White Rose Research Online

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk
Universities of Leeds, Sheffield and York

Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/id/eprint/158503/

Version: Accepted Version

Article:
Churchill, S., Angus, C., Purshouse, R. et al. (2020) Expanding attributable fraction applications to outcomes wholly attributable to a risk factor. Statistical Methods in Medical Research, 29 (9). pp. 2637-2646. ISSN: 0962-2802

https://doi.org/10.1177/0962280220907113

Churchill S, Angus C, Purshouse R, Brennan A, Sherk A. Expanding attributable fraction applications to outcomes wholly attributable to a risk factor. Statistical Methods in Medical Research. 2020;29(9):2637-2646. Copyright © 2020 The Author(s). DOI: https://doi.org/10.1177/0962280220907113. Article available under the terms of the CC-BY-NC-ND licence (https://creativecommons.org/licenses/by-nc-nd/4.0/).

Reuse
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

UNIVERSITY OF LEEDS
University of Sheffield
UNIVERSITY of York

Expanding attributable fraction applications to outcomes wholly attributable to a risk factor

Abstract

The problem central to this document is the estimation of change in disease attributable to an epidemiological exposure variable that stems from a change in the distribution of that variable. We require that both disease and exposure are quantifiable as real numbers, and then ask how to estimate the fraction of disease attributable to exposure, producing the general attributable fraction methodology. After the mathematical framework is in place, we explore the implications of a disease that is wholly attributable to a given risk factor, demonstrate why standard applications of the attributable fractions do not extend, and present general methodological considerations for this case. Finally, we demonstrate the methodology using the example of alcoholic psychoses.

Introduction

The population attributable fraction is a common tool for estimating the proportion of disease that is attributable to a given risk factor in epidemiology. Two extensions of the total population attributable fraction are the estimation of disease distribution among population subgroups defined by exposure patterns and the estimation of the change in disease incidence resulting from a change in exposure distribution. The first application is realized by isolating the components of the attributable fraction when the exposure distribution is categorical[1] and by changing the bounds of integration, thus forming exposure categories, when the exposure distribution is given by a continuous function. The second application is practically realized by altering the exposure distribution or the probability of disease given exposure in the generalized attributable fraction.[2] The classical formulations of the population attributable fraction contain either a term for incidence of disease among a reference group (typically an unexposed population), or measures probability of disease given exposure relative to such a reference group. Regardless, when the incidence of disease among the reference group is exactly 0, the population attributable fraction collapses to 1, as we will see in the following section. When incidence of disease among the reference group is 0, we say that the disease is *wholly attributable* to the risk

factor, as opposed to *partially attributable* when incidence among the reference group is strictly greater than 0 and strictly less than 1. The questions of disease distribution by exposure group and change in disease prevalence stemming from change in exposure are still relevant in the study of wholly attributable diseases, so we develop and demonstrate methodology to that end.

The aims of this paper are threefold:

1. To present a derivation of the generalized attributable fraction and the two applications described above,
2. To explore and resolve the situation for wholly attributable disease, to which the usual methodology does not extend,
3. To develop an alternative formulation for wholly attributable disease distribution that permits the two applications of interest.

We first derive the generalized population attributable fraction as a computable metric when both the distribution of exposure and its relationship to risk relative to a control are known. The applications of disease distribution over exposure patterns and change in disease incidence from change in exposure are then explored, we discuss the obstruction to expanding these applications to results wholly attributable to the exposure

variable, and propose a general method that is currently in use by the authors.[3-5] Finally, we present a worked example in the context of alcohol epidemiology.

Generalized attributable fraction

Derivation of the population attributable fraction is identical to that found in Eide & Heuch.[2] We begin by deriving an estimate for the fraction of disease attributable to a given exposure variable. Let X be a real-valued random exposure variable whose distribution over the population S is the cumulative distribution function $F(x) = Pr(X \leq x)$. Let D denote a disease that may be attributable to exposure to X. Then the probability of such a disease is denoted $Pr(D)$, and the conditional probability of disease as a function of exposure level is denoted $p(x) = Pr(D|X = x)$. The unconditional probability of disease is then:

