

Properties of 2 Counterfactual Effect Definitions of a Point Exposure

W. Dana Flanders* and Mitchel Klein†

Abstract: As recognized for more than 2 decades, the way to define effects of an exposure may be unclear if the effects are conditional on occurrence of prior events. Since age-specific rates are inherently conditional on survival to the age for which rates are calculated, age-specific rate ratios may be misleading. We consider this problem in the context of a point exposure and an unmeasured risk factor that is independent of the exposure, together with potential outcome models and associated counterfactual effect definitions. The methods apply to a recurring exposure that “tracks” over time, as well as to more complicated situations (although additional issues may then arise). We identify and evaluate 2 seemingly-natural ways that the population effects of a point exposure might be defined. At least one definition of the population effects of a point exposure is identifiable, while another natural definition is not identifiable. We describe possible implications of these definitions for the distortion of time-specific rate ratios that can occur with passage of time. We discuss interpretation of effects for each definition, and how the definitions are related to selection bias as recently defined by Hernán et al (*Epidemiology*. 2004;15:615–625). We present implications for study design, and make several recommendations. Problems may be reduced or avoided by starting follow-up before onset of exposure and by using survival curves to compare exposed with unexposed.

(*Epidemiology* 2007;18: 453–460)

Results of epidemiologic studies and clinical trials are often reported as estimations of effects over time.¹ Such estimations can involve age- or time-specific incidence rate differences or rate ratios, as with proportional hazards models.² Descriptive studies also estimate age- or time-specific rates and ratios.

The importance of properly estimating and accounting for effects over time is illustrated by the recent controversy surrounding hormone replacement therapy (HRT) in postmenopausal women.^{3,4} Here, the apparent benefit of HRT in reducing risk of cardiac events suggested by observational

studies was contradicted in clinical trials. Subsequently, Prentice et al compared hazard ratios from clinical trials with those from observational studies in similar populations, and suggested that discrepancies were largely explained by appropriately accounting for time since initiation of treatment and by adjusting for confounding.^{3,4}

Important limitations of time-specific measures, such as rate ratios, are recognized. For example, rate (hazard) ratios and conditional risk ratios (their discrete counterparts), are potentially biased estimates of effect if any risk factor is unmeasured even if initially independent of the exposure.^{5,6} This bias can be seen to arise because of conditioning on a collider.⁶ These and closely-related biases have been discussed in the context of frailty models^{7–10} or estimating the etiologic fraction.^{11–13} Bias can be severe enough that observed rate ratios are less than 1.0 in later time periods even if exposure is never beneficial.^{5–7}

Counterfactual reasoning has provided much insight and furthered understanding of causal reasoning, definitions of effect, confounding, and bias.^{1,6,13–15} For a dichotomous exposure, the counterfactual approach defines the effect of exposure as the contrast in disease occurrence among the exposed with what it would have been if, contrary to fact, the exposed had been unexposed. Such definitions are counterfactual because at least one of the contrasted conditions is contrary to fact and unobservable.

In this manuscript we use counterfactual reasoning to explore the biases and limitations inherent in estimating time-specific effects. We start from fundamentals by considering the definition of effects. We show that the effects in a seemingly natural definition are not estimable, and that the rate ratio can be a biased estimate of those effects, even in an experiment. An alternative definition yields a population effect that is validly estimated by contrasts of observable cumulative risks.

To motivate our consideration of effect definitions, we start with a hypothetical example patterned after observations of cardiac events reported from the Women's Health Initiative that included a randomized clinical trial (RCT) of estrogen plus progestin in postmenopausal women.^{3,4,16}

Example

We consider a hypothetical RCT in which half the participants are randomized to treatment, and the other half to placebo (Table 1). In this RCT, the numbers of events in the first period and the hazard ratios for the first 3 periods are similar to those reported in the WHI clinical trial.¹⁶ However, we have modified the “data” in several ways: the numbers in

Submitted 13 July 2006; accepted 2 January 2007; posted 30 April 2007.
From the Departments of *Epidemiology and Biostatistics, and †Environmental and Occupational Health, Rollins School of Public Health, Emory University, Atlanta, Georgia.

Correspondence: W. Dana Flanders, Department of Epidemiology and Biostatistics, Rollins School of Public Health, Emory University, 1518 Clifton Rd., Atlanta, GA 30322. E-mail: flanders@sph.emory.edu.

Copyright © 2007 by Lippincott Williams & Wilkins

ISSN: 1044-3983/07/1804-0453

DOI: 10.1097/01.ede.0000261472.07150.4f

TABLE 1. Example Summary Observations for a Hypothetical Clinical Trial

Measure	Treatment	Year 1–2	Year 3–4	Year 5–6	Year 7–8
No. events*/No. at risk†	Untreated	47/8300	44/8253	93/8209	70/8116
No. events*/No. at risk†	Treated	80/8300	55/8220	65/8165	60/8100
Cumulative risk	Untreated	0.0057	0.0110	0.0222	0.0306
	Treated	0.0096	0.0163	0.0241	0.0313
Risk ratio (cumulative risk)		1.70	1.48	1.09	1.02
Conditional risk (each time point)	Untreated	0.0057	0.0053	0.0113	0.0086
	Treated	0.0096	0.0067	0.0080	0.0074
Conditional risk ratio		1.70	1.26	0.70	0.86
Number of events advanced by treatment‡			←33	←44	←16

*Number of coronary events in each period.

