</p> | <p>Extension of formula 4 for use with multicategory exposures. Produces internally valid estimate when confounding exists and when, as a result, adjusted relative risks must be used. pd_i = proportion of cases falling into ith exposure level; RR_i = relative risk comparing ith exposure level with unexposed group ($i = 0$). See Bruzzi et al.⁵ and Miettinen⁹ for discussion and derivations; in Kleinbaum et al.²⁹ and Schlesselman.³⁰</p> |

tors.^{25(p608)} Such statements reflect misunderstanding about the meaning of the population attributable fraction. The proportion of patients exposed to the considered risk factor(s) is different from the population attributable fraction. In the Seidman et al. study, the proportions of breast cancer patients who had at least one of the considered factors were 0.76 in the 30 to 54-year age stratum and 0.82 in the 55 to 84-year age stratum.

Seidman et al. may have contributed to misinterpretations with the wording of their conclusion: "Given our current understanding of breast cancer risk factors, we are unable to identify...the 'causes' of more than about one-quarter of all cases."¹⁹ An average population attributable fraction

estimate of 0.25 across the two age strata means that 25% of the population risk of breast cancer would be eliminated if all 10 risk factors were to be eliminated from the population or, equivalently, that 25% of cases would be prevented following the risk factor eliminations. As just discussed, it does not mean that 25% of women who develop breast cancer will have one or more of the 10 risk factors; nor does it mean that epidemiologists can identify the cause(s) of breast cancer for a quarter of individuals with the disease. The population attributable fraction does not address probability of causation for a specific case of disease, nor does its estimation enable epidemiologists to discriminate between those cases caused by, and those not caused by,

the risk factors under consideration.

A more recent report on population attributable fractions and breast cancer risk factors has similarly been misinterpreted. Bruzzi et al.⁵ considered four established factors and estimated a population attributable fraction of 0.55. Referring to this estimate, a recent article included the following misstatement: "Another report estimates that 55 percent of breast cancers have one or more risk factors."^{26(p5)} In fact, as a result of the broad risk factor definitions used, 99% of the breast cancer cases in the Bruzzi et al. analysis involved one or more risk factors!

From a public health perspective, estimation of the population attributable fraction is of most use when the factor of interest is

clearly causally related to the end point and when there is consensus that the exposure is amenable to intervention. However, many researchers use risk factors that are surrogates for susceptibility attributes that may be unmodifiable (e.g., ethnicity, family history of cancer), as well as factors that are preclinical markers of disease (e.g., history of benign breast biopsy). Some factors included in population attributable fraction estimations are surrogates for more proximate exposures (e.g., poverty, educational level, marital status). Obviously, breast cancer risk will not be reduced by denying women a college education or a breast biopsy or by ensuring that all women marry, assuming that more causally proximate exposures and behaviors remain the same; however, these points are rarely discussed by investigators. The practical and logical limitations of including unmodifiable attributes, potential disease markers, and surrogate factors in population attributable fraction estimation are not always recognized.

Another issue related to interpretation of a population attributable fraction concerns specification of the exposed group. When modifiable risk factors are being considered in order to prioritize public health intervention strategies, the exposure cut point should be chosen so that the "unexposed" level is realistically attainable by those in the exposed category. Otherwise, the population attributable fraction may have theoretic value but will be of little practical public health value. Related to this point is Rose's observation that, for many chronic diseases, susceptibility for any disease is rarely confined to a high-risk minority within the population.²⁷ More typically, the majority of cases arise from the mass of the population with risk factor values around the population average. For many chronic diseases, population attributable fractions can be made high only by defining risk factors in such a way that almost the entire population is labeled "exposed" or "at elevated risk." The unrealistic implication of such broad exposure definitions is that virtually everyone in the population will need to be "shifted" to the lowest exposure category. This was the implication for the estimate presented by Bruzzi et al.⁵ It is also the case with the most recent estimate (0.41) of the summary population attributable fraction for three breast cancer risk factors²⁸; the authors estimate that 90% of US women have one or more of these established risk factors for breast cancer, and thus this large proportion of the population will need to be shifted on one or more of the factors if the estimated reduction in breast cancer burden is to be achieved.

A final philosophical point concerns the common practice of equating the population attributable fraction with the proportion of disease cases that are "explained" by the risk factors. For instance, after computing their population attributable fraction of 0.41, Madigan et al. stated that their estimates "suggest that a substantial proportion of breast cancer cases in the United States are explained by well-established risk factors."²⁸ This use of the word "explain" is somewhat misleading, since many readers probably equate "explain" with "cause." According to the Madigan et al. data, nearly the entire population of women in the United States has at least one of the considered risk factors. Since the vast majority of such "exposed" women will not develop breast cancer, stating that such factors explain a large proportion of breast cancer risk seems euphemistic. As an extreme example, if an age of greater than 15 years is considered a risk factor in a population attributable fraction estimation, virtually all cases of breast cancer can be "explained," in the technical sense of explaining variation in rates between the exposed (those more than 15 years of age) and the unexposed (those 15 years of age or younger); however, to imply that being more than 15 years of age "causes" breast cancer is of no value. Authors who present population attributable fractions should communicate clearly what they mean when they use phrases such as "explained by" and "attributable to," because there is potential for confusion on the part of both scientific and lay readers.

Conclusion

Many public health researchers are interested in evaluating the potential population impacts of identified risk factors. For some of these evaluations, estimation of the population attributable fraction is appropriate and valuable. The assumptions underlying valid population attributable fraction estimation include the following: a causal relationship between the risk factors and disease; the immediate attainment, among those formerly exposed, of the unexposed disease risk following elimination of the exposures; and independence of the considered risk factors from other factors that influence disease risk so that it is possible to conceive of changing the population distributions of the considered factors only. Such assumptions are often not justified. Those who present population attributable fractions have a duty to ensure that estimates are correctly computed and that their limited meaning is correctly communicated,

given the interest among researchers, clinicians, and the public in quantitative figures that attempt to summarize the state of etiologic knowledge about a disease. □

Acknowledgments

We thank Beth Gladen, Glinda Cooper, Bob Millikan, Bruce Levin, and the two anonymous reviewers for their thoughtful and very helpful reviews of the manuscript.

References

1. Northridge ME. Annotation: public health methods—attributable risk as a link between causality and public health action. *Am J Public Health*. 1995;85:1202–1203.
2. Coughlin SS, Benichou J, Weed DL. Attributable risk estimation in case-control studies. *Epidemiol Rev*. 1994;16:51–64.
3. Benichou J. Methods of adjustment for estimating the attributable risk in case-control studies: a review. *Stat Med*. 1991;10:1753–1773.
4. Greenland S, Robins JM. Conceptual problems in the definition and interpretation of attributable fractions. *Am J Epidemiol*. 1988;128:1185–1197.
5. Bruzzi P, Green SB, Byar DP, Brinton LA, Schairer C. Estimating the population attributable risk for multiple risk factors using case-control data. *Am J Epidemiol*. 1985;122:904–914.
6. Whittemore AS. Statistical methods for estimating attributable risk from retrospective data. *Stat Med* 1982;1:229–243.
7. Walter SD. Effects of interaction, confounding, and observational error on attributable risk estimation. *Am J Epidemiol*. 1983;117:598–604.
8. Levin ML. The occurrence of lung cancer in man. *Acta Union International Contra Cancrum*. 1953;9:531–541.
9. Miettinen O. Proportion of disease caused or prevented by a given exposure, trait, or intervention. *Am J Epidemiol*. 1974;99:325–332.
10. Wacholder S, Benichou J, Heineman EF, Hartge P, Hoover RN. Attributable risk: advantages of a broad definition of exposure. *Am J Epidemiol*. 1994;140:303–309.
11. Hsieh C-C, Walter SD. The effect of nondifferential exposure misclassification on estimates of the attributable and prevented fraction. *Stat Med*. 1988;7:1073–1085.
12. Walter SD. The estimation and interpretation of attributable fraction in health research. *Biometrics*. 1976;32:829–849.
13. Benichou J, Gail MH. Variance calculations and confidence intervals for estimates of the attributable fraction based on logistic models. *Biometrics*. 1990;46:991–1003.
14. Liao Y, Cooper RS, McGee DL, Mensah GA, Ghali JK. The relative effects of left ventricular hypertrophy, coronary artery disease, and ventricular dysfunction on survival among black adults. *JAMA*. 1995;273:1592–1597.
15. Donders GGG, Desmyter J, De Wet DH, Van Assche FA. The association of gonorrhoea and syphilis with premature birth and low birth-weight. *Genitourin Med*. 1993;69:98–101.

16. Hankin JH, Zhao LP, Wilkens LR, Kolonel LN. Attributable fraction of breast, prostate, and lung cancer in Hawaii due to saturated fat. *Cancer Causes Control*. 1992;3:17–23.
17. Mavalankar DV, Gray RH, Trivedi CR. Risk factors for preterm and term low birthweight in Ahmedabad, India. *Int J Epidemiol*. 1992;21:263–271.
18. Stroffolini T, Chiaramonte M, Tiribelli C, et al. Hepatitis C virus infection, HBsAg carrier state and hepatocellular carcinoma: relative risk and population attributable risk from a case-control study in Italy. *J Hepatol*. 1992;16:360–363.
19. Seidman H, Stellman SD, Mushinski MH. A different perspective on breast cancer risk factors: some implications of the nonattributable risk. *CA*. 1982;32:301–312.
20. Hahn RA, Eaker E, Barker ND, Teutsch SM, Sosniak W, Krieger N. Poverty and death in the United States—1973 and 1991. *Epidemiology*. 1995;6:490–497.
21. Rockhill B, Weinberg C. Error in population attributable risk calculation. *Epidemiology*. 1996;7:453.
22. Higginson J, Muir CS. Environmental carcinogenesis: misconceptions and limitations to cancer control. *JNCI*. 1979;63:1291–1298.
23. Wynder EL, Gori GB. Contribution of the environment to cancer incidence: an epidemiologic exercise. *JNCI*. 1977;58:825–832.
24. Freeman HP, Wasfie TJ. Cancer of the breast in poor black women. *Cancer*. 1989;63: 2562–2569.
25. Love SM. Use of risk factors in counseling patients. *Hematol Oncol Clin North Am*. 1989;3:599–610.
26. Garfinkel L. Perspectives on cancer prevention. *CA*. 1995;45:5–7.
27. Rose G. Sick individuals and sick populations. *Int J Epidemiol*. 1985;14:32–38.
28. Madigan MP, Ziegler RG, Benichou J, Byrne C, Hoover RN. Fraction of breast cancer cases in the United States explained by well-established risk factors. *JNCI*. 1995;87: 1681–1685.
29. Kleinbaum DG, Kupper LL, Morgenstern H. *Epidemiologic Research*. Belmont, Calif: Lifetime Learning Publications; 1982:163.
30. Schlesselman JJ. *Case-Control Studies: Design, Conduct, Analysis*. New York, NY: Oxford University Press Inc; 1982.

Lecture: Hazard Ratios



Hazard ratios

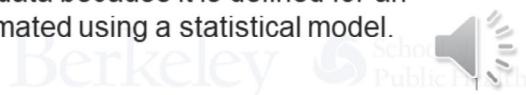
PHW250 B – Andrew Mertens



This video focuses on hazard ratios, and it will draw from an article by Hernán in 2010 called The Hazards of Hazard Ratios.

Quick review: hazard

- Hazard is defined as the **instantaneous potential for change in disease status** per unit of time at time t relative to the size of a disease-free population at time t
 - Synonyms for hazard: "force of morbidity", "force of mortality"
 - The **incidence density** averages over the hazard.
 - In a cohort study, the hazard at time t is defined as follows:
 - $$h(t) = \frac{P(\text{event in interval between } t \text{ and } [t + \Delta t] \text{ | at risk at } t)}{\Delta t}$$
- In practice, this means that it is calculated among survivors who did not develop the disease at earlier time points in a study.
- The hazard cannot be calculated directly from data because it is defined for an infinitely small period of time. Instead, it is estimated using a statistical model.



Let's start by quickly reviewing what the hazard is. First of all, hazard is a measure of disease, and it's defined as the instantaneous potential for change in disease status per unit of time at a specific time, t , relative to the size of the disease free population at time t .

Now that's quite a mouthful of a definition, so we're going to come back to that a little lower on the slide. But I want to note a few things that make the hazard different from incidence or risk. For one thing, it's *an instantaneous measure. It's not averaging over time. It's focusing on a very specific point in time that's infinitesimally small.

It's also relative to the disease-free population or the population at risk. Another way of calling that is the surviving population at a specific time.

Also to note, sometimes you'll see epidemiologists use the phrase force of morbidity, or force of mortality when they're referring to hazard.

The difference between the incidence density and the hazard is that the incidence density *averages over the hazard. The incidence density is spanning a longer period of time than the hazards since the hazard is an instantaneous measure.

So in a cohort study, how do we define the hazard?

*Well, hazard is indexed by a specific time, the specific time of follow up in the study. And so we denote that h (The h stands for hazard), h of t , and t is the index in time. And so h of t is equal to the probability that a disease event or mortality event occurs in the interval between t and t plus Δt . (So Δt is just a change in the time, so it might be from one day to two days or from one week to two weeks.) And as that Δt gets smaller and smaller, you'll get closer to the instantaneous hazard. So this numerator is the probability that an event occurs in that interval, conditional on the population at risk at time t , and then that conditional probability is divided by the change in Δt , so the change in time. This vertical bar on the right, this conditional part of this probability, "at risk at t " is written out in words below *here.

So this means that in practice, we're calculating our hazard among *survivors, people who did not develop the disease or did not die at earlier time points in the study. This is going to be a critical cause of some of the challenges we're going to talk about in this video when it comes to trying to make causal inferences about the hazard.

One other thing to just briefly review is that *the hazard can't be calculated directly from data, because it's defined for an infinitely small period of time. And instead, we use statistical models to estimate the hazard at specific times. So unlike for prevalence or cumulative incidence, you won't be hand-calculating hazards in this class.

Quick review: hazard ratios

- Hazard ratios = $\frac{\text{hazard among the exposed}}{\text{hazard among the unexposed}}$ 2 levels
- (For a binary, non-time varying exposure)
- Commonly estimated in cohort studies or trials in which the time to the disease or event of interest is measured as the outcome.



And then to further review using hazard as a measure of association, we can calculate something called a hazard ratio, which is just the ratio of the hazard among the exposed over the hazard among the unexposed.

Similarly, in a randomized trial, you could have the hazard in the treated divided by the hazard in the control group. This is defined for any binary non-time varying exposure. Binary exposure just means an *exposure of two levels, so exposed, unexposed as opposed to a continuous exposure, such as weight measured in kilograms. This measure of association is commonly estimated in cohort studies, and trials in which the outcome of interest is the time until the disease or event rather than just an indicator that the event occurred regardless of the time at which it occurred.

It is difficult to make causal inferences about hazard ratios even when:

1. There is no unmeasured confounding
2. There is no measurement error
3. The appropriate statistical modeling approach is used

This is because:

- Hazard ratios may change over time
- Hazard ratios have built in selection bias
 - Note: this can also be true of other measures of association (e.g. incidence density ratio) in cohort studies



The title of this paper, Hazards of Hazard Ratios, is getting at this fact that it's difficult to make causal inferences about hazard ratios, even when you meet several common criteria that from an epidemiologist standpoint make it possible to make a causal inference. The first is in a scenario where we have no unmeasured confounding, no measurement error, and when we use an appropriate statistical model.

Even if those three things are true, we still have to be very careful about making causal inferences about hazard ratios, and this is because of two main reasons. The first is that hazard ratios can change over time, and we're going to go into that in detail shortly. And the second is that hazard ratios have built in selection bias, and this gets back to that conditional statement that we saw in the previous slide in the definition for hazard.

Where it's measured at a specific time, t , conditional on the population that survived up to that time, and so some of the critiques that we're going to make of hazard ratio in this video are also true of other measures of association in cohort studies. And this is just because in cohort studies measuring hazard and incidence, we have selection bias built in because some people will be lost to follow up, or will die, or will no longer be at risk for one reason or another over time. And the people who remain in the study population may be systematically different from those who were lost at follow up.