$$Pr(D) = \int_{\mathbb{R}} p(x)dF(x) \quad (1)$$

We obtain the fraction of disease attributable to exposure to a given risk factor by considering a counterfactual exposure scenario with random exposure variable X^* on $\mathbb{R}$

with cumulative distribution function $F^*(x) = Pr(X^* \leq x)$. In particular, suppose we have a scenario where exposed E means $X > x_0$, and $Pr(E) = 0$. We define the special case of a counterfactual exposure distribution F^* in which no person is exposed to the risk factor, i.e. $Pr^*(E) = 0$. We denote this distribution by F_0, and F_0 is then defined by $F_0(x_0) = P(E') = 1$. We also assume that the relationship between exposure and probability of disease is unchanged, so the unconditional probability of a disease event in this exposure scenario is:

$$Pr_0(D) = \int_{\mathbb{R}} p(x) dF_0(x) \tag{2}$$

The excess probability of disease is then $Pr(D) - Pr_0(D)$ and the fraction of disease attributable to exposure, i.e. the population attributable fraction, is:

$$\lambda = \frac{Pr(D) - Pr_0(D)}{Pr(D)} \tag{3}$$

There is an unexposed population whose proportion is exactly $F(x_0) = \int_{x \leq x_0} dF(x)$ and an exposed population whose proportion is $1 - F(x_0) = \int_{x \geq x_0} dF(x)$. In the prospective scenario of no exposure, $F_0(x_0) = 1$ and $1 - F_0(x_0) = 0$ by definition. The risk of disease among the unexposed is given by

$$R' = \int_{x<x_0} p(x)dF(x) \tag{4}$$

We first consider the case where the disease is only partially attributable to the risk factor and exposure to the risk factor in the population is incomplete. Here, the unexposed are at risk of disease and there exists an unexposed population (i.e. $p(x) > 0$ and $F(x) > 0$ on a subinterval of $(-\infty, x_0]$), so $R' > 0$. We break down the components of λ as:

$$\begin{aligned} Pr(D) &= Pr(D \cap E) + Pr(D \cap E') \\ &= Pr(D \cap X \geq x_0) + Pr(D \cap X \leq x_0) \\ &= \int_{x \geq x_0} p(x)dF(x) + F(x_0) \cdot R' \end{aligned} \tag{5}$$

$$\begin{aligned} Pr_0(D) &= Pr_0(D \cap E) + Pr_0(D \cap E') \\ &= Pr_0(D \cap X \geq x_0) + Pr_0(D \cap X \leq x_0) \\ &= \int_{x \geq x_0} p(x)dF_0(x) + F_0(x_0) \cdot R' \\ &= 0 + 1 \cdot R' = R' \end{aligned} \tag{6}$$

We modify λ into an expression in terms of relative risk, taking $RR(x) = p(x)/R'$, so that only two quantities need to be known in order to practically compute λ: the relative risk function $RR(x)$ and the exposure distribution $F(x)$.

$$\lambda = \frac{Pr(D) - Pr_0(D)}{Pr(D)}$$

$$= \frac{\int_{x \geq x_0} p(x)dF(x) + F(x_0) \cdot R' - R'}{\int_{x \geq x_0} p(x)dF(x) + F(x_0) \cdot R'}$$

$$= \frac{\int_{x \geq x_0} p(x)dF(x) - \left(1 - F(x_0)\right) \cdot R'}{R' + \int_{x \geq x_0} p(x)dF(x) - \left(1 - F(x_0)\right) \cdot R'}$$

$$= \frac{\int_{x \geq x_0} p(x)dF(x) - \int_{x \geq x_0} R'dF(x)}{R' + \int_{x \geq x_0} p(x)dF(x) - \int_{x \geq x_0} R'dF(x)}$$

$$= \frac{\int_{x \geq x_0} (p(x) - R')dF(x)}{R' + \int_{x \geq x_0} (p(x) - R')dF(x)}$$

$$= \frac{\int_{x \geq x_0} (RR(x) - 1)dF(x)}{1 + \int_{x \geq x_0} (RR(x) - 1)dF(x)} \qquad (7)$$

Note that Equation (7) is the continuous version of Levin's formula from 1953.[6] This formulation readily permits our two applications of interest: computation of attributable fractions under prospective exposure scenarios, and calculation of disease distribution among population subgroups defined by exposure levels.