†Number at risk for event at start of each period.

‡Number of people in whom treatment advances the time of event by 1 period.

each group are exactly equal initially; additional follow-up is included so that censoring, (which could introduce additional complexity^{2,17}) occurs only after 8 years; and the time periods are equal.

Several measures of association are summarized in Table 1. The cumulative risk for each time period is the proportion of each group that has experienced an event during or before that time period. For example among the treated individuals, a total of 135 (80 + 55) people had experienced an event during or before year 3–4, so that the cumulative risk is 0.0163 (135/8300). Treatment in this example has a deleterious effect on occurrence of coronary events: cumulative risk is higher and survival is lower at every time among the treated than among the untreated (ignoring random variability).

If we were interested in questions such as whether the treatment continued to have a harmful effect among long-term users, we might be tempted to compare conditional risks (conditional risk calculated as the number of events at a given time divided by the total who were at risk for the event at that time). Such comparisons might suggest, apart from variability, that treatment has become beneficial among those who survive beyond year 5, since the conditional risk ratio is less than 1.0 in the last 2 periods (Table 1).

This conclusion could be erroneous as we now show. These same results could be observed even if treatment had only harmful effects. We assume exchangeability^{1,14,15} of the treated and untreated immediately after randomization. Exchangeability means that risks in the untreated group equal the risks that would have been seen in the treated group had the latter been (contrary to fact) untreated. In other words, the untreated group is comparable and provides “substitute” information.¹⁴ Exchangeability is made more plausible by the randomization, because risk factors should tend to be equally distributed across treatment groups. Alternatively, we would reach the same conclusions if the observations for the treated group represented the expected values, and the observations for the untreated were the counterfactual expected values (what would have happened among the treated if they had been untreated).

We suppose that treatment acted by advancing the time of the coronary event of some treated subjects, for example “by an acceleration of events in earlier years among susceptible women assigned to postmenopausal hormone therapy.”¹⁶ Using our exchangeability assumption (and ignoring random variation), we can see from Table 1 by referring to the untreated group, that 47 of those treated would have had the event in the years 1–2 had they been untreated. When numbers from the untreated group are applied to the treated group in this way, the numbers are “counterfactual” since they pertain to a condition that differs from the condition that actually occurred.¹⁵ The observed treatment effect could be produced by advancing time of occurrence of coronary events to earlier times. For example, a shift of 33 of the 44 events that occurred at year 3–4 to year 1–2 would produce the observed numbers of events in the treated group at year 1–2 (Table 1). Similarly, an advance of 44 of the 93 events at year 5–6, added to the 11 events remaining at year 3–4, (44–33) would produce the observed treatment effect (55 events) at year 3–4, and so forth. We have assumed that treatment has no effect for all others.

Note that the observed conditional risk ratios are less than 1.0 for years 5–6 and 7–8, even though treatment has had only deleterious effects in some and no effects in others. Further, had an observational study enrolled only those under treatment more than 4 years, only the latter patterns would be observed, and all conditional risk ratios would be less than one, possibly leading to the erroneous conclusion that treatment was beneficial. This design—enrolling some who have already been exposed for some time—may not be uncommon. For example, the design may be a feature of some cohort studies, including some studies of postmenopausal estrogen use.^{3,18} Age-specific rate ratios at older ages could potentially suffer the bias described here even in an experiment in which estrogen use had been randomized at an earlier age. Most case-control studies that seek to estimate age-specific rate ratios for an early exposure, perhaps with risk set sampling,¹⁹ could also be vulnerable to this bias.

The bias may also be understood by explicitly recognizing that the population is likely to be heterogeneous with

respect to one or more risk factors: a subset of the population may have an unmeasured risk factor U , while others do not. Because of randomization, the risk factor U should initially be unrelated to the randomly assigned treatment. However, among those who survive the passage of time, U can become unequally distributed with respect to treatment because survival is affected by both treatment and U .^{6,20} Here, survival is a common effect of both treatment and U , which is referred to as a collider in the DAG literature²⁰; conditioning on the collider will tend to induce an association.^{5,6,20} This unequal distribution of U tends to create confounding in later time periods^{5,6}, and illustrates how measures such as conditional risk ratios can be biased estimators of time-specific effects.^{5,9–12,21,22}

In this example, we argued for the existence of possible bias without explicitly defining the effect of interest. To understand the bias further, we now explicitly consider this issue by giving 2 possible effect definitions. Because of the central role of counterfactual reasoning (accepted by many^{14,15,23–25} but not all²⁶), we turn to the counterfactual perspective to define time-specific effects. We then investigate how the defined effects relate to conditional and other time-specific risk differences.