Example: Women's health initiative

- Randomized trial
- ~16,000 women
- Treatment: hormone replacement therapy
- Primary outcome: heart disease
- Discontinued early due to concerns about the safety of the treatment



We're going to use the Women's Health Initiative as a case study for examining the challenges of hazard ratios. This was a well-known randomized trial, very large in size. Had around 16,000 women, and the treatment was hormone replacement therapy.

The trial randomized participants to receive that therapy or to a placebo. The primary outcome of interest was coronary heart disease. The trial was actually discontinued early due to concerns about the safety of the treatment

Women's health initiative main findings

- Overall HR = 1.24 ← Averages over time specific HRs
- Follow-up time 1 HR = 1.81 ←
- Follow-up time 2 HR = 1.34
- Follow-up time 3 HR = 1.27
- Follow-up time 4 HR = 1.25
- Follow-up time 5 HR = 1.45 ← →
- Follow-up time 6+ HR = 0.70 ←



ESTABLISHED IN 1812 AUGUST 7, 2003 VOL. 349 NO. 6

Estrogen plus Progestin and the Risk of Coronary Heart Disease
Julian E. Manson, M.D., Dr.P.H.; Judith Hsu, M.D.; Karen L. Johnson, M.D., M.P.H.; Jacques E. Rossouw, M.D.,
Ammar Al-Sabagh, Ph.D.; Nancy G. Lasser, M.D., Ph.D.; Maurizio Trevisan, M.D.; Henry R. Black, M.D.,
Susan R. Heckbert, M.D., Ph.D.; Robert Detrano, M.D., Ph.D.; Ora L. Strickland, Ph.D.; Nathan D. Wong, Ph.D.;
John R. Crouse, M.D.; Evan Stein, M.D.; and Mary Cushman, M.D., for the Women's Health Initiative Investigators*

ABSTRACT

BACKGROUND
Recent observational clinical trials have suggested that estrogen plus progestin does not confer cardiac protection and may increase the risk of coronary heart disease (CHD). In this report, we provide the final results with regard to estrogen plus progestin and CHD from the Women's Health Initiative (WHI).

METHODS
The WHI included a randomized primary prevention trial of estrogen plus progestin in 16,608 postmenopausal women who were 50–79 years of age at baseline. Participants were randomly assigned to receive conjugated equine estrogens (0.625 mg per day) plus medroxyprogesterone acetate (0.25 mg per day) or placebo. The primary efficacy outcome of the trial was CHD (nonfatal myocardial infarction or death due to CHD).

RESULTS
After a mean follow-up of 5.2 years (planned duration, 8.5 years), the data and safety monitoring board recommended terminating the estrogen-plus-progestin trial because of a significant excess of CHD events among women assigned to the intervention, with a hazard ratio for CHD of 1.24 (90–95 percent confidence interval, 1.00 to 1.54). The proportion of risk was more than 100% of the baseline risk ratio, 1.81 (95 percent confidence interval, 1.59 to 3.01), although excess risks at baseline were low. Women taking lipid-lowering therapy, higher baseline levels of C-reactive protein, other biomarkers, and other clinical characteristics did not significantly modify the treatment-related risk.

CONCLUSIONS
Estrogen plus progestin does not confer cardiac protection and may increase the risk of CHD among generally healthy postmenopausal women, especially during the first year after the initiation of hormone use. This treatment should not be prescribed for the prevention of cardiovascular disease.

N Engl J Med 2003;349:51–59.
Copyright © 2003 Massachusetts Medical Society.
This journal is available online at www.nejm.org. Visit the journal at www.nejm.org.

Here's a screenshot of the front page of the article and the abstract. Right *here in red is the main result that was reported, so this is the high-level result that, for example, journalists and maybe policymakers would use as the main finding for the paper.

And it's a hazard ratio for coronary heart disease, comparing the intervention to the control, of *1.24. That's from one follow up year all the way up to six years or more. When we look at the hazard ratio stratified over time, we see that it starts out at 1.81 in the first to follow up time point. And then as we move down to *follow up time five, the hazard ratio is 1.45. Follow up time *six or more years, the hazard ratio is actually 0.7 which looks protective.

And so the key point here to take away is just that it's very common when we report the result of a trial or really any study to have a single number serve as our main finding. In this case, HR of 1.24. This hazard ratio though is *averaging over a wide period of time. And when we stratify time, we can see that the hazard ratio at specific follow up times varies quite a bit and even spans the null value, which suggests that this pooled measure is not really telling us a comprehensive story about what happened in this trial.

Values of HRs depend upon the follow-up time that they average over

- Overall HR = 1.24
- Follow-up time 1 HR = 1.81
- Follow-up time 2 HR = 1.34
- Follow-up time 3 HR = 1.27
- Follow-up time 4 HR = 1.25
- Follow-up time 5 HR = 1.45
- Follow-up time 6+ HR = 0.70

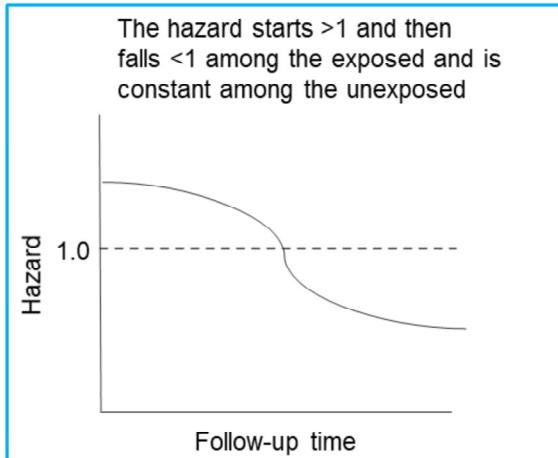
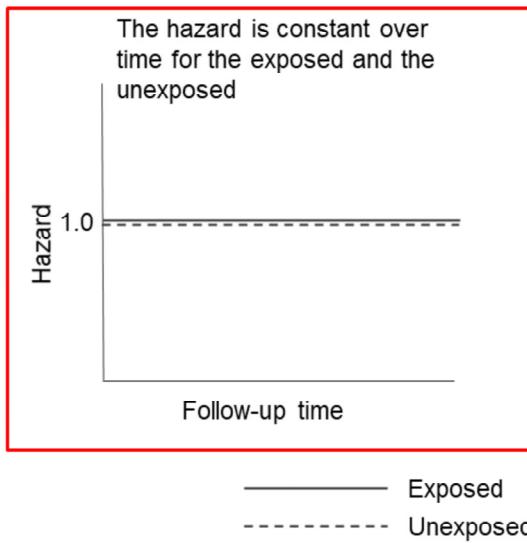
If there had only been 2 years of follow-up, the overall HR would have been ~1.58 instead of 1.24



Another way of thinking about this is the trial was designed to go for more than six years, but what if the investigators had only planned a two year follow up period? Well, if they had done that, then their overall hazard ratio would have been 1.58 instead of 1.24. So this is really just to show you that the value of a hazard ratio really depends upon the amount of follow up time because of the fact that an averaged hazard ratio is capturing hazard that may change greatly over the course of follow up.

So you can make a really different inference about a study, depending on what period of follow up you decide to average over. This makes it tricky to really come up with a clear picture for what happened in the trial.

In both plots, the HR = 1



Take home message: HRs that average over long periods of time may obscure important time patterns in the hazard by exposure group

Another way of looking at this is by plotting the hazard, here on the y-axis against the follow up time on the x-axis, and so *on the left, we have a plot where the hazard in the exposed and in the unexposed is constant over time. This gives us a hazard ratio of one that is constant over entire follow up time, but this is actually rarely what we see. It's much more common to see something like we have on the right *here where the hazard in one group, in this case, the exposed is fluctuating over time.

In this example, the hazard in the unexposed is flat. As a result, we can get an average hazard ratio of one because overall, the hazard in the exposed is sort of evenly above and below the null of one. But these two different plots have very different patterns, and so averaging over the hazard in this follow up time period is obscuring important time specific patterns in that hazard.

Can we just report period-specific HRs?

- No — they have selection bias built in.
- **Definition of hazard:** the instantaneous potential for change in disease status per unit of time at time t relative to the size of a disease-free population at time t
 - Overall HR = 1.24
 - Follow-up time 1 HR = 1.81
 - Follow-up time 2 HR = 1.34
 - Follow-up time 3 HR = 1.27
 - Follow-up time 4 HR = 1.25
 - Follow-up time 5 HR = 1.45
 - Follow-up time 6+ HR = 0.70

systematically
different

Is this truly a protective effect
of the intervention?



You may be thinking that one solution to this would be instead of reporting a pooled hazard ratio, we should just report period specific hazard ratios. But unfortunately, these period specific hazard ratios also have selection bias built in. And this can be seen in the definition of hazard.

The definition of hazard is the instantaneous potential for change in disease status per unit of time at time t relative to the size of a disease-free population at time t . This “disease-free population at time t ” is the important part.

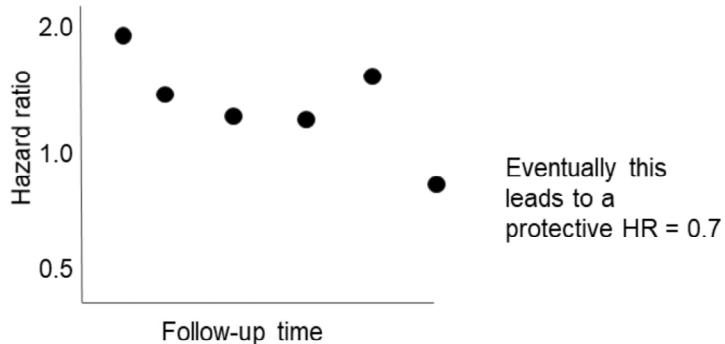
What we end up seeing *here is we get this protective effect of 0.7 for follow up time six or more years. And do we really believe that this is a truly protective effect of the intervention given the pattern that we see earlier? Is this biologically plausible?

Well, without getting into the details of the biology in this particular trial, from an epidemiologist standpoint, we are concerned that the people who remain in the study at six or more years are just *systematically different from those at previous follow up times. And this is because people who are still in the study at time six or more years are the ones who haven't developed coronary heart disease and haven't passed away. The people who are at the greatest risk of a deleterious effect of the intervention are likely to be lost to follow up or censored starting at time one, time two, or three, and then they're no longer in the population at times six or more. As a result, this hazard ratio of 0.7 could just be protective due to bias rather

than a truly protective hazard ratio.

HRs have selection bias built in, so just reporting period-specific HRs is not enough to account for this bias

Each year there are fewer women left who are susceptible to heart disease and who are harmed by the treatment

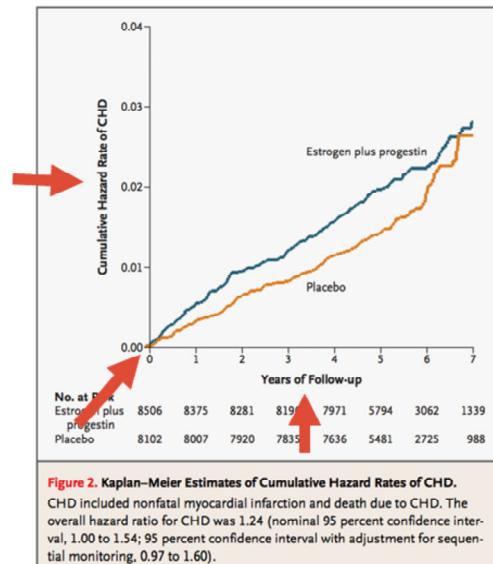


Another way of looking at this is just if we plot this here, so we have on the y-axis is the hazard ratio. And on the x-axis is follow up time, and basically, what happens is over time, the women who are left are the ones who are still susceptible to heart disease.

And each year, there are fewer and fewer women who are left who are susceptible to heart disease, and who may be harmed by the treatment. Those most likely to be harmed by the treatment we're already had coronary heart disease earlier on the study, and so that leads us to get this protective HR of 0.7.

What is the proper way to analyze hazard in a cohort study?

1. Plot Kaplan-Meier curve of cumulative risk, hazard, or survival
 - Plots the proportion of the study population that is disease-free by follow-up time
 - Provides information about the absolute risk that the HR does not
 - Typically unadjusted, but you can also adjust for confounders (see Hernan 2010)



Manson et al., 2003. N Engl J Med 2003;349:523-34.

So then what's the proper way to analyze hazard in a cohort study or a trial in this case?

There's something called a Kaplan Meier curve, and we see this here on the right. This is a figure from that same paper from the Women's Health Study. What this plot does is on the *y-axis, we have the cumulative hazard, or the cumulative risk, or sometimes survival. *And on the x-axis, we have follow up time.

It's cumulative, because we're looking at the total as it increases or decreases over time. In this particular plot, we start off at* time zero with zero individuals in the study with coronary heart disease. And then as additional cases are identified, the value goes up in the intervention group here in blue and in the placebo control group here in yellow.

And so it grows higher and higher, because it's a cumulative hazard rate of coronary heart disease. What's nice about these plots is *that it provides information about the absolute risk or the absolute hazard that the hazard ratio does not. So we see a different pattern here than we saw just by looking at the hazard ratios over time.

These plots are typically unadjusted for confounders, but there are methods available that allow you to adjust for confounders if you suspect that there is systematic differences between the exposed and unexposed groups. In this case,

this is randomized trial, so the plot is the unadjusted cumulative hazard.

But the hazard ratios themselves were adjusted for a few variables, and so that's why the pattern differs a little bit from what we see in this Kaplan Meier plot. You may remember the name Kaplan Meier from our measures of disease unit, because it's the method that we learned for calculating cumulative incidents. And essentially, the data that you get in a table that you would use to calculate cumulative incidence under the Kaplan Meier method could just be plotted like this, so those two methods kind of go hand in hand.

What is the proper way to analyze hazard in a cohort study?

2. Estimate and report a set of HRs that having an increasing period of follow-up

- HR for 1 year = 1.8
- HR for 1-2 years = 1.7
- HR for 1-5 years = 1.2
- (But it's still a good practice to complement this with a survival or hazard curve)



Another option for analyzing hazards in a cohort study or trial is to estimate and report a set of hazard ratios that have an increasing period of follow up. So instead of just giving the hazard ratio at year one, at year two, at year three, and so on, we could have a hazard ratio for year one. And then a hazard ratio of for years including one and two, and hazard ratio for years one through five.

And so if we do this, we see that it's still a hazard ratio that's getting closer to the null, but we're not seeing the numbers get protected quite yet. Even doing this though, it's a useful thing to do. It's still a good practice to complement this with a Kaplan Meier survival or hazard curve as we discussed in the last slide.

Summary of key points

- Hazard ratios are commonly the primary measure of association in cohort studies.
- They are calculated among survivors who did not develop the disease at earlier time points in a study. As a result, they have selection bias built in.
- As a result, we must be careful about making causal inferences about hazard ratios even when 1) there is no unmeasured confounding, 2) there is no measurement error, and 3) the appropriate statistical modeling approach is used.
- Pooled HRs must be reported and interpreted with caution because they can obscure meaningful variation in hazard over time.
- Kaplan-Meier curves provide helpful complementary information to HRs related to the absolute risk / survival of events over time.



To summarize, it's very common in cohort studies and in some randomized trials to estimate hazard ratios as the primary measure of association. And they're calculated among survivors, people who did not develop the disease at an earlier time point in the study. And as a result, selection bias is built into these measures of association.

So even when we have no one measured confounding, no measurement error, and we use the appropriate statistical modeling approach, you have to be really careful about trying to make causal inferences about these hazard ratios even in a trial.

Pooled hazard ratios must be reported and interpreted with great caution, because they can obscure meaningful variation in the hazard over time as we saw in some examples of plots in this video. And Kaplan Meier curves are a helpful set of complementary information that are good to report alongside hazard ratios themselves, because they provide information about the absolute risk or the absolute survival over time.



NIH Public Access

Author Manuscript

Epidemiology. Author manuscript; available in PMC 2013 May 14.