For the first application, we are interested in the new attributable fraction that would arise under a prospective exposure scenario, and that exposure scenario is represented by the distribution $F^*(x)$. The new distribution is still compared to the counterfactual of no

exposure, so the new attributable fraction computation carries F^* through to produce the scenario attributable fraction

$$\lambda^* = \frac{\int_{x \geq x_0} (RR(x) - 1) dF^*(x)}{1 + \int_{x \geq x_0} (RR(x) - 1) dF^*(x)} \tag{8}$$

There is another wrinkle if we want to further apply the attributable fraction to observed counts of disease events: the disease events were observed under actual exposure levels, so an adjustment must be made to total number of disease events in order to estimate the number attributable to the risk factor under the prospective exposure scenario. We perform this adjustment by making the assumption that the number of *non*-attributable disease events is invariant under the prospective exposure scenario. In other words, denoting by N the total number of observed disease events, the fraction λN is the number of events we estimate to be attributable to the risk factor, and its complement $(\lambda - 1)N$ is the number of events we estimate to be *not* attributable to the risk factor. If N^* denotes the total number of disease events under the prospective exposure scenario (a figure that is necessarily unobservable), then we expect the number of non-attributable events, given by $(\lambda^* - 1)N^*$, to be invariant under this change in exposure. This is represented by the relation

$$(\lambda - 1)N = (\lambda^* - 1)N^* \tag{9}$$

The metric we are trying to obtain for comparison is $\lambda^* N^*$, and that is obtained from Equation (9) as

$$N^*\lambda^* = N\frac{(1-\lambda)}{(1-\lambda^*)}\,\lambda^* \tag{10}$$

Now that we have access to this quantity, we can meaningfully compare estimates of risk attributable disease among different exposure scenarios.

The second application is realized by modifying the bounds of integration in the numerator of Equation (7). If a subpopulation S_1 is defined by exposure levels between x_1 and x_2, $x_1 < x_2$, then the proportion of disease suffered by population S_1 that is attributable to exposure is given by the quantity

$$\lambda_1 = \frac{\int_{x_1}^{x_2}(RR(x)-1)dF(x)}{1+\int_{x\geq x_0}(RR(x)-1)dF(x)} \tag{11}$$

Notably, this fraction is applied directly to the total number of disease events N, and direct comparisons can be made between subpopulations. For example, $\lambda_1 N$ would be the estimated number of disease events suffered by the population S_1 that are attributable to the risk factor, while λN would produce the estimated number of attributable disease events suffered by the whole population. Then λ_1/λ would be the proportion of attributable

disease suffered by population S_1. e.g. if $\lambda_1 = 0.3$ and $\lambda = 0.4$, then 75% of disease attributable to exposure to the risk factor would be felt by population S_1. This application is essentially a breakdown of the continuous attributable fraction into a categorical attributable fraction where we may directly compute variable components of the attributable fraction, in the language of Eide and Gefeller, Equation (4),[1] and later Eide and Heuch, Equation (20).[7]

Let us now consider the case where the chosen disease is wholly attributable to the risk factor, i.e. where $F(x_0) > 0$, and $p(x) = 0$ on $[0, x_0]$, so $R' = 0$. This occurs in practice in alcohol epidemiology, where there exists a population that is not exposed to alcohol and there exist diseases, such as alcoholic psychoses, that are not present among this unexposed population. In this case, the final line of Equation (7) is not achievable as it requires division by R', so consider instead the representation of λ found one line up:

$$\lambda = \frac{\int_{x \geq x_0} (p(x) - R')dF(x)}{R' + \int_{x \geq x_0} (p(x) - R')dF(x)} \tag{12}$$

Here we see that λ collapses to 1 when $R' = 0$ as expected: all of the disease is attributable to the risk factor because the disease is not found when the risk factor is absent. Furthermore, any alternate exposure scenario enacted by replacing $F(x)$ with a prospective

exposure distribution $F^*(x)$ will also produce an attributable fraction of 1. Our second application of categorizing subpopulations by exposure yields:

$$\lambda_1 = \frac{\int_{x_1}^{x_2} p(x)dF(x)}{\int_{x \geq x_0} p(x)dF(x)} \tag{13}$$

Here we see that the expression for the categorical attributable fraction is a computable quantity if we have access to $p(x)$, the conditional probability of disease as a function of exposure level. If we have access to such a function, then we can compute probabilities of disease directly from Equation (1). Furthermore, our two desired applications are now available. Estimation of disease incidence under a prospective exposure distribution is realized by replacing the exposure distribution F with the prospective exposure distribution F^*, and estimating the burden of disease among subpopulations defined by exposure intervals is realized by varying the bounds of integration. This is the direction explored in the next section, where we present an example of our applications in the realm of alcohol epidemiology.

