Original article

Concepts and pitfalls in measuring and interpreting attributable fractions, prevented fractions, and causation probabilities

Sander Greenland MA, MS, DrPH [a,b,*]

[a] Department of Epidemiology, University of California, Los Angeles, Los Angeles
[b] Department of Statistics, University of California, Los Angeles, Los Angeles

ABSTRACT

Measures of causal attribution and preventive potential appear deceptively simple to define, yet have many subtle variations and are subject to numerous pitfalls in conceptualization, interpretation, and application. This article reviews basic concepts, measures, and problems to serve as an introduction to more detailed literature. Allowing for validity and generalization (projection) issues, epidemiologic attribution measures can serve as useful policy inputs for contrasting expected caseloads or survival times under different well-defined interventions. Nonetheless, their application in these settings requires attention to effects of the interventions besides those on the study outcome. Their use as estimates of etiologic attribution requires assumptions beyond the usual validity and statistical assumptions; these further assumptions will usually have little support or plausibility when the mechanisms of action are unknown.

© 2015 Elsevier Inc. All rights reserved.

Introduction

Since Levin's [1] landmark article on what he termed "attributable proportion," there has been extensive growth of concepts, definitions, and measures of attribution and prevention, along with a proliferation of terminology that now includes attributable fraction, attributable risk, attributable risk percent, preventable fraction, prevented fraction, assigned shares, excess fraction, risk fraction, rate fraction, and etiologic fraction. The present article provides an elementary overview of these concepts and certain misunderstandings and errors that have been common in the literature. It is intended to provide a basis for approaching more detailed works explaining these problems and their solutions [2–14]. The underlying theme is that causal attribution is a far more complex task than basic formulas and statistical treatments make it appear even if one can eliminate or control for all well-recognized bias sources.

To provide a precise framework, the article begins by reviewing elements of abstract causal models. It then describes basic measures of attribution for survival time, risk, caseload, and rates, along with a brief overview of estimation issues. This review is followed by a discussion of the discrepancy between rate and caseload attribution and between excess and etiologic (causal) attribution. For brevity, I focus on comparing two exposure or treatment levels in one population; there is a large literature on extensions to situations involving multiple exposure levels and multivariate exposures and their interactions [15–31]. I only briefly mention the extremely important issue of connecting the resulting attribution measures to actual interventions [32–39]. I do not address validity problems particular to attributable-fraction estimation; there are many discussions [7,40–44]. I also do not address purely statistical aspects of attributable-fraction estimation, for which the literature is vast and continues to grow apace.

Some elements of potential-outcome causal models

The framework used here to make concepts precise will be the potential-outcome (counterfactual) model of causation. For those already familiar with these models, this section need be only consulted for notation.

Suppose we have a treatment or intervention variable X, which may have more than two possible levels. If X is a quantity, what follows will assume that X is scaled and centered so that level 0 denotes a reference treatment level and level 1 denotes a targeted level for administration. For example, X might represent the

* Corresponding author. Department of Epidemiology and Department of Statistics, University of California, Los Angeles, CA
E-mail address: lesdomes@ucla.edu.

prescribed dose given of the drug treatment, with 0 representing a placebo, and 1 representing an active treatment of 40 mg/day; then 1.5 would represent a 60 mg/day. Suppose we also have an outcome variable Y, which may be of any form (e.g., binary, polytomous, quantitative, multidimensional, or multivariate).

A model for the regression of Y on X is simply a model for the mean ("expectation") of Y given X, $E(Y|x)$, describing how that mean varies as one moves across subgroup with different values x for X. Such models are used for passive prediction or description of association. Causal (or "structural") models replace Y with at least two distinct potential outcomes: Y_1 = outcome if X = 1 is administered, and Y_0 = outcome if X = 0 is administered. Because only one level of X can be administered, at most only one of these potential outcomes can be observed; others remain unobserved or "missing." For example, if X = 1 is the actual treatment, only Y_1 is observable; X = 0 is then the counterfactual treatment and Y_0 is missing. On the other hand, if X = 0 is the actual treatment, only Y_0 is observable; X = 1 is then the counterfactual treatment and Y_1 is missing.

In the special case in which X can take only two levels (e.g., X = 1 if a booster vaccination is received, 0 if not), we can summarize the difference between ordinary regression and causal modeling as follows: the traditional (descriptive) single-outcome variable used in ordinary regression analysis is $Y = XY_1 + (1-X)Y_0$. Causal modeling instead treats the two-dimensional variable $Y = (Y_1, Y_0)$ as the outcome, with the treatment variable X indicating which component of Y we may measure once treatment is given.

In general, X and Y may take any form (binary, polytomous, quantitative, and multivariate). For example, X may represent a whole treatment protocol, in which case it would be multivariate, and any particular value x for X would encode particulars of treatment (such as dose level and dose timing). To allow for this extension, we may write Y_x for the outcome that would follow X = x, where now X may vary over multiple dimensions. Second, the observable outcome variables Y_x may be replaced by parameters θ_x of a distribution of the outcome, leaving the observed outcome to represent a random draw from the distribution determined by θ_x; this extension is called a stochastic potential-outcome model and is useful for analyzing probability of causation [11].