Published in final edited form as:

Epidemiology. 2010 January ; 21(1): 13–15. doi:10.1097/EDE.0b013e3181c1ea43.

The Hazards of Hazard Ratios

Miguel A. Hernán

Department of Epidemiology, Harvard School of Public Health, 677 Huntington Avenue, Boston, MA 02115

Miguel A. Hernán: mhernan@hsph.harvard.edu

The hazard ratio (HR) is the main, and often the only, effect measure reported in many epidemiologic studies. For dichotomous, non-time-varying exposures, the HR is defined as the hazard in the exposed groups divided by the hazard in the unexposed groups. For all practical purposes, hazards can be thought of as incidence rates and thus the HR can be roughly interpreted as the incidence rate ratio. The HR is commonly and conveniently estimated via a Cox proportional hazards model, which can include potential confounders as covariates.

Unfortunately, the use of the HR for causal inference is not straightforward even in the absence of unmeasured confounding, measurement error, and model misspecification. Endowing a HR with a causal interpretation is risky for 2 key reasons: the HR may change over time, and the HR has a built-in selection bias. Here I review these 2 problems and some proposed solutions. As an example, I will use the findings from a Women's Health Initiative randomized experiment that compared the risk of coronary heart disease of women assigned to combined (estrogen plus progestin) hormone therapy with that of women assigned to placebo.¹ By using a randomized experiment as an example, the discussion can focus on the shortcomings of the HR, setting aside issues of confounding and other serious problems that arise in observational studies.

The Women's Health Initiative followed over 16,000 women for an average of 5.2 years before the study was halted due to safety concerns. The primary result from the trial was a HR. As stated in the abstract 1 and shown in Table 1 of the article, “Combined hormone therapy was associated with a hazard ratio of 1.24.”¹ In addition, Table 2 provided the HRs during each year of follow-up: 1.81, 1.34, 1.27, 1.25, 1.45, and 0.70 for years 1, 2, 3, 4, 5, and 6 or more, respectively. Thus, the HR reported in the abstract and Table 1 can be viewed as some sort of weighted average of the period-specific HRs reported in Table 2.

This brings us to Problem 1: although the HR may change over time, some studies report only a single HR averaged over the duration of the study's follow-up. As a result, the conclusions from the study may critically depend on the duration of the follow-up. For example, the average HR in the WHI would have been 1.8 if the study had been halted after 1 year of follow-up, 1.7 after 2 years,² 1.2 after 5 years, and—who knows—perhaps 1.0 after 10 years. The 24% increase in the rate of coronary heart disease that many researchers and journalists consider as *the* effect of combined hormone therapy is the result of the arbitrary choice of an average follow-up period of 5.2 years. A trial with a shorter follow-up could have reported an 80% increase, whereas a longer trial might have found little or no increase at all.

The magnitude of the average HR depends on the length of follow-up because the average HR ignores the distribution of events during the follow-up. The average HR can take the value 1.0 if the hazard in the exposed is identical to the hazard in the unexposed during the entire follow-up, or if the hazard in the exposed is higher during, say, the first 5 years and

NIH-PA Author Manuscript NIH-PA Author Manuscript NIH-PA Author Manuscript

lower afterward. Incidentally, the same problem arises whether the average HR is directly estimated in a cohort study, as discussed here, or estimated via the odds ratio of a properly designed case-control study with incidence density sampling. One might then conclude that we should forget about the average HR and restrict our attention to the period-specific HRs, which seem to capture the potentially time-varying magnitude of the effect. This brings us to Problem 2: the period-specific HRs have a built-in selection bias. To describe the bias, consider that the (discrete-time) hazard during period t is defined as the risk of the outcome during period t among those who reached period t free of the outcome. In the Women's Health Initiative, the calculation of the HR during year t was restricted to women who did not develop coronary heart disease—the “survivors”—between baseline and the beginning of year t . The HR after year 5 was 0.7, which means that the disease rate after year 5 was lower in the treatment arm (the hazard in the numerator of the HR) than in the placebo arm (the hazard in the denominator).

However, this apparently protective effect of hormone therapy after year 5 is hardly surprising if one bears in mind that women vary in their susceptibility to heart disease. A certain proportion of all women enrolled in the trial were particularly prone to develop heart disease if they were exposed to hormone therapy or other factors (for simplicity, let's refer to them as the “susceptible women”). The proportion of susceptible women in the trial was of course unknown but, because of randomization, it was expected to be the same in both the treatment and placebo arms at baseline. However, these susceptible women were preferentially excluded from the treatment arm as they developed heart disease over time—precisely because they were assigned to a therapy with harmful effects to which they were susceptible (all other factors to which they were susceptible were expected to be equally distributed between the 2 arms). With time, the proportion of susceptible women progressively increased in the placebo arm compared with the treatment arm. The bias due to the differential selection of less susceptible women over time, because of differential depletion of susceptibles, is the built-in selection bias of period-specific HRs. This bias may explain that the HR after year 5 is less than 1.0 even if hormone therapy has no truly preventive effect in any woman at any time. This built-in selection bias of the HR has also been described using causal diagrams.^{3,4}

In short, the average HR may be uninformative because of potentially time-varying period-specific HRs, and because the period-specific HRs may be time-varying because of built-in selection bias. These problems can be overcome by summarizing the study findings as appropriately adjusted survival curves, where the survival at time t is defined as the proportion of individuals who are free of disease through time t . Another alternative not discussed here is the comparison of the distribution of survival times between the exposed and the unexposed, which can be accomplished by using accelerated failure time models⁵ rather than Cox models.

Because of the shortcomings of the HR, the analysis of randomized experiments routinely include Kaplan-Meier survival curves—or their complement, the cumulative risk curve (see Figure 2 of the Women's Health Initiative trial report¹). In contrast (and despite multiple warnings in the epidemiologic literature^{3–6}), the analysis of observational follow-up studies are commonly summarized by HRs only. A possible explanation for this practice in observational studies is the need to deal with confounding. The HRs presented in observational studies are not simply the hazard in the exposed divided by the hazard in the unexposed. Rather, these HRs are adjusted for measured confounders by using regression models, inverse probability weighting, or other methods. Unadjusted HRs would be of little use for causal inference from observational data, as would unadjusted survival curves. It is not unexpected that most epidemiologic articles include HRs only, because epidemiology students are traditionally taught to estimate adjusted HRs but not adjusted survival curves.⁷

The next paragraph sketches a general procedure to obtain survival curves adjusted for baseline confounders.

First, fit a discrete-time hazards model (eg, a pooled logistic model with relatively short periods) that estimates, at each time and for each person, the conditional probability of remaining free of the outcome given exposure, baseline covariates, and time of follow-up. Allow for time-varying hazards by modeling the variable “time of follow-up,” using a flexible functional form (eg, cubic splines), and for time-varying HRs by adding product terms between exposure and “time of follow-up.” Second, for each subject, multiply the model’s predicted values through time t to estimate the survival at t for subjects with their same combination of covariate values. One can then construct conditional (adjusted) survival curves under the conditions of exposure and no exposure for each observed combination of values of the baseline covariates (in randomized trials, the survival curves are unconditional or marginal, ie, averaged over all the individuals irrespective of their covariate values). Third, predict the survival at time t for each subject both under exposure and under no exposure, regardless of the subject’s exposure status. Fourth, separately average the conditional survivals under exposure and under no exposure, over all subjects. This last step effectively standardizes the curves to the empirical distribution of the covariates in the study, and results in 2 marginal survival curves: one under exposure, another under no exposure.

The above procedure can be extended in a number of ways. In settings with time-varying exposures and confounders, the procedure can be combined with inverse probability weighting of the hazards model. This procedure has been used to present adjusted survival curves under continuous use (“always exposed”) and no use of hormone therapy (“never exposed”) in the analysis of both observational studies⁸ and randomized experiments⁹ in which time-varying exposures arise when considering adherence-adjusted analyses. In settings with continuous rather than dichotomous exposures, the procedure requires the choice of a finite number of levels of exposure to be compared (“always versus never exposed” will not do).¹⁰ One may then construct as many survival curves as there are exposure levels of interest. For continuous and time-varying exposures one needs to be especially careful about dose-response assumptions. Sensitivity analyses can be used to evaluate the possibility of model extrapolation beyond the observed data. Confidence intervals for the survival curves can be obtained by bootstrapping.

So should we outlaw the use of HRs in epidemiologic studies? Of course not. A single average HR through t may be misleading, as explained above, but a single survival probability at t could be as misleading because both measures ignore the distribution of events between baseline and t . On the other hand, a series of average HRs for increasingly longer periods of follow-up is informative. For example, in the WHI the average HRs for 1, 2, and 5 years were approximately 1.8, 1.7, and 1.2, which indicates that hormone therapy increases the cumulative risk of heart disease in the early part of the follow-up but probably not much over longer periods. The same conclusion is drawn from the survival curves for the treatment and placebo groups, which converge after 8 years. In mortality studies with sufficiently long follow-up, the survival probabilities in both groups are ensured to reach the value 0, and the average HR is ensured to reach the value 1.

An advantage of the survival curves over a series of average HRs is that the survival curves provide information about the absolute risks. For example, in the Women’s Health Initiative, the average HR of 1.8 during year 1 means that the one-year risk was about 0.49% in the treatment group and 0.28% in the placebo group rather than, say, 49% versus 28%. An advantage of the average HRs over the survival curves is the readiness with which confidence intervals can be computed in standard software. What about period-specific

HRs? Their built-in selection bias makes them difficult to interpret as a measure of time-varying effect. For example, in the Women's Health Initiative, the HR goes from greater than 1.0 to less than 1.0 after year 5—that is, the hazards of the treatment and the placebo groups cross at about year 5. However, this crossing of hazards is essentially meaningless from a practical standpoint. What really matters is that the survival is lower in the placebo group compared with the treatment group until at least year 8. Hazards may cross at some point during the follow-up because of depletion of susceptibles even if the survival curves never cross. Cumulative measures, such as a series of average HRs or survival curves, are needed to summarize the data in a meaningful way. On the other hand, period-specific HRs are useful as an intermediate step to estimate survival curves in the procedure described above.

In summary, survival curves are more informative than HRs and can be easily generated. It would not be a bad thing to see them more widely used in observational studies.

References

1. Manson JE, Hsia J, Johnson KC, et al. Estrogen plus progestin and the risk of coronary heart disease. *N Engl J Med.* 2003; 349:523–534. [PubMed: 12904517]
2. Prentice RL, Pettinger M, Anderson GL. Statistical issues arising in the Women's Health Initiative. *Biometrics.* 2005; 61:899–911. [PubMed: 16401257]
3. Hernan MA, Hernandez-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology.* 2004; 15:615–625. [PubMed: 15308962]
4. Flanders WD, Klein M. Properties of 2 counterfactual effect definitions of a point exposure. *Epidemiology.* 2007; 18:453–460. [PubMed: 17473709]
5. Hernan MA, Cole SR, Margolick JB, Cohen MH, Robins JM. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiol Drug Saf.* 2005; 14:477–491. [PubMed: 15660442]
6. Greenland S. Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology.* 1996; 7:498–501. [PubMed: 8862980]
7. Cole SR, Hernan MA. Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed.* 2004; 75:45–49. [PubMed: 15158046]
8. Hernan MA, Alonso A, Logan R, et al. Observational studies analyzed like randomized experiments: An application to postmenopausal hormone therapy and coronary heart disease (with discussion). *Epidemiology.* 2008; 19:766–779. [PubMed: 18854702]
9. Toh SG, Hernandez-Díaz S, Logan R, Rossouw JE, Hernan MA. Coronary heart disease in postmenopausal users of estrogen plus progestin hormone therapy: Does the increased risk ever disappear? *Ann Intern Med.* In Press.
10. Zhang Y, Thamer M, Cotter D, Kaufman J, Hernán MA. Estimated effect of epoetin dosage on survival among elderly hemodialysis patients in the United States. *Clin J Am Soc Nephrol.* 2009; 4:638–644. [PubMed: 19261818]

Lecture: Causal Perspective on Measures of Association



Causal perspective on measures of association

PHW250 B – Andrew Mertens



To close out this unit on measures of association, we'll talk about the traditional perspectives and the modern perspectives on making causal inferences with different types of measures of association.

Traditional vs. modern perspectives on measures of association and causal inference

- • In the traditional perspective:
 - • Relative measures (RR, OR) provide information about etiology and causation
 - • Absolute measures (RD) provide information about public health impact
 - • These "epidemiologic traditions" have been "handed down from one generation to the next, without citation or critical reflection, as though their truth were self-evident." — Poole, 2010. *Epidemiology* 21(1).
 - • This assumption was tested in a heated controversy about smoking and lung cancer in the early 1950's and 1960's.
- • In the modern perspective, epidemiologists use a simple model of the different "types" of people in the population to distinguish between the utility of the RR vs. the RD in making causal inferences.



*Traditionally, epidemiologists have been taught that relative measures, such as *the relative risk or the odds ratio, provide the best information about the etiology or causal process for disease. *On the other hand, they've been taught that absolute measures, such as the risk difference provide the best information about public health impact, but aren't what we should be using to try to make causal inferences. *Poole has this nice quote that says, "These epidemiologic traditions have been handed down from one generation to the next without citation or critical reflection as though their truth were self-evident."

In one of the assigned readings that goes along with this module, his article from 2010 talks about how *this assumption behind this traditional perspective led to a pretty heated controversy in the smoking and lung cancer research that was going on in the early 1950s and 1960s. It's actually quite a nice case study of how different types of measures can lead to different inferences.

*In this video, we're going to talk about the modern perspective. And we're going to use a simple model for making causal inferences from relative versus absolute scale measures to distinguish their utility.

Simple model to distinguish causation from association

| Type | Response ^a under | | Description | Proportion of types |
|------|-----------------------------|-------------|------------------------|---------------------|
| | Exposure | Nonexposure | | |
| 1 | 1 | 1 | Doomed | p_1 |
| 2 | 1 | 0 | Exposure is causal | p_2 |
| 3 | 0 | 1 | Exposure is preventive | p_3 |
| 4 | 0 | 0 | Immune | p_4 |

- • “Doomed”: disease occurs no matter what
- • “Causal”: disease occurs if exposed
- • “Preventive”: disease occurs if unexposed
- • “Immune”: disease never occurs



Rothman et al., *Modern Epidemiology*, 3rd Ed.

In this model, there are four types of people. *Type one is someone who we will call doomed. And these are people who, unfortunately, will get disease no matter whether they are exposed or unexposed.

So this is indicated *here in the exposure column where we're seeing a 1. That means that they experience disease. They also have a 1 under the nonexposure column, which means they also experience disease when they're not exposed. So this is a doomed person. And we can label the proportion of them in the population as P1.

*Next is the second causal type called the causal group. And in this group, the exposure is causal. *So this person will develop disease if they are exposed and will not develop disease if they're unexposed. The proportion of people who are this causal type is labeled as P2.

*We call type three, the preventive type. These are a relatively unusual group.

*They only develop disease if they are unexposed. If they're exposed, they do not have disease. And we label the proportion of preventive types as P3.

*Then there's the last type who are immune. No matter what, whether they're exposed or unexposed, they never get the disease. *And the proportion of immune people is labeled as P4.

Calculate the incidence for the exposed and unexposed

| Type | Exposure | Nonexposure | Description | Proportion of types | R_{exp} |
|------|----------|-------------|------------------------|---------------------|-----------------------------------|
| 1 | 1 | 1 | Doomed | p_1 | $p_1 + p_2 + p_3 + p_4$ |
| 2 | 1 | 0 | Exposure is causal | p_2 | $p_1 + p_2$ |
| 3 | 0 | 1 | Exposure is preventive | p_3 | |
| 4 | 0 | 0 | Immune | p_4 | |

So let's calculate *the risk under the exposure. So if we sum up these different people in the population-- P_1 plus P_2 plus P_3 plus P_4 , what happens is * P_3 and P_4 don't develop the disease when they're exposed, and so they drop out of the formula. And so the risk under the exposure is P_1 plus P_2 .