Two Counterfactual Effect Definitions

We consider discrete time (say 0, 1, 2, and 3+) and a dichotomous point exposure received at baseline (time 0). We give 2 alternative definitions for the effects at time 2, using counterfactual reasoning. The 2 definitions differ because of differences in the population of interest, the target population,¹⁴ each of which is a different subgroup of the exposed cohort. The definitions are:

Target Population 1 (“Full” Population)

All in the cohort who are exposed at time 0 without regard to subsequent survival.

Definition 1. The causal risk difference for the time-2 population effect of exposure (among the exposed) is the time-2 risk among the exposed in the target population, compared with what the risk would have been in the target population had they been unexposed.

Target Population 2 (Surviving Population)

All in the cohort who are exposed at time 0 and die at time 2 or later (exposed survivors).

Definition 2. The causal risk difference for the time-2 population effect of exposure (among the exposed survivors) is the time-2 risk in the target population (exposed survivors), compared with what this risk would have been in the target population had they been unexposed.

Both definitions are clearly stated using counterfactuals. In the next section, we define a potential outcome model, and use it to explore properties of these 2 definitions and relationships with conditional risk differences.

Potential Outcome Model and Effects Over Time

We now use a counterfactual (potential outcome) model to evaluate the implications of these definitions and to

express the defined effects in terms of the model parameters. (References 1, 15, 25, 27, 28 have useful definitions and applications). We consider a situation much like that in the example above—a harmful, dichotomous point exposure, randomized and given at time 0, with follow-up to identify deaths at times 1, 2, and 3+. Each individual m has a time of death T_{me} that would occur when exposure $E = e$, for $e = 0$ and 1, although we can observe time of death only under one exposure condition. Nine response types are possible (3 death times for each exposure condition, or 3^2), as summarized in Table 2. For example, one possible type is that individual m would die at time 1 ($T_{m,1} = 1$) if exposed and also if unexposed ($T_{m,0} = 1$; Type “A,” Table 2). We label the response types as A, B, . . . , or I. To further explicate Table 2 and the potential outcome model, consider another individual, say n , with response type B (Table 2, line 3). Individual n would die at time 1 if exposed ($T_{n,1} = 1$) and at time 2 if unexposed ($T_{n,0} = 2$). Thus, exposure advances time of death from 2 to 1 for those with response type B. The population frequency of response type B is p_B among the exposed and q_B among the unexposed. Three types (labeled G, H, and I in Table 2) are excluded by the assumption that E is never beneficial—which we now adopt. In Appendix 1, we give an equivalent probabilistic causal model.

To express population effects using the potential outcome model, we relate these counterfactual response types for individuals to those in the target population through the population frequency of each type (eg, the p ’s and q ’s in Table 2, similar to the p ’s and q ’s, in reference 15). In particular, we now express the causal risk differences corresponding to Definition 1 or 2 in terms of parameters of this potential outcome model.

Let $R_{i,j,k}$ denote the risk in target population k at time i , if $E = j$, for i, j , and k in $\{0,1\}$. Some of the risks are directly observable and are factual, but those under exposure conditions other than those that actually occurred are counterfactual. Under the potential outcome model, the risk at time 1 among target population 1 (the entire exposed population) is:

$$R_{1,1,1} = p_A + p_B + p_C, \quad (1)$$

and had they been unexposed would be:

$$R_{1,0,1} = p_A. \quad (2)$$

Note that we use “ p ’s” rather than “ q ’s” in these equations, since the target population is the exposed.¹⁵

The risk at time 2 among target population 1 (the entire exposed population), is:

$$R_{2,1,1} = p_D + p_F, \quad (3)$$

and had they been unexposed:

$$R_{2,0,1} = p_B + p_D. \quad (4)$$

Denote the time- i effect of exposure measured using risk differences, under definition k by $RD_{i,k}$, for i and k in $\{1,2\}$. Thus, the population effects of exposure at times 1 and

TABLE 2. Potential Outcome Model for Exposure (E) and Time of Death (T), Assuming Exposure (E) is Never Harmful

Response Type	Population Proportion for E = 1*	Population Proportion for E = 0†	Counterfactual Death Time‡ for E = 0 or 1: T _{m,e}		Comment
			T _{m,1}	T _{m,0}	
A	p _A	q _A	1	1	E has no effect
B	p _B	q _B	1	2	E advances death from time 2 to 1
C	p _C	q _C	1	3	E advances death from time 3 to 1
D	p _D	q _D	2	2	E has no effect
E	p _E	q _E	3	3	E has no effect
F	p _F	q _F	2	3	E advances death from time 3 to 2
G	0	0	2	1	Excluded, E beneficial
H	0	0	3	1	Excluded, E beneficial
I	0	0	3	2	Excluded, E beneficial

*Population proportion for E = 1 is the proportion of the exposed population (E = 1) with each counterfactual response type at time 0: p_A is the proportion with response type A among the exposed, p_B is the proportion with type B, and so on.

†Population proportion for E = 0 is the proportion of the unexposed population (E = 0) with each counterfactual response type at time 0: q_A is the proportion of type A among the unexposed, q_B is the proportion with type B, and so on.