Basics of attribution

Consider two men labeled patient A1 and A2, who at age 60 years are placed on 40 mg/day atorvastatin (treatment, indicated by X = 1), then die at ages 70 and 74 years. Let T be survival time past age 60 years, with T playing the role of the outcome variable Y. Then $T_1 = 10$ and $T_1 = 14$ are the survival times given X = 1 for these two patients, with an average value for T_1 (survival past 60 years when untreated) of 12. We might ask several questions about whether and how treatment was involved in survival; the answers would shape our measure of attribution.

A basic question is "what would have happened had they simply not been prescribed the treatment or any substitute?" That is, what if they had received 0 mg/day of statins (X = 0)? If we were interested only in survival time, regardless of cause of death, we might attempt to estimate or impute the unobserved average survival time at 0 mg/day (the average value of T_0) from men taking 0 mg/day matched as closely as possible on birth year and baseline factors affecting survival such as clinical measurements, comorbidities, and so on at age 60 years. Suppose there were two such matched control patients, with the one matched to patient A1 dying at 67 years (call him patient B1) and the other matched to patient A2 dying at 73 years (call him patient B2). Then the estimated average T_0 (survival past 60 years when untreated) for the treated patients A1 and A2 is $(7 + 13)/2 = 10$, which is the average T_0 among the untreated controls.

There are several ways we could compare and combine the 12-year average survival observed with treatment and the 10-year average survival without. We could estimate the extra average survival among the treated attributable to treatment as $12 - 10 = 2$ years, or as $12/10 = 1.20$-fold increase, or as $(12 - 10)/10 = 0.20 = 20\%$ increased survival attributable to treatment or as $(12 - 10)/12 = 0.17 = 17\%$ of average survival attributable to treatment. These are estimated average effects among the treated.

We may also attempt to estimate effects in very small groups or individuals. Under ideal conditions, the death age of patient B1 (67, which makes $T_0 = 7$ for B1) would equal the death age patient A1 would have had if untreated (making the unobserved T_0 for A1 equal 7), in which case we could say that treatment delayed the death of A1 by 3 years, from age 67 to age 70 years. In parallel, because B2 died at 73, treatment delayed the death A2 from age 73 to age 74. Thus, we could estimate $T_1 - T_0$ as $10 - 7 = 3$ years for A1, and as $14 - 13 = 1$ year for A2.

The preceding estimates assume, however, that we are interested in the effect 40 mg/day had in comparison to 0 mg/day; that is, they assume that the treated state is the target (study) condition with the untreated state as the reference (control) condition. We might instead be interested in the reverse question, namely the effect of 0 mg/day in comparison to 40 mg/day, in which case the untreated state is the target, with 40 mg/day as the referent. We could then estimate the average survival reduction among the untreated attributable to failure to treat (nontreatment, indicated by X = 0) as $10 - 12 = -2$ years, or two expected years of life lost (YLL) from failure to treat; or as $10/12 = 0.83$-fold decrease in average survival (relative average survival); or as a $(10 - 12)/12 = -0.17$ or 17% reduction in average survival attributable to failure to treat. Under ideal conditions, we could further estimate the individual YLL as $T_0 - T_1$ as $7 - 10 = -3$ years for B1, and as $13 - 14 = -1$ year for B2.

Allowing that conditions are usually less than ideal because of problems such as inexact matching (residual confounding), there is nothing controversial or particularly subtle about any of the above measures. Epidemiology textbooks however often have little or no coverage of effect measures based on survival time. Instead, they focus on counting the number of deaths under each treatment and dividing that by one of two types of denominators, which lead to two different types of attributable fractions.

Most easy to understand correctly are cohort measures which divide the number of deaths by the number starting follow-up (a count denominator), showing the proportion of that number dying by various time points during follow-up. These proportions are examples of incidence proportions, also known as average risks or simply risks. In the example, the incidents being counted are deaths, and the proportion dying (mortality proportion) among the untreated is $0/2 = 0$ until 7 years, then $1/2 = 0.5$ until 13 years, then $2/2 = 1$ after that; among the treated, the proportion is $0/2 = 0$ until 10 years, then $1/2 = 0.5$ until 14 years, then $2/2 = 1$ after that. Thus, the difference in proportions is 0 until 7 years, then 0.5 until 10 years, then 0 again until 13 years, then 0.5 again until 14 years, and finally 0 again after that (when everyone has died).

The attributable fraction for the proportion, or risk fraction, divides these differences by the proportion dying in the untreated (when i.e., not zero), expressing the excess proportion from failure to treat as a proportion of the total incidence (which here is mortality). In the example, the risk fraction is undefined until 7 years, when it becomes $(0.5 - 0)/0.5 = 1$ until 10 years, then $(0.5 - 0.5)/0.5 = 0$ until 13 years, then $(1 - 0.5)/1 = 0.5$ until 14 years, and then $(1 - 1)/1 = 0$ thereafter.

More generally, let R_t the proportion ("risk") of outcome events occurring over a period in a target cohort (here, the untreated) and let R_r be the proportion of events that would have occurred over the