Calculate the incidence for the exposed and unexposed

| Type | Exposure | Nonexposure | Description | Proportion of types | $R_{exp} = p_1 + p_2 + p_3 + p_4$ |
|------|----------|-------------|------------------------|---------------------|--|
| 1 | 1 | 1 | Doomed | p_1 | $= p_1 + p_2$ |
| 2 | 1 | 0 | Exposure is causal | p_2 | |
| 3 | 0 | 1 | Exposure is preventive | p_3 | $R_{unexp} = p_1 + \cancel{p_2} + \cancel{p_3} + \cancel{p_4}$ |
| 4 | 0 | 0 | Immune | p_4 | $= p_1 + p_3$ |

We can do the same thing for the unexposed scenario. We start with the same four summed up together, but then only P1 and P3 develop the disease when they are not exposed.

Calculate the incidence for the exposed and unexposed

| Type | Exposure | Nonexposure | Description | Proportion of types |
|------|----------|-------------|------------------------|---------------------|
| 1 | 1 | 1 | Doomed | p_1 |
| 2 | 1 | 0 | Exposure is causal | p_2 |
| 3 | 0 | 1 | Exposure is preventive | p_3 |
| 4 | 0 | 0 | Immune | p_4 |

$$R_{\text{exp}} = p_1 + p_2 + p_3 + p_4$$

$$= p_1 + p_2$$

$$R_{\text{unexp}} = p_1 + p_2 + p_3 + p_4$$

$$= p_1 + p_3$$

$$\text{RR} = (p_1 + p_2) / (p_1 + p_3)$$

$$\text{RD} = (p_1 + p_2) - (p_1 + p_3) = p_2 - p_3$$

From here, we can use these quantities to estimate the relative risk and the risk difference. So the relative risk compares the risk among the exposed to the risk in the unexposed and divides them-- so P1 plus P2 divided by P1 plus P3. And the same is true for the risk difference except, instead of dividing, we subtract. The P1's cancel out, which leaves us with P2 minus P3 for the risk difference.

Simple model to distinguish causation from association

| Type | Response ^a under | | Description | Proportion of types |
|------|-----------------------------|-------------|------------------------|------------------------|
| | Exposure | Nonexposure | | |
| 1 | 1 | 1 | Doomed | $p_1 = 10\%$ |
| 2 | 1 | 0 | Exposure is causal | $p_2 = 50\%$ |
| 3 | 0 | 1 | Exposure is preventive | $p_3 = 10\%$ |
| 4 | 0 | 0 | Immune | $p_4 = 30\%$ |

If all exposed (N=100)

| | Disease | No disease |
|-----------|---------|------------|
| Exposed | 10+50 | 30+10 |
| Unexposed | 0 | 0 |

If all unexposed (N=100)

| | Disease | No disease |
|-----------|---------|------------|
| Exposed | 0 | 0 |
| Unexposed | 10+10 | 30+50 |

Summing up causal types:

$$RD = (p_1 + p_2) - (p_1 + p_3) = p_2 - p_3 = 0.5 - 0.1 = 0.4$$

Using 2x2 tables:

$$RD = 60/100 - 20/100 = 0.4$$



Rothman et al., Modern Epidemiology, 3rd Ed.

Let's go over an example that links this layout to *the 2 by 2 tables that were more used to. *So let's say that the proportion of each type is 10% for doomed, 50% for the causal type, 10% for preventive, and 30% for those who are immune. At the first 2 by 2 table on the right side, let's imagine a scenario in which everyone is exposed. So this is sort of a counterfactual scenario. And let's say that our sample size is 100.

In our exposed scenario, *we're only looking at the top row of the 2 by 2 table because everyone is exposed. And then the number of people who are diseased is equal to P1 times 100, which is 10, plus P2 times 100, which is 50. And the number who are not diseased is P3 times 100, which is 10, and P4 times 100, which is 30.

We can do the same thing for the second 2 by 2 table in which *everyone is unexposed with the same sample of 100 people. In this scenario, the number of people who are diseased when they're not exposed is the sum of P1 times 100, which is 10 and P3 times 100, which is also 10. And then the remaining of people in the population do not get the disease.

*Then, over here on the left, we can sum up the causal types, which is what we did on the previous slide. The risk difference, for example, is P2 minus P3. So that's 0.5 minus 0.1 equals 0.4. *To get the risk difference using the 2 by 2 table method that we're more familiar with, we take the risk in the exposed, which is 60-- summing up 10 and 50 out of 100-- minus the risk in the unexposed, which is 10

plus 10-- 20/100. That gives us 0.4. We get the same answer from these two different methods. And this slide was just to illustrate how the 2 by 2 tables link to the way that we're doing our calculation from this video using the causal types.

Interpreting “no effect” under this model

| Type | Response ^a under | | Description | Proportion of types |
|------|-----------------------------|-------------|------------------------|---------------------|
| | Exposure | Nonexposure | | |
| 1 | 1 | 1 | Doomed | p_1 |
| 2 | 1 | 0 | Exposure is causal | p_2 |
| 3 | 0 | 1 | Exposure is preventive | p_3 |
| 4 | 0 | 0 | Immune | p_4 |

$$RR = (p_1 + p_2)/(p_1 + p_3)$$

$$RD = p_2 - p_3$$

- If there are equal causal and preventive types, $p_2 = p_3$ and $RR = 1$ and $RD=0$
- Thus $RR=1$ and $RD=0$ does not necessarily mean no effect but rather equality of causal types and preventive types.
- No effect would require $p_2 = p_3 = 0$.
- $RR=1$ or $RD=0$ are more accurately described as “no net effect” than “no effect”.



Rothman et al., *Modern Epidemiology*, 3rd Ed.

When we formulate the relative risk and the risk difference using this model, a few things come to light. First is that, if there are an equal number of causal types and preventive types-- so that's the P2 and the P3 types. That would mean that P2 equals P3, and as a result, the relative risk would be equal to 1, and the risk difference would be equal to 0.

Now, often, when we get a relative risk of 1 or a risk difference of 0, we just immediately interpret that as no effect. But in this formulation, what we can see is there may still be a P2 greater than 0 and a P3 greater than 0. It's just that they're equal, and that's why we're getting a null finding. This brings light to the fact that, when a relative risk is equal to 1, or a risk difference is equal to 0, it doesn't necessarily mean there's no effect, but rather, that we have an equality of causal types and preventive types.

To have no effect we require us to have P2 and P3 equal to 0. It's more accurate to describe a relative risk of 1 or a risk difference of 0 as no net effect rather than no effect.

Influence of the “doomed” category on the RR

Absolute measures

| Type | Response ^a under | | Description | Proportion of types |
|------|-----------------------------|-------------|------------------------|---------------------|
| | Exposure | Nonexposure | | |
| 1 | 1 | 1 | Doomed | p_1 |
| 2 | 1 | 0 | Exposure is causal | p_2 |
| 3 | 0 | 1 | Exposure is preventive | p_3 |
| 4 | 0 | 0 | Immune | p_4 |

Types amenable to intervention

Example 1: $p_1 = 0.75$, $p_2 = 0.19$, $p_3 = 0.01$, $p_4 = 0.05$

- $RR = (0.75 + 0.19) / (0.75 + 0.01) = 1.24$
- $RD = 0.19 - 0.01 = 0.18$

Example 2: $p_1 = 0.05$, $p_2 = 0.19$, $p_3 = 0.01$, $p_4 = 0.75$

- $RR = (0.05 + 0.19) / (0.05 + 0.01) = 4$
- $RD = 0.19 - 0.01 = 0.18$

$$RR = (p_1 + p_2) / (p_1 + p_3)$$

$$RD = p_2 - p_3$$

- The RR depends not only on the people amenable to intervention (types 2 and 3) but also on the people who are doomed.
- The RD only depends on people amenable to intervention.
- If the proportion of the doomed in the population is high, the causal effect is diluted compared to when it is lower. In both cases, the RD stays the same.
- For this reason, some have argued that the RD is a better measure of causal impact than the RR.

This slide gets to the key point behind the modern view on the benefits of absolute measures.

Looking at our table here on the left, type two and type three are really the people we're most interested in. These are the people who are amenable to intervention. If they're exposed, that will affect their disease risk, and if they're not exposed, it will also affect their disease risk. Whereas, for the doomed people, they'll be disease no matter what. The immune people will never be diseased.

From a public health perspective, we're just much more interested in type two and type three because there isn't anything we can do to change the outcome for type one and type four. Now, keeping that in mind, let's look at our formulas for the relative risk and the risk difference.

The relative risk here has P_1 , the doomed people, both in the numerator and in the denominator. And that means that the relative risk depends, not only on the types that we're interested in, the causal and preventive types, but it also depends on the proportion of people in the population who are doomed. But when we look at the risk difference, it's just restricted to P_2 and P_3 , the people who are amenable to intervention.

As a result of this mathematical fact, if the proportion of people who are doomed in the population is high, the causal effect may be diluted in a relative risk compared

to when it is lower. But in both cases, the risk difference stays the same. For this reason, some have argued that the risk difference is a better measure of causal impact than the relative risk. Let's go through an example to illustrate this.

So in example one, the black color P1 and P4 are the numbers that we're changing. And then we're holding P2 and P3 constant. So P2 and P3, in red, are the tapes that are amenable to intervention. In the first example, P1 is high, and P4 is low. And in the second example, we reverse those numbers. When we fill in our formula here, we get a relative risk of 1.24 and a risk difference of 0.18 in example one. In example two, the risk difference is the same, which is not surprising. The relative risk is 4.

This discrepancy in results highlights how a high proportion of people in the population who are doomed can cause the relative risk to be closer to the null value of 1. But the risk difference stays the same regardless of the proportion of people in the population who are doomed. This is a nice feature of the risk difference. And that's really the key reason why, in the modern perspective, many people feel that more and more studies should be reporting the risk difference in addition to the relative risk instead of just the relative risk alone when the goal is to make causal inference.

Summary of key points

- In the traditional perspective: relative scale measures provide information about etiology and causation and absolute scale measures provide information about public health impact
- Some have argued that the RD is a better measure of a causal effect than the RR because the RR depends on the proportion of people who are doomed to get disease (and cannot be affected by intervention), whereas the RD isolates the impact on those who are amenable to intervention.



To summarize, in the traditional perspective, epidemiologists learned that the relative scale measures, such as the relative risk, provide information about etiology and causation. Whereas, absolute scale measures provide information about public health impact.

These days, many epidemiologists argue that the risk difference is a better measure of a causal effect than the relative risk because the relative risk depends on the proportion of people in the population who are doomed to get the disease and can't be affected by any public health intervention. Whereas, the risk difference isolates the impact on those who are amenable to intervention.

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,

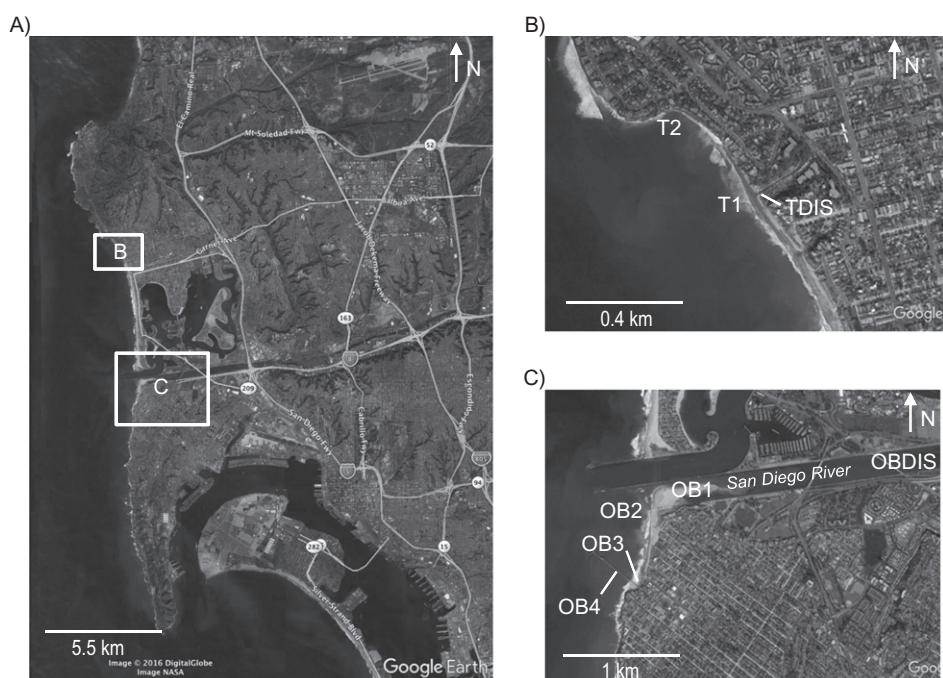


Figure 1. Monitoring beach water quality sampling locations in San Diego, California, winters of 2013–2014 and 2014–2015. Shown are the locations of the 2 monitored beaches along the San Diego coastline (A) and the water quality sampling sites at Tourmaline Surfing Park (B) and Ocean Beach (C). Samples 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-

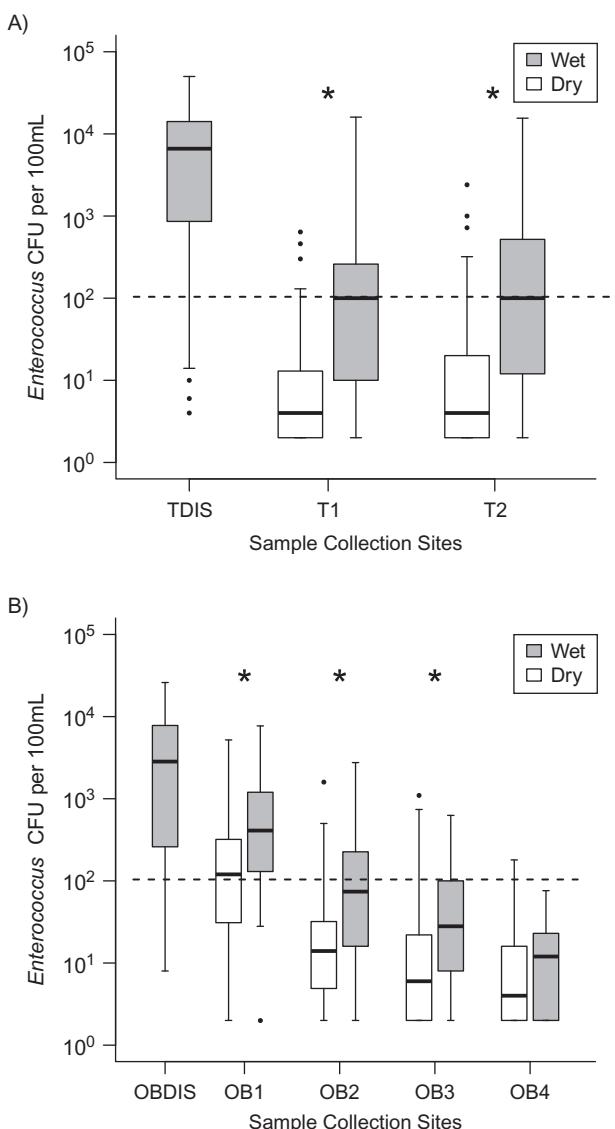


Figure 2. *Enterococcus* levels during dry and wet weather at the sampling locations at Tourmaline Surfing Park (A) and Ocean Beach (B) mapped in Figure 1. Boxes mark interquartile ranges, vertical lines mark 1.5 times the interquartile range, and points mark outliers. Horizontal dashed lines mark the 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 Surf, San Diego, California, 2013–2015

| Fecal Indicator Bacteria and Illness Symptom | Unadjusted | | | | | | | | | | | | Adjusted ^a | |
|--|-------------|--------------|-------------|--------------|-------------|-------------|-------------|-------------|----------------------|-------------|-------------|------|-----------------------|--|
| | Dry Weather | | Wet Weather | | Dry Weather | | Wet Weather | | P Value ^b | Dry Weather | | | | |
| | Episodes | Days at Risk | Episodes | Days at Risk | IRR | 95% CI | IRR | 95% CI | | IRR | 95% CI | IRR | | |
| 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 | 1.16 | 1.16 | |
| 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 | 2.05 | 2.05 | |
| 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 | 1.16 | 1.16 | |
| 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 | 1.16 | 1.16 | |
| 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 | 2.05 | 2.05 | |
| 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 | 0.99 | 0.99 | |
| 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 | 3.00 | 3.00 | |
| 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 | 1.16 | 1.16 | |
| 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 | 2.05 | 2.05 | |
| Fecal 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 | 2.05 | 2.05 | |
| 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 | 3.00 | 3.00 | |
| 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 | 1.16 | 1.16 | |
| 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 | 1.16 | 1.16 | |
| 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 | 4.00 | 4.00 | |
| 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 | 0.99 | 0.99 | |
| 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 | 6.00 | 6.00 | |
| 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 | 2.05 | 2.05 | |
| 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 | 3.00 | 3.00 | |
| 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 | 1.16 | 1.16 | |
| 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 | 1.16 | 1.16 | |
| 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 | 1.16 | 1.16 | |
| 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 | 1.16 | 1.16 | |
| 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 | 2.05 | 2.05 | |
| 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 | 1.16 | 1.16 | |
| 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 | 9.00 | 9.00 | |
| Upper respiratory illness ^a | 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 | 1.16 | 1.16 | |
| 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 | 3.00 | 3.00 | |

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 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 infections, 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 symptom 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.