‡Counterfactual death time, T_{m,e}, is the time, for individual m, at which death will occur under exposure condition E = e (e = 0, 1). For example, those of response type U = B would die at time 2 when exposed (T₁ = 2) and at time 1 when unexposed (T₀ = 1). For those with response type B, Exposure advances death from time 2 to time 1.

2, respectively, under this model among target population 1 (the entire exposed population) are:

$$RD_{1,1} = R_{1,1,1} - R_{1,0,1} = (p_B + p_C); \quad (5)$$

$$RD_{2,1} = R_{2,1,1} - R_{2,0,1} = p_F - p_B. \quad (6)$$

We can also use the potential outcome model to express the causal risk differences inherent in effect definition 2—for the surviving population. According to definition 2, the desired contrast is the proportion of the exposed cohort surviving past time 1 who dies at time 2, compared with the corresponding proportion if unexposed. The number surviving time 1 among the exposed (target) is proportional to p_D + p_E + p_F, so that the proportion of these survivors who would die at time 2 if exposed is (p_D + p_F)/(p_D + p_E + p_F). On the other hand, the number of these same survivors who would die in period time 2 had they been unexposed is proportional to p_D so that the proportion who would die if unexposed is p_D/(p_D + p_E + p_F). Thus, the time-2 effect under Definition 2 for survivors past time 1 is:

$$RD_{2,2} = (p_D + p_F)/(p_D + p_E + p_F) - (p_D)/(p_D + p_E + p_F) = p_F/(p_D + p_E + p_F). \quad (7)$$

The effect in Equation (7) is inherently conditional, because it is defined as a risk among the subgroup surviving the first period. We can also obtain Equation (7) using a probabilistic causal model in Appendix 1, consistent with an approach described by Pearl.²⁵

Attempts to define effects using conditional risks as in definition 2 for survivors, but with the target as the unexposed who survive time 1, lead to potential inconsistencies. The possible inconsistency arises because some of those who

would survive time 1 if unexposed may not survive if exposed (by assumption, the exposure is only harmful). Questions about survival after time 1 might then involve asking about future mortality of people already deceased. Hernán et al discussed this issue in the context of administrative censoring, accelerated time failure models and g-estimation.²⁹

Estimation

Having expressed the causal risk differences in terms of model parameters, we now consider estimation of these measures from observations of the cohort. We assume that the cohort is closed, observed from time 0–3, and large enough that sampling variability is ignorable. Thus, observed risks equal the corresponding expression as expressed in terms of model parameters. The target population is either all the exposed (definition 1) or the exposed survivors (definition 2) in the cohort.

The potential outcome model (Table 2) and the risk measures (Equations 1–7) derived from it hold in the unexposed subpopulation with p's replaced by the corresponding q's. We now assume that the unexposed are exchangeable with the exposed, so that p_x = q_x for each response type: A, . . . , F. (In the terminology of Maldonado and Greenland,¹⁴ the unexposed subpopulation provides substitute information.)

Using an underline to denote observed risk differences, the observed time-2 risk difference is: $\underline{RD}_2 = p_D + p_F - q_B - q_D$. Under the assumption of exchangeability, the expression simplifies to $\underline{RD}_2 = p_F - p_B$, which is the population effect of exposure at time 2 according to definition 1 for the entire exposed population. Similar conclusions hold for $\underline{RD}_1$ so that time-specific effects in definition 1 are identified.

We now consider estimation of the effects in definition 2 for exposed survivors. The observed (conditional) risk at time 2 among the unexposed who survive past time 1, still assuming exchangeability initially and no sampling variability, is: $(p_B + p_D)/(p_B + p_C + p_D + p_E + p_F)$, and the difference in observed conditional risks is:

$$cRD_2 = (p_D + p_F)/(p_D + p_E + p_F) - (p_B + p_D)/(p_B + p_C + p_D + p_E + p_F). \quad (8)$$

Comparing Equation (8), the difference of observed, conditional risks at time 2, with Equation (7), the time-2 exposure effect under definition 2 for survivors, shows that the desired effect is not estimated. Conceptually, this occurs because those who survive past time 1 when exposed differ from those who survive when unexposed, but definition 2 requires that we contrast risks in the same people (the target)—those who would survive if exposed. Differences in the survivors lead to different denominators in Equation (8) compared with Equation (7).

In the example, one might have thought that the observed conditional risk difference, or its ratio counterpart, would estimate treatment effects among survivors, as specified by definition 2. Our results show that the conditional risk difference and ratio may not estimate such effects.

Thus, effects in definition 1 for the entire population are estimated by contrasts of the corresponding observable risks, assuming exchangeability initially and assuming that exposure is never beneficial, whereas those in definition 2 for survivors are not estimated by such contrasts.