The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812

AUGUST 7, 2003

VOL. 349 NO. 6

Estrogen plus Progestin and the Risk of Coronary Heart Disease

JoAnn E. Manson, M.D., Dr.P.H., Judith Hsia, M.D., Karen C. Johnson, M.D., M.P.H., Jacques E. Rossouw, M.D., Ann Louise R. Assaf, Ph.D., Norman L. Lasser, M.D., Ph.D., Maurizio Trevisan, M.D., Henry R. Black, M.D., Susan R. Heckbert, M.D., Ph.D., Robert Detrano, M.D., Ph.D., Ora L. Strickland, Ph.D., Nathan D. Wong, Ph.D., John R. Crouse, M.D., Evan Stein, M.D., and Mary Cushman, M.D., for the Women's Health Initiative Investigators*

ABSTRACT

BACKGROUND

Recent randomized clinical trials have suggested that estrogen plus progestin does not confer cardiac protection and may increase the risk of coronary heart disease (CHD). In this report, we provide the final results with regard to estrogen plus progestin and CHD from the Women's Health Initiative (WHI).

METHODS

The WHI included a randomized primary-prevention trial of estrogen plus progestin in 16,608 postmenopausal women who were 50 to 79 years of age at base line. Participants were randomly assigned to receive conjugated equine estrogens (0.625 mg per day) plus medroxyprogesterone acetate (2.5 mg per day) or placebo. The primary efficacy outcome of the trial was CHD (nonfatal myocardial infarction or death due to CHD).

RESULTS

After a mean follow-up of 5.2 years (planned duration, 8.5 years), the data and safety monitoring board recommended terminating the estrogen-plus-progestin trial because the overall risks exceeded the benefits. Combined hormone therapy was associated with a hazard ratio for CHD of 1.24 (nominal 95 percent confidence interval, 1.00 to 1.54; 95 percent confidence interval after adjustment for sequential monitoring, 0.97 to 1.60). The elevation in risk was most apparent at one year (hazard ratio, 1.81 [95 percent confidence interval, 1.09 to 3.01]). Although higher base-line levels of low-density lipoprotein cholesterol were associated with an excess risk of CHD among women who received hormone therapy, higher base-line levels of C-reactive protein, other biomarkers, and other clinical characteristics did not significantly modify the treatment-related risk of CHD.

CONCLUSIONS

Estrogen plus progestin does not confer cardiac protection and may increase the risk of CHD among generally healthy postmenopausal women, especially during the first year after the initiation of hormone use. This treatment should not be prescribed for the prevention of cardiovascular disease.

From the Division of Preventive Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston (J.E.M.); the Department of Medicine, George Washington University, Washington, D.C. (J.H.); the Department of Preventive Medicine, University of Tennessee Health Science Center, Memphis (K.C.J.); the Program Office, National Heart, Lung, and Blood Institute, Bethesda, Md. (J.E.R.); Memorial Hospital, Brown Medical School, Pawtucket, R.I. (A.R.A.); the Preventive Cardiology Program, New Jersey Medical School, Newark (N.L.L.); the Department of Social and Preventive Medicine, University at Buffalo, Buffalo, N.Y. (M.T.); the Department of Preventive Medicine, Rush-Presbyterian-St. Luke's Medical Center, Chicago (H.R.B.); the Department of Epidemiology, University of Washington, Seattle (S.R.H.); the Division of Cardiology, Harbor-UCLA Research and Education Institute, Torrance, Calif. (R.D.); the Woodruff School of Nursing, Emory University, Atlanta (O.L.S.); the Heart Disease Prevention Program, University of California, Irvine (N.D.W.); the Department of Medicine, Wake Forest University, Winston-Salem, N.C. (J.R.C.); Medical Research Laboratories International, Highland Heights, Ky. (E.S.); and the Departments of Medicine and Pathology, University of Vermont, Burlington (M.C.). Address reprint requests to Dr. Manson at the Division of Preventive Medicine, Brigham and Women's Hospital, 900 Commonwealth Ave., Boston, MA 02215, or at jmanson@rics.bwh.harvard.edu.

*The Women's Health Initiative (WHI) investigators are listed in the Appendix.

N Engl J Med 2003;349:523-34.

Copyright © 2003 Massachusetts Medical Society.

OUR UNDERSTANDING OF THE EFFECT of postmenopausal hormone therapy on the risk of coronary heart disease (CHD) has recently undergone a major change. Although previous observational studies had suggested that postmenopausal hormone therapy was associated with a reduction of 40 to 50 percent in the risk of CHD,^{1,2} recent randomized clinical trials have provided no evidence of cardiac protection and even some evidence of harm with postmenopausal hormone therapy.³⁻⁸ The primary findings of the Estrogen plus Progestin trial of the Women's Health Initiative (WHI) suggested an overall increase in the risk of CHD (hazard ratio, 1.29) among women randomly assigned to combined hormone therapy as compared with those assigned to placebo.⁸ The trial was stopped early, after an average of 5.2 years of follow-up, because it was found that the health risks associated with estrogen plus progestin exceeded the benefits.

It has been hypothesized that divergent findings from observational studies and randomized clinical trials may be at least partially attributable to differences in the clinical characteristics of the study populations, including differences in age, years since menopause, and underlying risk of CHD, as well as methodologic limitations of observational studies.^{9,10} Moreover, certain biomarkers, including base-line levels of lipoproteins, inflammatory markers, and thrombotic factors, may identify women for whom postmenopausal hormone therapy confers a higher or lower risk of coronary events.¹¹⁻¹⁴

In this article, we present the final results of the WHI trial of the relation between the use of estrogen plus progestin and the risk of CHD. We provide an updated analysis of coronary end points reached through the termination of the trial on July 7, 2002 (previous analyses included end points reached through April 2002). We use centrally adjudicated end points for the primary coronary outcome of nonfatal myocardial infarction or death due to CHD (previous analyses were based on local adjudication) to enhance the uniformity of documentation of outcomes. We also provide results for additional coronary end points, including angina, acute coronary syndromes, and congestive heart failure, and provide detailed analyses of subgroups of women defined according to clinical characteristics and biomarker levels to further elucidate the primary findings.

METHODS

STUDY POPULATION, RECRUITMENT, STUDY REGIMENS, AND FOLLOW-UP

Detailed information about the study population, recruitment methods, study regimens, randomization, blinding, follow-up, data and safety monitoring, and quality assurance has been published previously.^{8,15} Briefly, eligible women were 50 to 79 years of age at the time of initial screening, were postmenopausal, and were likely to be residing in the same geographic area for at least three years.

Postmenopausal women with an intact uterus at screening were eligible for the trial of combined estrogen and progestin; women who had undergone hysterectomy were eligible for the trial of estrogen alone. The protocol and consent forms were approved by the institutional review boards of the participating institutions, and written informed consent was obtained from all participants. The sample analyzed here consists of the 16,608 women with an intact uterus at base line who were enrolled in the double-blind trial comparing estrogen plus progestin with placebo. The study regimen of combined estrogen and progestin was provided in one daily tablet containing 0.625 mg of oral conjugated equine estrogen and 2.5 mg of medroxyprogesterone acetate (Prempro, Wyeth). The control group received matching placebo.

ASCERTAINMENT OF OUTCOMES

CHD was defined as acute myocardial infarction necessitating overnight hospitalization, death due to CHD, or silent myocardial infarction identified on serial electrocardiography.¹⁶ The diagnosis of acute myocardial infarction was documented by a review of the medical records according to an algorithm that was adapted from standardized criteria,^{8,17} including cardiac pain, cardiac enzyme and troponin levels, and electrocardiographic readings. Death due to CHD was defined as death consistent with an underlying cause of CHD plus one or more of the following factors: hospitalization for myocardial infarction within 28 days before death, previous angina or myocardial infarction, death due to a procedure related to CHD, or a death certificate consistent with an underlying cause of CHD. Silent myocardial infarction¹⁶ was diagnosed through the comparison of base-line and follow-up electrocardiograms at three and six years. Additional coronary end points included coronary revascularization (coronary-artery bypass grafting [CABG] or

percutaneous transluminal coronary angioplasty [PTCA]) confirmed by a review of the medical records, angina necessitating hospitalization (hospital admission for chest pain or other symptoms determined to be due to angina), confirmed angina (hospitalization for angina, with myocardial ischemia confirmed by stress testing or obstructive coronary disease [luminal narrowing of >70 percent] confirmed by coronary angiography), acute coronary syndromes (hospitalization for angina, Q-wave infarction, or non-Q-wave infarction), and congestive heart failure (necessitating hospitalization, with a physician's diagnosis of congestive heart failure and pertinent abnormalities on diagnostic testing corroborated by a review of the medical records). Acute myocardial infarctions and deaths due to CHD were confirmed by central physician-adjudicators and other coronary end points by local adjudicators, all of whom were unaware of the treatment-group assignments. The rate of concordance between the local and central reviews was 90 percent for myocardial infarction and 97 percent for death due to atherosclerotic CHD.

ANALYSES OF BIOMARKERS

Blood was drawn at base line after a fast lasting a minimum of 10 hours. Serum and plasma samples were shipped to a central repository and stored at -70°C.¹⁸ In a random sample of 8.6 percent of participants (oversampled for women from minority groups), the lipid profile was obtained and glucose and insulin were measured at base line, year 1, and year 3. The assay methods have been described previously.¹⁸

A nested case-control study of biomarkers, treatment-group assignment, and risk of CHD was also conducted. A total of 205 cases of myocardial infarction or death due to CHD occurring between randomization and February 28, 2001, were included. Controls were selected from the hormone-therapy trial and were matched to the cases according to age, date of randomization, presence or absence of CHD at base line, hysterectomy status, and follow-up time. Additional controls selected for cases of stroke or venous thrombosis were also included; the total number of controls was 513. Methods of testing for the inflammatory and thrombotic markers have been described previously.¹⁸ Data analysis was performed with the use of logistic regression.

STATISTICAL ANALYSIS

Primary analyses used time-to-event methods based on the intention-to-treat principle. For coronary

outcomes, the time to the event was defined as the number of days between randomization and the first diagnosis after randomization. Comparisons with regard to the primary outcome are presented as hazard ratios with 95 percent confidence intervals that were calculated from Cox proportional-hazards analyses,¹⁹ stratified according to age, presence or absence of CHD at base line, and randomization status in the low-fat-diet trial (as in the original report⁸), and adjusted for the presence or absence of previous CABG or PTCA. Because CHD was the primary outcome of the hormone trial and was an important consideration for stopping the trial early⁸ (the trial was terminated after the 10th semiannual interim analysis), both nominal 95 percent confidence intervals and 95 percent confidence intervals adjusted for sequential monitoring are provided for the primary coronary end point. For other coronary end points, both nominal confidence intervals and confidence intervals adjusted for multiple (seven) trial outcomes are presented. Secondary analyses included women who adhered fully to the study medication.

Cox models for subgroup analyses were stratified according to age and the presence or absence of CHD at base line, and the consistency of treatment effects among subgroups was assessed by formal tests of interaction. Because of the large number of subgroups considered (at least 36), the results should be interpreted with caution, since some significant findings (at least one or two, based on a 0.05 nominal level of statistical significance) could have occurred by chance alone. All reported P values are two-sided.

RESULTS

BASE-LINE CHARACTERISTICS

As described in the original report,⁸ the base-line characteristics were nearly identical in the two treatment groups. The only base-line variable that differed significantly between the groups was a history of coronary revascularization (present in 1.1 percent of the women in the hormone group and 1.5 percent of those in the placebo group, P=0.04), so this variable was included as a covariate in the Cox models. A total of 8506 women were randomly assigned to estrogen plus progestin, and 8102 were assigned to placebo. The mean (\pm SD) age was 63.3 \pm 7.1 years; 16 percent of the women were members of minority groups; and one quarter of the women had previously used postmenopausal hormone therapy. Approximately 2.4 percent of the women reported

previous CHD (myocardial infarction, a coronary revascularization procedure, or both) and 4.4 percent reported previous CHD, stroke, or transient cerebral ischemia. Thus, the prevalence of previous cardiovascular disease was low, and women with such a history were analyzed separately in secondary analyses. The base-line levels of cardiovascular risk factors (36 percent of the women had hypertension, 13 percent were being treated for hypercholesterolemia, 4.4 percent were being treated for diabetes, and 10.5 percent were current smokers) were consistent with those in a generally healthy population of postmenopausal women.

through July 7, 2002 (after an average of 5.6 years of follow-up [as compared with 5.2 years in the earlier report⁸] and a maximum of 8.6 years). As previously reported,⁸ 42 percent of women randomly assigned to estrogen plus progestin and 38 percent of women randomly assigned to placebo stopped taking the study drugs during follow-up — rates that compare favorably with community-based adherence to hormone therapy.²⁰ The cumulative “drop-in” rate — the rate of hormone use initiated by the woman’s clinician — was 6.2 percent in the estrogen-plus-progestin group and 10.7 percent in the placebo group by year 6.

FOLLOW-UP AND ADHERENCE

Vital status was known for 16,067 women who underwent randomization (96.7 percent), including 485 (2.9 percent) who were known to be deceased. Information on outcomes was up to date for 15,582 women (93.8 percent); for the 541 women (3.3 percent) who were lost to follow-up or who stopped providing information on outcomes before the trial ended, we include all available information. The present report updates information on outcomes

INTERMEDIATE BIOMARKERS AND RISK FACTORS FOR CHD

The results of assessments of CHD biomarkers, including fasting blood lipid, glucose, and insulin levels, in an 8.6 percent subsample of women at base line and at year 1 are shown in Figure 1. Women randomly assigned to estrogen plus progestin had greater reductions in the total cholesterol, low-density lipoprotein (LDL) cholesterol, glucose, and insulin levels and greater increases in the high-density

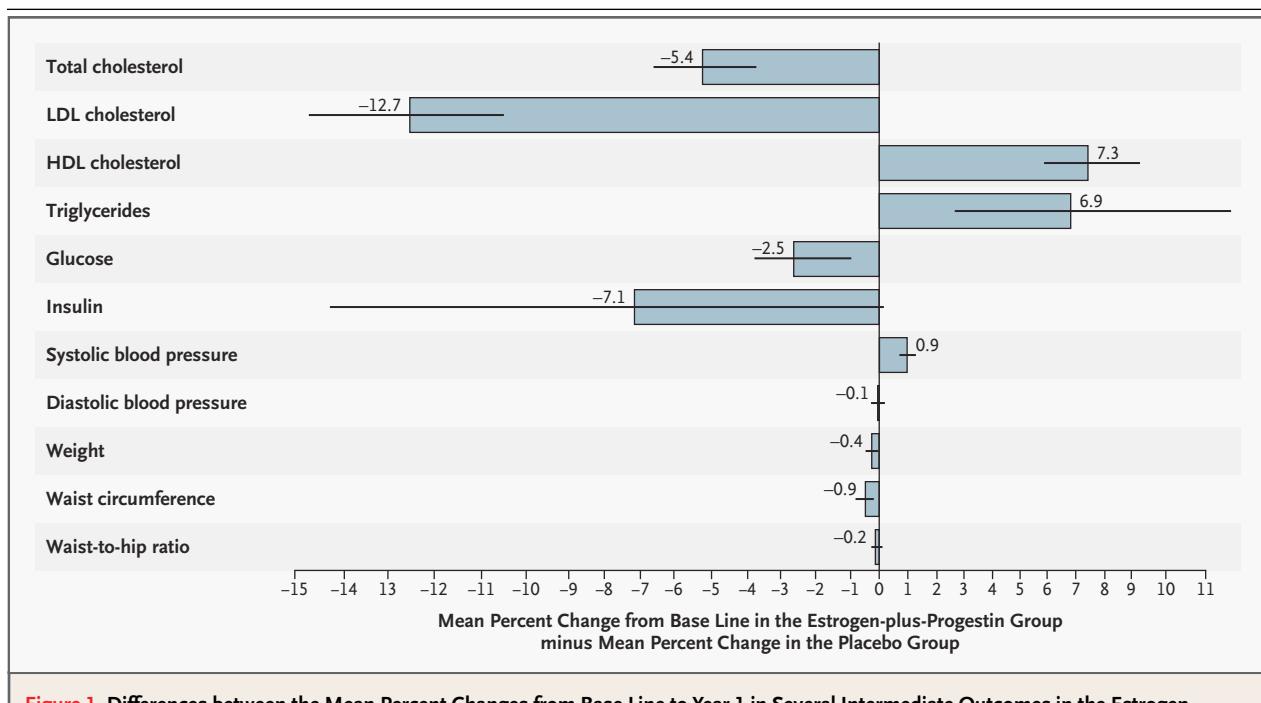


Figure 1. Differences between the Mean Percent Changes from Base Line to Year 1 in Several Intermediate Outcomes in the Estrogen-plus-Progestin Group as Compared with the Placebo Group.

Horizontal lines represent the 95 percent confidence intervals. The differences between the groups were significant ($P < 0.05$) for total cholesterol, low-density lipoprotein (LDL) cholesterol, high-density lipoprotein (HDL) cholesterol, triglycerides, glucose, systolic blood pressure, weight, and waist circumference.

lipoprotein (HDL) cholesterol and triglyceride levels than women in the placebo group. Systolic blood pressure at year 1 was 1 mm Hg higher among women receiving hormones than among those receiving placebo (remaining 1 to 2 mm Hg higher during follow-up), although diastolic blood pressure did not differ materially between groups. Body weight and waist circumference at follow-up were slightly lower among women in the hormone group than among those in the placebo group, although the ratio of the waist circumference to the hip circumference did not differ appreciably (Fig. 1). Results at year 3 (data not shown) were nearly identical to those at year 1.

CLINICAL CORONARY OUTCOMES

Table 1 shows the rates of CHD (nonfatal myocardial infarction, including silent myocardial infarction, and death due to CHD), coronary revasculariza-

tion, angina, and congestive heart failure. In adjusted analyses, women randomly assigned to estrogen plus progestin had a risk of CHD that was 24 percent higher than that among women randomly assigned to placebo (hazard ratio, 1.24 [nominal 95 percent confidence interval, 1.00 to 1.54; 95 percent confidence interval with adjustment for sequential monitoring, 0.97 to 1.60]). The hazard ratios were 1.28 for nonfatal myocardial infarction and 1.10 for death due to CHD (total cases of CHD, 335, as compared with 286 in the earlier report⁸). Absolute rates of CHD were 39 cases per 10,000 person-years and 33 cases per 10,000 person-years for hormone therapy and placebo, respectively. No significant differences were observed with regard to coronary revascularization, hospitalization for angina, confirmed angina, acute coronary syndrome, or congestive heart failure.

Additional analyses were conducted to examine

Table 1. Coronary Outcomes among Women Randomly Assigned to Estrogen plus Progestin, as Compared with Those Assigned to Placebo.*

| Variable | Estrogen-plus-Progestin Group (N=8506) | Placebo Group (N=8102) | Adjusted Hazard Ratio | Nominal 95% CI | Adjusted 95% CI |
|--------------------------------------|---|---------------------------|-----------------------|----------------|-----------------|
| Mean follow-up time (mo) | 67.8 | 66.8 | | | |
| no. of cases (annualized percentage) | | | | | |
| CHD | 188 (0.39) | 147 (0.33) | 1.24 | 1.00–1.54 | 0.97–1.60 |
| Nonfatal MI | | | | | |
| Including silent MI | 151 (0.31) | 114 (0.25) | 1.28 | 1.00–1.63 | 0.96–1.70 |
| Excluding silent MI | 147 (0.31) | 109 (0.24) | 1.30 | 1.01–1.67 | 0.97–1.74 |
| Death due to CHD | 39 (0.08) | 34 (0.08) | 1.10 | 0.70–1.75 | 0.65–1.89 |
| CHD, revascularization, or angina | 369 (0.77) | 356 (0.79) | 1.00 | 0.86–1.15 | 0.82–1.22 |
| CABG or PTCA | 214 (0.45) | 205 (0.45) | 1.01 | 0.83–1.22 | 0.77–1.31 |
| Hospitalization for angina | 172 (0.36) | 195 (0.43) | 0.86 | 0.70–1.05 | 0.65–1.13 |
| Confirmed angina | 106 (0.22) | 126 (0.28) | 0.82 | 0.63–1.06 | 0.57–1.17 |
| Acute coronary syndrome | 322 (0.67) | 299 (0.66) | 1.03 | 0.88–1.21 | 0.83–1.28 |
| Congestive heart failure | 113 (0.23) | 109 (0.24) | 0.99 | 0.76–1.29 | 0.69–1.42 |

* CHD includes acute myocardial infarction (MI) necessitating hospitalization, silent myocardial infarction as determined by serial electrocardiography, and death due to CHD. Hazard ratios and nominal 95 percent confidence intervals (CIs) are stratified according to age, presence or absence of a previous coronary event, and randomly assigned diet-modification group and are adjusted for the presence or absence of previous coronary-artery bypass grafting (CABG) or percutaneous transluminal coronary angioplasty (PTCA). The adjusted 95 percent confidence interval includes control for the above variables and further control for sequential monitoring (for the primary coronary end points) and for multiple (seven) trial outcomes (for the secondary coronary end points). Confirmed angina includes hospitalization for angina with myocardial ischemia confirmed by stress testing or obstructive coronary disease (luminal narrowing of >70 percent) confirmed by coronary angiography. Acute coronary syndromes include Q-wave myocardial infarction, non-Q-wave myocardial infarction, and hospitalization for angina. The numbers of events do not add up to the totals for the categories because some women had more than one event.

the sensitivity of these results to the actual use of study medications. Because a substantial proportion of women stopped taking study pills during follow-up, analyses were performed that censored the data on a woman's history of coronary events six months after she stopped taking the pills (or began taking less than 80 percent of them) or six months after she began nonstudy hormone therapy. These analyses produced higher estimates of the excess risk with estrogen plus progestin. For CHD, the adjusted hazard ratio was 1.50 (95 percent confidence interval, 1.14 to 1.97), and for CHD, revascularization, or angina, the hazard ratio was 1.09 (95 percent confidence interval, 0.90 to 1.31). If discontinuation of treatment and initiation of nonstudy hormone therapy occurred independently of the risk of CHD, it would suggest that the intention-to-treat analyses may underestimate the effect. Such "adherence-based" analyses, however, have limitations and should be interpreted with caution.

TEMPORAL TRENDS

The cumulative hazard rates of CHD (nonfatal myocardial infarction or death due to CHD) in the two treatment groups are provided in Figure 2. An elevated risk of CHD with estrogen plus progestin appeared to emerge soon after randomization, and the cumulative rates did not begin to converge until year 6.

Hazard ratios for CHD for one-year intervals of follow-up are presented in Table 2. A substantial elevation in the risk of CHD with estrogen plus progestin occurred in year 1 (hazard ratio, 1.81 [95 percent confidence interval, 1.09 to 3.01]), and a smaller and nonsignificant excess risk occurred in years 2 through 5. In year 6 and beyond, the increased rates in the placebo group resulted in an apparent risk reduction. The trend toward a decreasing relative risk over time was statistically significant. For CHD, revascularization, or angina, the hazard ratio was 1.48 (95 percent confidence interval, 1.03 to 2.11) at one year, but no elevation in the risk was apparent in subsequent years.

SUBGROUP ANALYSES

To determine whether certain subgroups of women were at particularly high or low risk for CHD (non-fatal myocardial infarction or death due to CHD) with estrogen plus progestin, we examined several demographic and clinical characteristics. In addition, base-line levels of several lipid, inflammatory, and thrombotic biomarkers were assessed as potential modulators of risk. Overall, no subgroup of women except those with higher base-line LDL cholesterol levels had evidence of a pattern of hazard ratios for CHD with postmenopausal hormone therapy that was different from the pattern found among all women. Subgroup analyses were planned a priori; the results of analyses of variables whose influence has greater biologic plausibility are shown in Figures 3 and 4, and the remainder are summarized in Table 3 or below.

Results of evaluations of the roles of age and the time since menopause in modulating the risk of treatment are shown in Figure 3. No significant interaction between age and treatment was observed. For women in whom menopause had begun less than 10 years previously, 10 to 19 years previously, and 20 or more years previously, the hazard ratios for CHD associated with postmenopausal hormone therapy were 0.89, 1.22, and 1.71, respectively, but the interaction was nonsignificant. Moreover, the

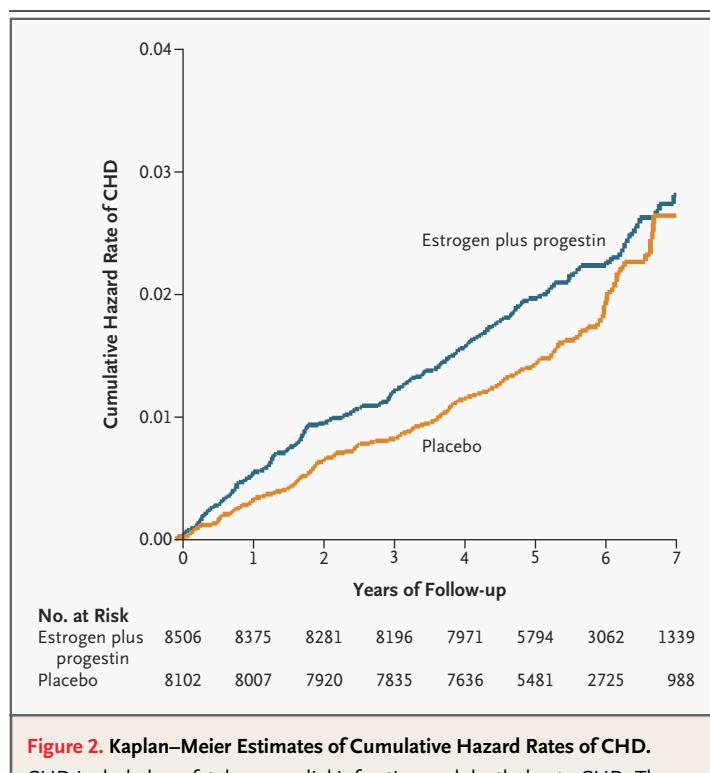


Figure 2. Kaplan-Meier Estimates of Cumulative Hazard Rates of CHD.

CHD included nonfatal myocardial infarction and death due to CHD. The overall hazard ratio for CHD was 1.24 (nominal 95 percent confidence interval, 1.00 to 1.54; 95 percent confidence interval with adjustment for sequential monitoring, 0.97 to 1.60).

presence or absence of vasomotor symptoms (hot flashes, night sweats, or both) was not significantly related to the risk of CHD associated with postmenopausal hormone therapy, either among women 50 to 59 years of age (Fig. 3) or in the total cohort (hazard ratios, 1.26 and 1.25, respectively). Previous use of hormone therapy did not appreciably or consistently modify risk, regardless of the duration or temporal proximity of this use. Body-mass index (the weight in kilograms divided by the square of the height in meters) and other anthropometric measures (the waist circumference and the waist-to-hip ratio) did not clearly modulate the risk associated with postmenopausal hormone therapy, nor did the use of aspirin (≥ 80 mg per day) or statin therapy (Fig. 3).

The hazard ratios for CHD with estrogen plus progestin did not differ substantially according to ethnic group, level of education, or CHD-risk-factor status (Table 3) or according to the past use or nonuse of oral contraceptives or levels of physical activity (data not shown). Women who were current smokers or who had a history of hypertension or diabetes, a higher number of risk factors for CHD, or preexisting CHD or other cardiovascular disease did not have a significantly greater excess risk of subsequent coronary events with postmenopausal hormone therapy than did women without these risk factors (Table 3).

Women with higher base-line LDL cholesterol levels appeared to have a greater excess risk of CHD with hormone therapy (P for interaction=0.01, after adjustment for age, year of randomization, previous CHD, and use of statins at base line) (Fig. 4), but this finding may have been due to chance, given the large number of comparisons tested. No other subgroup defined according to biomarker levels, including the C-reactive protein level, had a risk of CHD with postmenopausal hormone therapy that differed significantly from the risk among all women (Fig. 4).

DISCUSSION

Our findings in predominantly healthy postmenopausal women 50 to 79 years of age document that combined estrogen and progestin does not confer cardiac protection and may slightly increase the risk of coronary events. These findings extend the information that has been published previously⁸ by including updated and centrally adjudicated primary

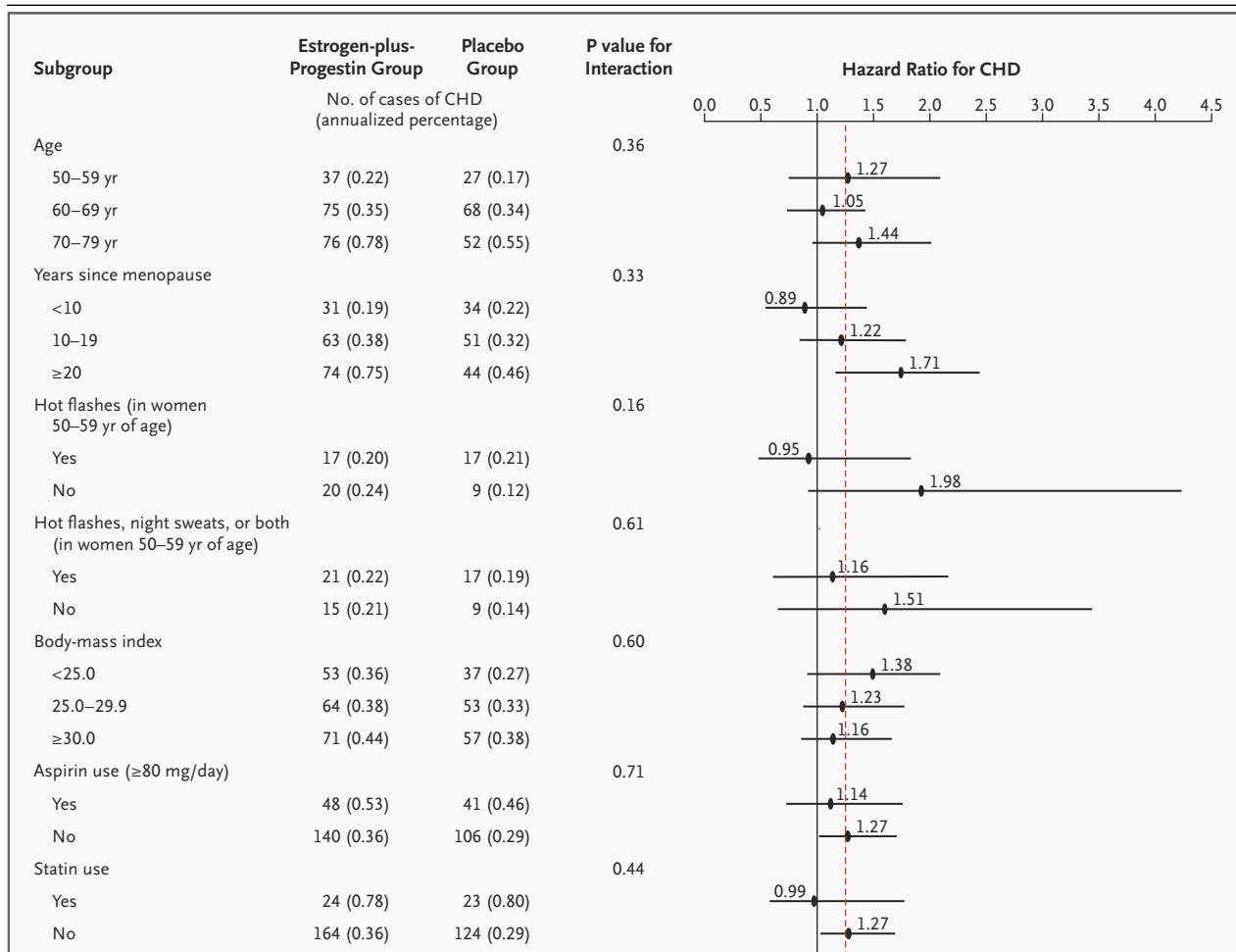
coronary end points, providing results for additional coronary outcomes, and examining risk in subgroups of women. The apparent slight increase in risk occurred predominantly for myocardial infarction, with no material increase in the risk of coronary revascularization, angina, or congestive heart failure.