Model Identifiability

Although we can estimate the causal contrasts inherent in definition 1 with the stated assumptions, all model parameters are not identified without additional assumptions—different parameter values lead to the same observable risks. In particular, we cannot estimate both p_B and p_F , the proportions of the unexposed population for whom exposure would advance time of death from time 2 to 1, and from time 3 to 2, respectively. For example, observable risks are the same, but p_F differs if either: ($p_A = 0.10$, $p_B = 0.05$, $p_C = 0$, $p_D = 0.05$, $p_E = 0.65$, $p_F = 0.15$), or if: ($p_A = 0.10$, $p_B = 0$, $p_C = 0.05$, $p_D = 0.10$, $p_E = 0.65$, $p_F = 0.10$). This nonidentifiability differs from the single-period situation in which the model is identifiable given the assumptions used here—E is never beneficial and exposed and unexposed subpopulations are exchangeable.¹⁵ Nonidentifiability for the multiple-time situation can be understood intuitively by observing that the number of deaths at time 2 among the exposed could be less than the number in the unexposed merely because exposure served to advance the time of diagnosis from time 2 to time 1 (as in response type B). Thus, a relative deficit of deaths may be seen at time 2 among the exposed, only because some who would have died at time 2 were already dead. Unless we make further assumptions, the additional complexity and parameters associated with time-specific effects leads to nonidentifiability of important causal parameters.^{11,12,15}

Despite the lack of complete identifiability, some bounds on model parameters are possible¹¹ (assuming, as throughout, that exposure is not beneficial). For example, if

$RD_{1,1} > 0$ (which indicates a deleterious effect at time 1), then the proportion of cases advanced from time 2 to time 1 is at most $RD_{1,1}$. If $RD_{2,1} > 0$, then direct substitution shows that the proportion of the population in whom exposure causes death at time 2, p_F , is bounded by:

$$\max(0, RD_{2,1}) \leq p_F \leq RD_{1,1} + RD_{2,1}. \quad (9)$$

Limitation of Time-Specific Risk Differences

The observed conditional risk difference (cRD_2 , Equation [8]) and corresponding population measure have an important limitation—the differences can be positive, zero, or negative, even if the exposure has no effect at time 2.^{5,6} (This possibility is easily demonstrated by taking $p_F = 0$, assuring that E has no time 2 effect among those who would survive if exposed. Then if $p_A = 0$, $p_B = 0.10$, $p_C = 0.5$, $p_D = 0.10$, $p_E = 0.3$, and $p_F = 0$, then $cRD_2 = +0.05$. On the other hand, if $p_A = 0$, $p_B = 0.15$, $p_C = 0.5$, $p_D = 0.05$, $p_E = 0.3$, and $p_F = 0$, then $cRD_2 = -0.057$. Thus, E has no individual time-2 effect on disease among survivors, but the conditional risk difference could be either positive or negative.)

This limitation is partially shared by the (unconditional) risk difference (Equation [6]). The risk difference can be either negative or zero when E has no time 2 effect on those who would survive past time 1 if exposed ($p_F = 0$), as can be seen by inspection of Equation (6). Thus, measures of time-specific effects have a limitation not shared by measures of effects for a single, cumulative time—they can be negative even if the exposure is never beneficial.

Additional Assumption Leading to Identifiability

We temporarily adopt the additional assumption that exposure to E does not influence survival status at time 2 (or later) of any subject who would survive past time 1 when exposed. That is, E has no effects on mortality among the exposed who survive to time 2.

With this additional assumption, $p_F = 0$ and all parameters are now identifiable. For example, the observed risk difference, RD_2 now equals $-p_B$, apart from variability. Others have shown identifiability with alternative assumptions (eg, reference 17, which considers informative censoring).

DISCUSSION

It might seem natural to interpret rate ratios at older ages as estimates of effects among survivors, as in definition 2. For example, we might be tempted to interpret the reduced associations of obesity,³⁰ smoking,³¹ or hypertension³² with mortality at older ages, as implying that the effects of these factors had diminished or even reversed at older ages. The counterfactual approach here shows that such rate comparisons do not necessarily reflect causal effects among survivors^{6,11,12} and that a seemingly natural, carefully defined causal effect may not be estimated by them.

Moreover, the observed “unconditional,” time-specific risk ratio or difference (Equation [6]) can be in the opposite direction from the true causal effect. That is, this time-

specific risk difference, in contrast to cumulative risk differences, can be negative or zero even if exposure is never beneficial, similar to the more general limitations noted here and elsewhere,^{5,6,9,10} for conditional risk differences or ratios. The underlying mechanism for bias in conditional measures can be conceptualized as reflecting lack of exchangeability of the exposed and unexposed survivors in the presence of an unmeasured risk factor.⁶ In the potential outcome framework, the unmeasured risk factor might be viewed as an unmeasured predictor of specific counterfactual response types. Of course, conditional risks are the discrete counterpart of hazards so these limitations have implications for these measures as well.

We have also shown that time-specific, population causal parameters are only partially identifiable, even assuming initial exchangeability of exposed and unexposed, and assuming that exposure is only harmful^{11,12}. In particular, we cannot identify the proportion of the population for whom exposure would advance time of death from time 2 to time 1. This situation differs from that in which one considers only a single time; with multiple times, harmful effects at one time may seem to appear as a deficit of cases at other times. Further, we have shown that time-specific effects are identifiable if we additionally assume that exposure has no effect on those who would survive the first time if exposed. Essentially, this eliminates a parameter (ie, $p_F = 0$), allowing identification of the remaining parameters of interest. Assumptions allowing identification of other time-specific effects have been reported.^{11,12}

One of the definitions we considered, and perhaps a particularly natural one (definition 2), is metaphysical in that it cannot be estimated, even in a perfectly randomized trial of exposure. This definition specifically targets effects in a specific group of survivors, but, as noted, the unexposed and exposed survivors may not be exchangeable. Thus, the effect does not correspond to observable contrasts of exposed and unexposed, even in a randomized experiment. This reflects a more general problem in which effects may be difficult to define or interpret when their existence depends on occurrence of prior events, such as time-varying covariates² or survival. The other causal definition (definition 1), for the entire exposed group corresponds to effects that are identifiable and directly observable. For a point exposure, definition 1 or contrasts of cumulative risk or survival correspond to effects typically estimated in randomized trials and in approaches based on structural models.^{11–13}

To be useful, definitions, models, and effects should be related to observables. Although effects defined for those who survive past time 1 (definition 2) may not be directly observable in a simple, randomized experiment,^{6,11,12} they might be observable if, for example, a marker were available that accurately indicated the response type or an intervention that could immediately nullify any future exposure effects. The definition has relevance, however, because it focuses on effects that one might attempt to estimate in a cohort study in which subjects were enrolled sometime after a past exposure. Further, the definition does address research questions such as: “Would the exposed survivors to time t have been better off beyond time t if they had been exposed?” However, our

results show that the defined effect is not identifiable without additional assumptions or other study designs, and that the difference in observed conditional risks differs from the defined effect. Thus, the practical utility of defining effects as in definition 2 is uncertain despite philosophical appeal. Definition 1, on the other hand, addresses a different research question that might be phrased as “Would the exposed have been less likely to die at exactly this time, had they been unexposed?” It has the virtue of being directly estimable from well-designed studies.

Our purpose here is to illustrate how definitions of effect can differ, how they are related to parameters of a potential outcome model, and how they reflect limitations of time-specific measures such as the conditional risk difference. We offer advice for avoiding the associated problems.^{11,29,33–35}

We have considered a particularly simple situation in which the outcome is all-cause mortality. Additional complexity arises with censoring^{2,17,29,36} and for other outcomes such as nonfatal disease. Then, time-specific risks need not add to 1.0 over the lifetime, competing risks can affect the risk of disease occurrence, and dependent events can obscure the true effects of exposure.³⁶ We have also simplified by considering a point exposure. With an ongoing exposure, further issues arise, including the possibilities of time-varying confounding and reverse causality. Analyses may then need to incorporate a structural model or other approach.^{17,22,34,37} Despite having treated only a relatively simple case, this situation can approximate (or actually be) a special case of the more general. For example, an ongoing exposure could have effects predominantly at a young age with later exposures having little impact so that effects are approximated by an early point exposure. Thus, conclusions drawn here have implications for more complex situations.

These conclusions naturally lead to several practical considerations and recommendations. First, if a study is conducted by enrolling only those who were exposed before the start of follow-up, then all risk measures are inherently conditional on having survived to the start of follow-up. Attempts to estimate time-specific effects would be susceptible to the problems noted here, including lack of estimability and identifiability.¹² These problems may be reduced or avoided by starting follow-up early, preferably before exposure, and studying associations using a measure other than conditional risk differences, such as contrasts of cumulative survival or survival curves. For an ongoing, variable exposure, starting follow-up early might also allow the use of available approaches, such as one based on structural equations potentially accounting for time-varying confounding.

This recommendation—to start follow-up at or prior to exposure—reinforces, perhaps for different reasons, recent recommendations of Ray³⁸ (referenced by Prentice et al³) Prentice and colleagues noted that one factor contributing to discrepancies between results of observational studies and clinical trials of HRT may have been the enrollment in the cohort of women who had started HRT long before start of follow-up. In this case, a deleterious effect present soon after starting treatment may have been diluted by inclusion of those who had already been on therapy for some time. Our

results illustrate that such dilution could be entirely an artifact.¹⁶ Perhaps the dilution and discrepancies would have been less with enrollment of new users, as described by Ray³⁸ and extended by Hernán et al more generally to time-varying covariates.²²

If an early start of follow-up can not be adopted, but the goal is to estimate effects in the entire population as initially composed, (in accordance with definition 1), then the exposed and unexposed may be exchangeable at the time of exposure. However, they do not necessarily have to be comparable at the later time after follow-up has begun, as exposure can induce differences among survivors. That is, for estimating such effects the conceptual goal of adjusting for confounders should be to assure exchangeability of exposed and unexposed populations at the time of exposure. This goal may be aided by use of inverse probability of treatment estimators, if survival probabilities for the initial period and other relevant information were available. With this approach one could, if feasible, attempt to reconstruct the initial population. On the other hand, if the goal is to estimate effects among survivors as in definition 2, one must realize that the effect may not be estimable, even in randomized trials of exposure.