Although the trend toward a decreasing risk of CHD over time with estrogen plus progestin was statistically significant, these results must be interpreted with caution. Hazard ratios for CHD were above 1.0 through year 5 among women assigned to postmenopausal hormone therapy, with particularly elevated rates in year 1. Results in subsequent years were limited by smaller numbers and lower rates of adherence to study medication and were confined to women who were still at risk for a first coronary event. Thus, results in later years could be artificially lowered by an acceleration of events in earlier years among susceptible women assigned to postmenopausal hormone therapy. In addition, an increase in the rates of events in the placebo

Table 2. Estrogen plus Progestin and the Risk of CHD, According to Year of Follow-up.*

| Year of Follow-up | CHD | | Hazard Ratio for CHD (95% CI) |
|--------------------------------------|-------------------------------|---------------|-------------------------------|
| | Estrogen-plus-Progestin Group | Placebo Group | |
| no. of cases (annualized percentage) | | | |
| 1 | 42 (0.50) | 23 (0.29) | 1.81 (1.09–3.01) |
| 2 | 38 (0.45) | 28 (0.35) | 1.34 (0.82–2.18) |
| 3 | 19 (0.23) | 15 (0.19) | 1.27 (0.64–2.50) |
| 4 | 32 (0.39) | 25 (0.32) | 1.25 (0.74–2.12) |
| 5 | 29 (0.41) | 19 (0.28) | 1.45 (0.81–2.59) |
| ≥6 | 28 (0.37) | 37 (0.56) | 0.70 (0.42–1.14) |

* CHD includes acute myocardial infarction (MI) necessitating hospitalization, silent myocardial infarction as determined by serial electrocardiography, and death due to CHD. There were nine silent myocardial infarctions (four in the estrogen-plus-progestin group and five in the placebo group). Hazard ratios are stratified according to age, presence or absence of a previous coronary event, and randomly assigned diet-modification group and are adjusted for previous coronary-artery bypass grafting or percutaneous transluminal coronary angioplasty. The z score for trend was -2.36 ($P=0.02$); the test for trend was based on Cox proportional-hazards models with time-dependent treatment effects. The 95 percent confidence intervals (CIs) are nominal.

**Figure 3.** Estrogen plus Progestin and the Risk of CHD in Various Subgroups.

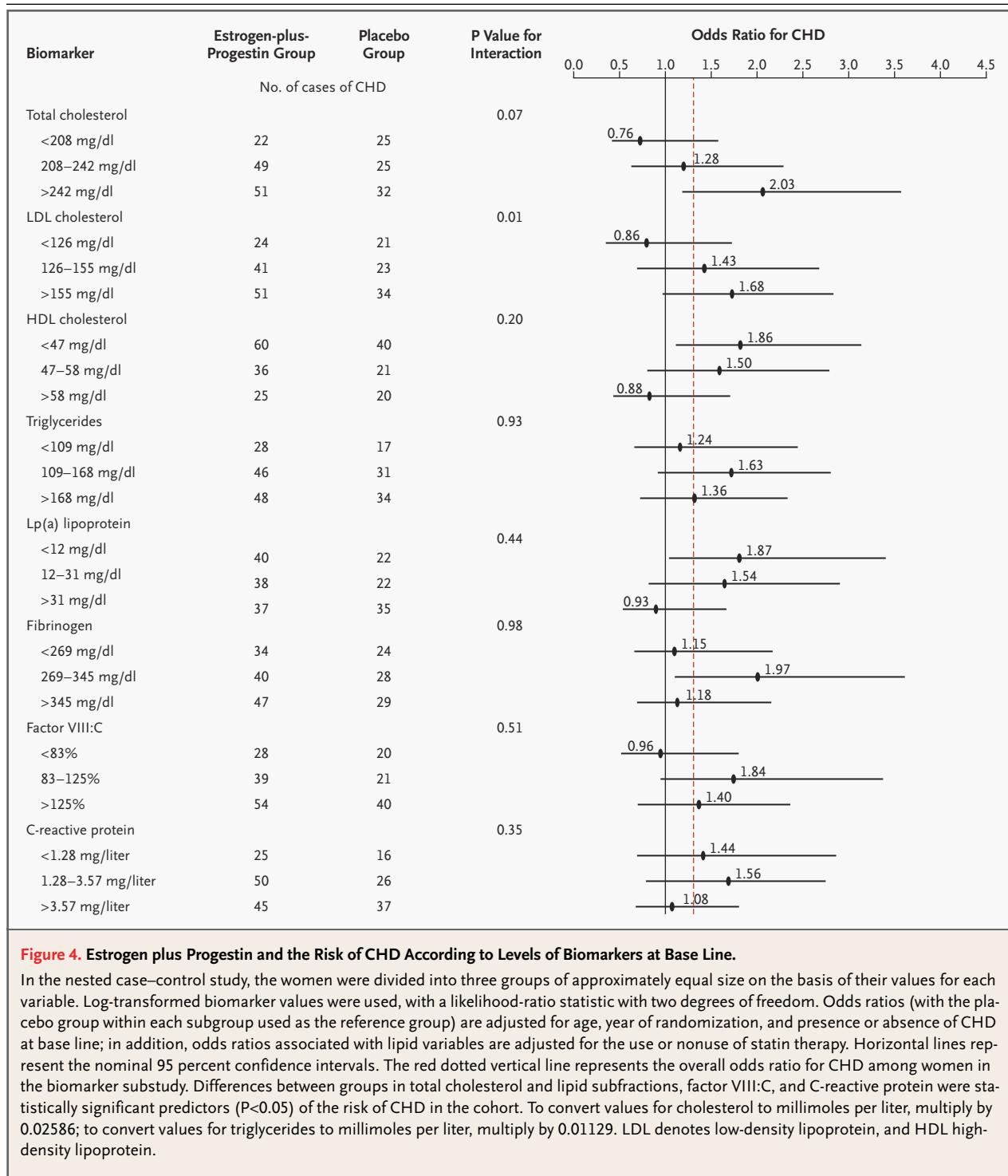
CHD includes nonfatal myocardial infarction and death due to CHD. Hazard ratios are adjusted for age (except for those associated with age and years since menopause) and the presence or absence of CHD at baseline. Horizontal bars represent nominal 95 percent confidence intervals. The red dotted vertical line represents the hazard ratio for CHD in the overall cohort. Because of missing data on some variables, the numbers of cases do not always add up to the total number of cases in the treatment group.

group contributed to the apparently lower hazard ratio in year 6 and beyond. Moreover, the increased risk of breast cancer with a longer duration of treatment⁸ and the adverse overall benefit-to-risk profile would outweigh any coronary benefit that might be seen with longer follow-up.

No subgroup of women except those with higher base-line LDL cholesterol levels had evidence of a risk of CHD with estrogen plus progestin that differed significantly from that observed for all women, and the findings related to LDL cholesterol may have been due to chance. Age, time since menopause, body-mass index, presence or absence of

vasomotor symptoms at baseline, coronary-risk-factor status, and other variables were not significantly related to the risk of CHD with hormone therapy. Base-line levels of C-reactive protein, fibrinogen, and other biomarkers also did not appear to modulate the risk. None of these variables should be used at this time for risk stratification or for the identification of women who may be more or less vulnerable to an adverse coronary outcome when given hormone therapy.

The absence of the provision of cardiac protection by estrogen plus progestin in our study is consistent with recent findings from randomized tri-

**Figure 4.** Estrogen plus Progestin and the Risk of CHD According to Levels of Biomarkers at Base Line.

In the nested case-control study, the women were divided into three groups of approximately equal size on the basis of their values for each variable. Log-transformed biomarker values were used, with a likelihood-ratio statistic with two degrees of freedom. Odds ratios (with the placebo group within each subgroup used as the reference group) are adjusted for age, year of randomization, and presence or absence of CHD at base line; in addition, odds ratios associated with lipid variables are adjusted for the use or nonuse of statin therapy. Horizontal lines represent the nominal 95 percent confidence intervals. The red dotted vertical line represents the overall odds ratio for CHD among women in the biomarker substudy. Differences between groups in total cholesterol and lipid subfractions, factor VIII:C, and C-reactive protein were statistically significant predictors ($P<0.05$) of the risk of CHD in the cohort. To convert values for cholesterol to millimoles per liter, multiply by 0.02586; to convert values for triglycerides to millimoles per liter, multiply by 0.01129. LDL denotes low-density lipoprotein, and HDL high-density lipoprotein.

Table 3. Estrogen plus Progestin and the Risk of CHD in Various Subgroups.*

| Variable | CHD | | Adjusted Hazard Ratio (95% CI) | P Value for Interaction |
|---|--|------------------|-----------------------------------|-------------------------|
| | Estrogen- plus- Progestin Group | Placebo Group | | |
| no. of cases (annualized percentage) | | | | |
| Race or ethnic group | | | | 0.41 |
| Non-Hispanic white | 165 (0.41) | 124 (0.33) | 1.28 (1.02–1.62) | |
| Non-Hispanic black | 13 (0.42) | 10 (0.32) | 1.22 (0.53–2.81) | |
| Hispanic | 6 (0.24) | 4 (0.18) | 1.33 (0.37–4.75) | |
| Level of education | | | | 0.86 |
| ≤High school or GED | 63 (0.51) | 50 (0.42) | 1.26 (0.87–1.83) | |
| >High school | 124 (0.35) | 96 (0.29) | 1.22 (0.94–1.60) | |
| Cigarette smoking | | | | 0.64 |
| Never smoked or former smoker | 153 (0.36) | 116 (0.29) | 1.27 (1.00–1.62) | |
| Current smoker | 31 (0.63) | 25 (0.53) | 1.10 (0.64–1.87) | |
| Hypertension | | | | 0.49 |
| No | 81 (0.26) | 66 (0.23) | 1.14 (0.82–1.58) | |
| Yes | 107 (0.65) | 81 (0.50) | 1.32 (0.99–1.76) | |
| Diabetes | | | | 0.51 |
| No | 155 (0.34) | 123 (0.29) | 1.20 (0.94–1.52) | |
| Yes, medication-treated | 27 (1.33) | 22 (1.15) | 1.31 (0.73–2.34) | |
| Yes (all cases) | 32 (1.20) | 24 (0.96) | 1.45 (0.84–2.51) | |
| No. of risk factors for CHD | | | | 0.96 |
| None | 20 (0.15) | 17 (0.12) | 1.19 (0.62–2.28) | |
| 1–2 | 62 (0.43) | 38 (0.27) | 1.59 (1.06–2.37) | |
| ≥3 | 59 (1.24) | 52 (1.12) | 1.15 (0.79–1.68) | |
| Presence of cardiovascular disease at base line | | | | 0.64 |
| No | 156 (0.34) | 118 (0.28) | 1.23 (0.97–1.56) | |
| Yes | 29 (1.64) | 24 (1.19) | 1.45 (0.84–2.49) | |
| Presence of CHD at base line | | | | 0.66 |
| No | 163 (0.35) | 124 (0.29) | 1.23 (0.97–1.55) | |
| Yes | 22 (2.18) | 18 (1.65) | 1.44 (0.77–2.70) | |

* CHD includes nonfatal myocardial infarction and death due to CHD. Hazard ratios (with nominal 95 percent confidence intervals [CIs]) are adjusted for age and the presence of CHD at base line. P values are for the interaction between the subgroup variable and treatment. Hypertension was defined as treated hypertension or a measured blood pressure of 140/90 mm Hg or higher. Risk factors for CHD included current cigarette smoking, hypertension, diabetes, high cholesterol levels, and a parental history of myocardial infarction (at <55 years of age in the father or <65 years of age in the mother). The presence of cardiovascular disease at base line was defined as a history of myocardial infarction, coronary-artery bypass grafting (CABG), percutaneous transluminal coronary angioplasty (PTCA), stroke, or transient cerebral ischemia. The presence of CHD at base line was defined as a history of myocardial infarction, CABG, or PTCA. Because of missing data on some variables, the numbers of cases do not always add up to the total number of cases in the treatment group. GED denotes general equivalency diploma.

als of postmenopausal hormone therapy in women with CHD. In the Heart Estrogen/Progestin Replacement Study (HERS), estrogen plus progestin had no overall effect on the risk of recurrent coronary events after 4.1³ and 6.8²¹ years of follow-up, although the finding of an increased risk after the initiation of treatment was similar to findings in our study. In two angiographic trials,^{4,6} neither estrogen plus progestin nor estrogen alone was associated with inhibition of the progression of coronary atherosclerosis. The Papworth trial,⁵ which tested transdermal 17 β -estradiol with or without norethindrone, and a trial of estradiol valerate (without progestin) in women with a history of myocardial infarction²² also demonstrated no cardioprotection with postmenopausal hormone therapy. Moreover, the Women's Estrogen for Stroke Trial, which tested oral 17 β -estradiol (without progestin), found no overall effect of estrogen on the risk of recurrent stroke and an increase in the risk of fatal stroke.²³ Thus, although most of these trials tested the hormone regimen we studied (oral conjugated equine estrogen and medroxyprogesterone acetate), the trials testing transdermal or oral 17 β -estradiol or estradiol valerate had similar results.

Previous randomized trials have elucidated several favorable and unfavorable effects of exogenous hormone therapy on intermediate biomarkers. Estrogen therapy reduces plasma levels of LDL cholesterol and increases levels of HDL cholesterol, improves endothelial vascular function, and reduces the levels of fibrinogen, Lp(a) lipoprotein, plasminogen-activator inhibitor type 1, and insulin.^{11,12,24,25} However, estrogen also has adverse physiological effects, including increasing the plasma levels of triglycerides; small, dense LDL particles; C-reactive protein; and thrombotic markers such as factor VII, prothrombin fragment 1+2, and fibrinopeptide A.^{11,12,26,27} The addition of a progestin attenuates some of the lipid benefits of estrogen, particularly the increase in HDL cholesterol, but does not seem to counter the prothrombotic effects.^{12,24}

Whether or not certain clinical characteristics of the study population or base-line levels of selected biomarkers predict the coronary effects of postmenopausal hormone therapy is an important area of inquiry. Previous trials have identified few factors that modulate risk. In the HERS trial, despite extensive subgroup analyses, results were found to be generally similar regardless of age and coronary-risk-factor status.^{3,21,28} Hormone therapy appeared to have a less adverse coronary effect on women who

were taking statins than on those who were not, but the differences were not significant.²¹ Although the findings are of interest in view of antiinflammatory and C-reactive-protein–lowering effects of statins,²⁹ the available data do not support the use of such agents to attenuate the risk of CHD associated with postmenopausal hormone therapy unless and until clinical trials demonstrate such a benefit. Finally, the results of HERS suggested a possible reduction in the risk of CHD with hormone therapy among women with elevated base-line Lp(a) lipoprotein levels¹³; we did not observe clear evidence of cardiac protection by postmenopausal hormone therapy in this subgroup.

Some limitations of our trial deserve consideration. The WHI tested only a single regimen of estrogen plus progestin. Thus, our results do not necessarily apply to other formulations, doses, or routes of administration of these hormones, and the trial could not distinguish the effects of estrogen from those of progestin. However, randomized trials of oral or transdermal estrogen alone, to date, have had results similar to those of the WHI with regard to CHD.^{4–6,22} Another limitation is the relatively high rate of discontinuation of hormone therapy in the trial, which tends to decrease the observed treatment effects and may lead to an underestimate of adverse cardiovascular effects. Finally, because of the small size of many of the subgroups examined (which limits the statistical power to detect interactions) and the number of comparisons made (approximately 36 tests for interaction), the findings should be interpreted with caution.

In conclusion, our trial documents that estrogen

plus progestin does not have a beneficial effect on the risk of CHD among healthy postmenopausal women. Overall, the risks of treatment outweighed the benefits during 5.6 years of treatment. In view of the combined excess risk of CHD, stroke, venous thromboembolism, and breast cancer, which was not offset by the reduced risk of hip fracture and colorectal cancer,⁸ this treatment is not a viable intervention for primary prevention. Estrogen-plus-progestin therapy should not be initiated or continued for the prevention of cardiovascular disease. These conclusions are consistent with those of recently published guidelines.^{30–32} The trial did not address the role of estrogen plus progestin for the short-term treatment of menopausal symptoms, which remains the only clear indication for the use of this regimen.^{10,31} Information provided in this report about subgroups of women are exploratory and provide direction for future inquiry. In the interim, women with indications for treatment, such as menopausal symptoms, need to consider with their clinicians the suggestion of a slight overall increase in the risk of CHD and information on the risks of other outcomes in making decisions about the use of estrogen-plus-progestin therapy.

Supported by the National Heart, Lung, and Blood Institute, National Institutes of Health, Department of Health and Human Services.

Dr. Assaf is an employee of Pfizer and reports holding stock options in the company; Dr. Wong reports having received grant support from Bristol-Myers Squibb and AstraZeneca.

We are indebted to the participants, investigators, and staff of the WHI for their outstanding dedication and commitment; to Mary Pettinger, M.S., and Philomena Quinn for their expert assistance; and to Medical Research Laboratories (Highland Heights, Ky.) and the Laboratory for Clinical Biochemistry Research (University of Vermont, Burlington) for performing the biomarker assays.

APPENDIX

The following persons are investigators in the Women's Health Initiative: Program Office (National Heart, Lung, and Blood Institute, Bethesda, Md.); B. Alving, J.E. Rossouw, L. Pottern, S. Ludlam, J.A. McGowan. Clinical Coordinating Center (Fred Hutchinson Cancer Research Center, Seattle); R. Prentice, G. Anderson, A. LaCroix, R. Patterson, A. McTiernan, B. Cochrane, J. Hunt, L. Tinker, C. Kooperberg, M. McIntosh, C.Y. Wang, C. Chen, D. Bowen, A. Kristal, J. Stanford, N. Urban, N. Weiss, E. White; (Bowman Gray School of Medicine, Winston-Salem, N.C.); S. Shumaker, P. Rautaharju, R. Prineas, M. Naughton; (Medical Research Laboratories, Highland Heights, Ky.); E. Stein, P. Laskarzewski; (University of California at San Francisco, San Francisco); S. Cummings, M. Nevitt, M. Dockrell; (University of Minnesota, Minneapolis); L. Harnack; (McKesson BioServices, Rockville, Md.); F. Cammarata, S. Lindenfelser; (University of Washington, Seattle); B. Psaty, S. Heckbert.

Clinical Centers (Albert Einstein College of Medicine, Bronx, N.Y.): S. Wassertheil-Smoller, W. Frishman, J. Wylie-Rosett, D. Barad, R. Freeman; (Baylor College of Medicine, Houston); J. Hays, R. Young, J. Anderson, S. Lithgow, P. Bray; (Brigham and Women's Hospital, Harvard Medical School, Boston); J. Manson, J. Buring, J.M. Gaziano, K. Rexrode, C. Chae; (Brown University, Providence, R.I.); A.R. Assaf, R. Carlton (deceased), C. Wheeler, C. Eaton, M. Cyr; (Emory University, Atlanta); L. Phillips, M. Pedersen, O. Strickland, M. Huber, V. Porter; (Fred Hutchinson Cancer Research Center, Seattle); S.A.A. Beresford, V.M. Taylor, N.F. Woods, M. Henderson, M. Kestin; (George Washington University, Washington, D.C.); J. Hsia, N. Gaba, J. Ascensao, S. Laowattana; (Harbor-UCLA Research and Education Institute, Torrance, Calif.); R. Chlebowski, R. Detrano, A. Nelson, J. Heiner, J. Marshall; (Kaiser Permanente Center for Health Research, Portland, Oreg.); C. Ritenbaugh, B. Valanis, P. Elmer, V. Stevens, N. Karanjan, J. Marshall; (Kaiser Permanente Division of Research, Oakland, Calif.); B. Caan, S. Sidney, G. Bailey, J. Hirata; (Medical College of Wisconsin, Milwaukee); J. Morley Kotchen, V. Barnabei, T.A. Kotchen, M.A.C. Gilligan, J. Neuner; (MedStar Research Institute, Howard University, Washington, D.C.); B.V. Howard, L. Adams-Campbell, M. Passaro, M. Rainford, T. Agurs-Collins; (Northwestern University, Chicago and Evanston, Ill.); L. Van Horn, P. Greenland, J. Khandekar, K. Liu, C. Rosenberg; (Rush-Presbyterian-St. Luke's Medical Center, Chicago); H. Black, L. Powell, E. Mason; (Stanford Center for Research in Disease Prevention, Stanford University, Stanford, Calif.); M.L. Stefanick, M.A. Hlatky, B. Chen, R.S. Stafford, L.C. Giudice; (State University of New York at Stony Brook, Stony Brook); D. Lane, I. Graneck, W. Lawson, G. San Roman, C. Messina; (Ohio State University, Columbus); R. Jackson, R. Harris, D. Frid, W.J. Mysiw, M. Blumenfeld; (University of Alabama at Birmingham, Birmingham); C.E. Lewis, A. Oberman, M.N. Fouad, J.M. Shikany, D. Smith West;

(University of Arizona, Tucson and Phoenix): T. Bassford, J. Mattox, M. Ko, T. Lohman; (University at Buffalo, Buffalo, N.Y.): M. Trevisan, J. Wactawski-Wende, S. Graham, J. Chang, E. Smit; (University of California at Davis, Sacramento): J. Robbins, S. Yasmeen, K. Lindfors, J. Stern; (University of California at Irvine, Orange): A. Hubbell, G. Frank, N. Wong, N. Greep, B. Monk; (University of California at Los Angeles, Los Angeles): H. Judd, D. Heber, R. Elashoff; (University of California at San Diego, La Jolla and Chula Vista): R.D. Langer, M.H. Criqui, G.T. Talavera, C.F. Garland, R.E. Hanson; (University of Cincinnati, Cincinnati): M. Gass, S. Wermke, N. Watts; (University of Florida, Gainesville and Jacksonville): M. Limacher, M. Perri, A. Kaunitz, R.S. Williams, Y. Brinson; (University of Hawaii, Honolulu): D. Curb, H. Petrovitch, B. Rodriguez, K. Masaki, S. Sharma; (University of Iowa, Iowa City and Davenport): R. Wallace, J. Torner, S. Johnson, L. Snetselaar, B. VanVoorhis; (University of Massachusetts, Fallon Clinic, Worcester): J. Ockene, M. Rosal, I. Ockene, R. Yood, P. Aronson; (University of Medicine and Dentistry of New Jersey, Newark): N. Lasser, N. Hymowitz, V. Lasser, M. Safford, J. Kostis; (University of Miami, Miami): M.J. O'Sullivan, L. Parker, R. Estape, D. Fernandez; (University of Minnesota, Minneapolis): K.L. Margolis, R.H. Grimm, D.B. Huntinghake, J. LaValleur, K.M. Hall; (University of Nevada, Reno): R. Brunner, S. St. Jeor, W. Graettinger, V. Oujevolk; (University of North Carolina, Chapel Hill): G. Heiss, P. Haines, D. Ontjes, C. Sueta, E. Wells; (University of Pittsburgh, Pittsburgh): L. Kuller, A. Caggiula, J. Cauley, S. Berga, N.C. Milas; (University of Tennessee, Memphis): K.C. Johnson, S. Satterfield, R.W. Ke, J. Vile, F. Tylavsky; (University of Texas Health Science Center, San Antonio): R. Brzyski, R. Schenken, J. Trabal, M. Rodriguez-Sifuentes, C. Moutou; (University of Wisconsin, Madison): C. Allen, D. Laube, P. McBride, J. Mares-Perlman, B. Loewinger; (Wake Forest University School of Medicine, Winston-Salem, N.C.): G. Burke, R. Crouse, L. Parsons, M. Vitolins; (Wayne State University School of Medicine, Hutzel Hospital, Detroit): S. Hendrix, M. Simon, G. McNeely, P. Gordon, P. Makela.

REFERENCES

- Grodstein F, Stampfer M. The epidemiology of coronary heart disease and estrogen replacement in postmenopausal women. *Prog Cardiovasc Dis* 1995;38:199-210.
- Grady D, Rubin SM, Petitti DB, et al. Hormone therapy to prevent disease and prolong life in postmenopausal women. *Ann Intern Med* 1992;117:1016-37.
- Hulley S, Grady D, Bush T, et al. Randomized trial of estrogen plus progestin for secondary prevention of coronary heart disease in postmenopausal women: Heart and Estrogen/progestin Replacement Study (HERS) Research Group. *JAMA* 1998;280:605-13.
- Herrington DM, Reboussin DM, Brosnihan KB, et al. Effects of estrogen replacement on the progression of coronary-artery atherosclerosis. *N Engl J Med* 2000;343:522-9.
- Clarke SC, Kelleher J, Lloyd-Jones H, Slack M, Schoefield PM. A study of hormone replacement therapy in postmenopausal women with ischaemic heart disease: the Papworth HRT atherosclerosis study. *BJOG* 2002;109:1056-62.
- Waters DD, Alderman EL, Hsia J, et al. Effects of hormone replacement therapy and antioxidant vitamin supplements on coronary atherosclerosis in postmenopausal women: a randomized controlled trial. *JAMA* 2002;288:2432-40.
- Hemminki E, McPherson K. Impact of postmenopausal hormone therapy on cardiovascular events and cancer: pooled data from clinical trials. *BMJ* 1997;315:149-53.
- Writing Group for the Women's Health Initiative Investigators. Risks and benefits of estrogen plus progestin in healthy postmenopausal women: principal results from the Women's Health Initiative randomized controlled trial. *JAMA* 2002;288:321-33.
- Risks of postmenopausal hormone replacement. *JAMA* 2002;288:2819-25.
- Grodstein F, Clarkson TB, Manson JE. Understanding the divergent data on postmenopausal hormone therapy. *N Engl J Med* 2003;348:645-50.
- Manson JE, Martin KA. Postmenopausal hormone-replacement therapy. *N Engl J Med* 2001;345:34-40.
- Chae CU, Manson JE. Postmenopausal hormone replacement therapy. In: Hennekens CH, ed. Clinical trials in cardiovascular disease: a companion to Braunwald's Heart Disease. Philadelphia: W.B. Saunders, 1999:399-414.
- Slipak MG, Simon JA, Vittinghoff E, et al. Estrogen and progestin, lipoprotein(a), and the risk of recurrent coronary heart disease events after menopause. *JAMA* 2000;283:1845-52.
- Psaty BM, Smith NL, Lemaitre RN, et al. Hormone replacement therapy, prothrombotic mutations, and the risk of incident nonfatal myocardial infarction in postmenopausal women. *JAMA* 2001;285:906-13.
- Design of the Women's Health Initiative clinical trial and observational study: the Women's Health Initiative Study Group. *Control Clin Trials* 1998;19:61-109.
- Rautaharju PM, Park LP, Chaitman BR, Rautaharju F, Zhang ZM. The Novacode criteria for classification of ECG abnormalities and their clinically significant progression and regression. *J Electrocardiol* 1998;31(3):157-87.
- Ives DG, Fitzpatrick AL, Bild DE, et al. Surveillance and ascertainment of cardiovascular events: the Cardiovascular Health Study. *Ann Epidemiol* 1995;5:278-85.
- Wassertheil-Smoller S, Hendrix S, Limacher M, et al. Effect of estrogen plus progestin on stroke in postmenopausal women: the Women's Health Initiative. *JAMA* 2003;289:2673-84.
- Cox DR. Regression models and life-tables. *J R Stat Soc [B]* 1972;34:187-220.
- Pilon D, Castilloux AM, Lelorier J. Estrogen replacement therapy: determinants of persistence with treatment. *Obstet Gynecol* 2001;97:97-100.
- Grady D, Herrington D, Bittner V, et al. Cardiovascular disease outcomes during 6.8 years of hormone therapy: Heart and Estrogen/progestin Replacement Study follow-up (HERS II). *JAMA* 2002;288:49-57. [Erratum, *JAMA* 2002;288:1064.]
- Cherry N, Gilmour K, Hannaford P, et al. Oestrogen therapy for prevention of reinfarction in postmenopausal women: a randomised placebo controlled trial. *Lancet* 2002;360:2001-8.
- Viscoli CM, Brass LM, Kernan WN, Sarrel PM, Suissa S, Horwitz RI. A clinical trial of estrogen-replacement therapy after ischemic stroke. *N Engl J Med* 2001;345:1243-9.
- Effects of estrogen or estrogen/progestin regimens on heart disease risk factors in postmenopausal women: the Postmenopausal Estrogen/Progestin Interventions (PEPI) Trial. *JAMA* 1995;273:199-208. [Erratum, *JAMA* 1995;274:1676.]
- Mendelsohn ME, Karas RH. The protective effects of estrogen on the cardiovascular system. *N Engl J Med* 1999;340:1801-11.
- Cushman M, Meilahn EN, Psaty BM, Kuller LH, Dobs AS, Tracy RP. Hormone replacement therapy, inflammation, and hemostasis in elderly women. *Arterioscler Thromb Vasc Biol* 1999;19:893-9.
- Cushman M, Legault C, Barrett-Connor E, et al. Effect of postmenopausal hormones on inflammation-sensitive proteins: the Postmenopausal Estrogen/Progestin Interventions (PEPI) Study. *Circulation* 1999;100:717-22.
- Furberg CD, Vittinghoff E, Davidson M, et al. Subgroup interactions in the Heart and Estrogen/Progestin Replacement Study: lessons learned. *Circulation* 2002;105:917-22.
- Albert MA, Danielson E, Rifai N, Ridker PM. Effect of statin therapy on C-reactive protein levels: the pravastatin inflammation/CRP evaluation (PRINCE): a randomized trial and cohort study. *JAMA* 2001;286:64-70.
- Mosca L, Collins P, Herrington DM, et al. Hormone replacement therapy and cardiovascular disease: a statement for healthcare professionals from the American Heart Association. *Circulation* 2001;104:499-503.
- The North American Menopause Society (NAMS). Report from the NAMS Advisory Panel on Postmenopausal Hormone Therapy. Cleveland: North American Menopause Society, October 3, 2002.
- U.S. Preventive Services Task Force. Postmenopausal hormone replacement therapy for primary prevention of chronic conditions: recommendations and rationale. *Ann Intern Med* 2002;137:834-9.

Copyright © 2003 Massachusetts Medical Society.