In summary, we have described 2 time-specific effect definitions, and found that one is inherently metaphysical and defines an effect that is not estimated by contrasts of the corresponding conditional risks. The magnitude of the discrepancy between this metaphysical effect and the conditional risk ratio or difference is not identified without additional, unverifiable assumptions. These considerations lead to the practical advice of carefully defining effects of interest, starting follow-up before exposure, and using measures such as survival curves to describe effects so that the limitations might be avoided.

ACKNOWLEDGMENTS

We thank Dr. Miguel Hernán who provided helpful comments as we revised the manuscript.

REFERENCES

- Rothman KJ, Greenland S, eds. *Modern Epidemiology*. Philadelphia, PA: Lippincott; 1998.
- Kalbfleisch JD, Prentice RL. *The Statistical Analysis of Failure Time Data*. New York: John Wiley & Sons; 1980.
- Prentice RL, Langer R, Stefanick ML, et al. Combined postmenopausal hormone therapy and cardiovascular disease: toward resolving the discrepancy between observation studies and the Women's Health Initiative Clinical Trial. *Am J Epidemiol*. 2005;162:404–414.
- Prentice RL, Langer RD, Stefanick ML, et al. Combined analysis of women's health initiative observational and clinical trial data on postmenopausal hormone treatment and cardiovascular disease. *Am J Epidemiol*. 2006;163:589–599.
- Greenland S. Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology*. 1996;114:498–501.
- Hernán MA, Hernandez-Diaz S, Robins JM. A structural approach to selection bias. *Epidemiology*. 2004;15:615–625.
- Manton Poss SS, Wing S. The black/white mortality crossover by a model of mortality selection. *Human Biol*. 1979;19:291–300.
- Manton KG, Stallard E. Methods for evaluating the heterogeneity of aging processes in human populations using vital statistics data: explaining the black–white mortality crossover by a model of mortality selection. *Human Biol*. 1981;47–67.
- Aalen O. Heterogeneity in survival analysis. *Stat Med*. 1988;7:1121–1137.
- Vaupel JW, Yashin AI. Heterogeneity's ruses: some surprising effects of selection on population dynamics. *Am Stat*. 1985;39:176–185.
- Robins JM, Greenland S. The probability of causation under a stochastic model for individual risk. *Biometrics*. 1989;45:1125–1138.
- Robins JM, Greenland S. Estimability and estimation of excess and etiologic fractions. *Stat Med*. 1989;8:845–859.
- Robins JM. Should compensation schemes be based on the probability of causation or expected years of life lost. *J Law Policy*. 2004;12:537–548.
- Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol*. 2002;31:422–429.
- Greenland S, Robins JM. Identifiability, exchangeability, and epidemiological confounding. *Int J Epidemiol*. 1986;15:413–419.
- Manson JE, Hsia J, Johnson KC, et al. Estrogen plus progestin and the risk of coronary heart disease. *New Engl J Med*. 2003;349:523–534.
- Robins JM. An analytic method for randomized trials with informative censoring: Pt 1. *Lifetime Data Anal*. 1995;1:241–254.
- Grodstein F, Manson JE, Stampfer MJ. Postmenopausal hormone use and secondary prevention of coronary events in the Nurse's Health Study. *Ann Int Med*. 2000;135:1–8.
- Lubin JH. Case-control methods in the presence of multiple failure times and competing risks. *Biometrics*. 1985;41:49–54.
- Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14:300–306.
- Greenland S. Discussion on Statistical issues in the women's health initiative. *Biometrics*. 2005;61:920–921.
- Hernán MA, Robins JM, Rodriguez LAG. Discussion on statistical issues in the women's health initiative. *Biometrics*. 2005;61:922–920.
- Rothman KJ. *Epidemiology*. New York: Oxford University Press; 2002.
- Robins J, Greenland S. Comment on causal inference without counterfactuals. *J Am Stat Assoc*. 2000;95:477–482.
- Pearl J. *Causality*. New York: Cambridge University Press; 2000.
- Dawid J. Causal inference without counterfactuals. *J Am Stat Assoc*. 2000;95:407–424.
- Greenland S, Brumback B. An overview of relations among causal modeling methods. *Int J Epidemiol*. 2002;312:1030–1037.
- Rubin DB. Comment: Neyman. (1923) and causal inference in experiments and observational studies. *Stat Sci*. 1990;5:472–480.
- Hernán MA, Cole SR, Margolick J, et al. Structural accelerated failure time models for survival analysis in studies with time-varying treatments. *Pharmacoepidemiol Drug Saf*. 2005;14:477–491.
- Stevens J, Cai J, Pamuk ER, et al. The effect of age on the association between body-mass index and mortality. *New Engl J Med*. 1998;338:1–7.
- Thun M, Myers DG, Day-Lally C, et al. Age and the exposure-response relationships between cigarette smoking and premature death in cancer prevention study II. In: *Changes in Cigarette-Related Disease Risks and Their Implication for Prevention and Control*. NIH; 1997: 383–413. Monograph No 97-4213.
- Prospective studies collaboration. Age-specific relevance of usual blood pressure to vascular mortality: a meta-analysis of individual data for one million adults in 61 prospective studies. *Lancet*. 2002;360:1903–1913.
- Robins JM, Greenland S. Identifiability and exchangeability for direct and indirect effects. *Epidemiology*. 1992;3:143–155.
- Robins JM. Causal Inference from complex longitudinal data. In: *Latent Variable Modeling and Applications to Causality. Lecture Notes in Statistics*. Vol 120. New York: Springer-Verlag; 1997:69–117.
- Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol*. 1987;125:761–768.
- Slud E, Byar D. How dependent causes of death can make risk factors appear protective. *Biometrics*. 1988;44:265–269.
- Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11:550–560.
- Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol*. 2003;158:915–920.

APPENDIX

We can equivalently express the potential outcome model used in the main text as a probabilistic causal model.²⁵ We denote the time of disease occurrence by $T \in \{1, 2, 3\}$, specify a dichotomous exposure by $E \in \{0, 1\}$, and define the categorical variable U to have a unique value in $\{A,$

$B, \dots, F\}$. U is an unobserved, “background” variable, such as the response type in the potential outcome model. We have omitted an unmeasured factor that determines E , as it is not used explicitly here. $P(U = u) = p_u$, for $u \in \{A, B, \dots, F\}$. We define the function f mapping $\{U, E\} \rightarrow T$ by:

$$\begin{aligned} T = f(U, E) &= 1, \text{ for } (U, E) \in \{(A, 1), (A, 0), (B, 0), (C, 0)\}, \\ &= 2, \text{ for } (U, E) \in \{(B, 1), (D, 1), (D, 0), (F, 0)\}, \\ &= 3, \text{ for } (U, E) \in \{(C, 1), (E, 1), (E, 0), (F, 1)\}. \end{aligned}$$

Relationships are depicted in Appendix Figure 1. With definition 2, we are targeting those who survive past time 1 when exposed, implying that $f(U, 1) > 1$. These are precisely those with $U \in \{D, E, F\}$. The subset of these people who would die at time 2 if exposed are those with $U = D$ or F . Thus, the conditional risk is $(p_D + p_F)/(p_D + p_E + p_F)$, the corresponding conditional risk in this same group of survivors if unexposed is $(p_D)/(p_D + p_E + p_F)$. The contrast for the

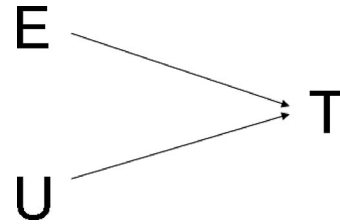


FIGURE 1. Causal relationships: Time to death T , a dichotomous exposure E , and an unmeasured risk factor U .

effect in definition 2 corresponds to the difference in these conditional risks, again yielding Equation (7).

A compact expression for these conditional risks is $P(f(U, 1) = 2 | f(U, 1) > 1)$ and $P(f(U, 0) = 2 | f(U, 1) > 1)$. We see that the conditional risk difference does correspond to a comparison of parameters in this probabilistic causal model, but that the comparison of observed risks may not correspond directly to an effect such as that in definitions 1 or 2.

ERRATA

Flanders DW, Klein M. Properties of 2 counterfactual effect definitions of a point exposure. *Epidemiology*. 2007;18:453–460.

The title of Table 2 should be:

Potential Outcome Model for Exposure (E) and Time of Death (T), Assuming Exposure (E) is Never Beneficial.

The function f in the Appendix should have been defined as:

$$\begin{aligned} T = f(U,E) &= 1, \text{ for } (U,E) \in \{(A,1), (A,0), (B,1), (C,1)\}, \\ &= 2, \text{ for } (U,E) \in \{(B,0), (D,1), (D,0), (F,1)\}, \\ &= 3, \text{ for } (U,E) \in \{(C,0), (E,1), (E,0), (F,0)\}. \end{aligned}$$

In addition, the first sentence of the last paragraph of the Appendix should be:

A compact expression for these conditional risks is $P(f(U,1) = 2 \mid f(U,1) > 1)$ and $P(f(U,0) = 2 \mid f(U,1) > 1)$.

Akbaraly TN (originally, Akbaraly NT), Hininger-Favier I, Carrière I, et al. Plasma selenium over time and cognitive decline in the elderly. *Epidemiology*. 2007;18:52–58.

The name of the first author should be Tasnime N. Akbaraly.

Abstract ISEE-681. Social status of Bulgarian population in the years of economic transition and coronary disease risk. *Epidemiology*. 2007;18:S164.

The author names were ordered incorrectly. The correct order of authors is listed here along with their affiliations.

Karaslavova E,* Turnovska T,† Semerdjieva M,‡ Petrova N,‡ Todorova D,‡ Chompalova B.§

*Department of Social Medicine and Health Management, Medical University, Plovdiv, Bulgaria; †Department of Hygiene and Ecomedicine, Medical University, Plovdiv, Bulgaria; ‡Department of Social Medicine and Health Management, Medical University, Sofia, Bulgaria; and §MHAT Plovdiv, Bulgaria.