Application in Alcohol Epidemiology

The formulation of the attributable fraction presented above is well suited to alcohol epidemiology, where relative risk functions and exposure distributions are readily available. Continuous relative risk functions that produce the conditional probability of suffering a disease event as a function of mean daily alcohol exposure are typically produced by combining analyses of risks across different population strata, summarizing constituent studies regarding each disease into a single function.[8] A comprehensive list of relative risk functions is compiled for the International Model of Alcohol Harms and Policies.[4] Prevalence of alcohol exposure is modeled by a scaled gamma distribution whose mean and standard deviation are linearly related,[9] yielding an exposure distribution determined entirely by population mean daily alcohol exposure.

Attributable fractions for diseases partially attributable to alcohol are therefore obtainable via the generalized attributable fraction, as are fractions for the two applications discussed in the previous section. For diseases wholly attributable to alcohol, we propose a different methodology that is based on the expression of expected value seen in Equation (1) and presented again below with exposure distribution determined by population mean alcohol exposure parameter μ. We also introduce an upper bound on our alcohol exposure distribution as a practical consideration in line with previous burden of disease estimates.[5, 10] The expected value expression is then given by Equation (14):

$$Pr(D) = \int_{x_0}^{x_1} p(x) dF(x; \mu) \tag{14}$$

If we assume that the risk among the unexposed for a given disease is zero, then all disease events are attributable to alcohol. In this case we are primarily concerned with the total number of disease events N in the population S, i.e. the incidence of disease $N/|S|$, which is exactly $Pr(D)$. To determine incidence in a prospective exposure scenario, we use the alternative mean daily exposure μ^* to generate the exposure distribution:

$$Pr^*(D \cap CD) = \int_{x_0}^{x_1} p(x) dF(x; \mu^*) \tag{15}$$

To determine disease distribution among population subgroups defined by intervals of mean daily exposure, we change the bounds of integration.

These applications require the function $p(x)$, a continuous function that provides the conditional probability of disease event as a function of mean daily alcohol exposure. In alcohol epidemiology, such functions are not available in the literature at large, so proximate *absolute risk functions* are generated as needed in applications such as the Sheffield Alcohol Policy Model[3, 11] and the International Model of Alcohol Harms and Policies[4].

For an example of generating an absolute risk functions, consider alcoholic psychoses, a condition that does not occur outside of current drinkers of alcohol. We make two assumptions that provide a general form for the absolute risk function:

1. Moderate to severe alcohol use disorder is a factor in diagnosis of alcoholic psychoses,[12] so we assume that the disease does not occur below a fixed threshold parameter t.
2. A common technique in estimating relative risk curves is the fractional polynomial technique[13] and it is often the case that fractional polynomials collapse to loglinear curves.[4, 14, 15] In short, the majority of relative risk curves for alcohol-related conditions are loglinear, so we assume that $p(x)$ increases exponentially in x.

These assumptions lead to the following form that depends on the parameters k and t:

$$p(x;k,t) = Pr(D|X=x) = \begin{cases} 0, & x < t \\ \exp\big(k(x-t)\big) - 1\,, & x \geq t \end{cases} \tag{16}$$

Here t is a fixed threshold parameter and k is an unknown slope parameter. The unconditional probability of a disease event is the incidence of events among the population, i.e. $N/|S|$, so we have the following relation:

$$\int_{x_0}^{x_1} p(x;k,t)dF(x;\mu) = \frac{N}{|S|} \tag{17}$$

The left hand side of the equation is an unbounded strictly increasing function in k, for which

$$\int_{x_0}^{x_1} p(x;0,t)dF(x;\mu) = 0 \tag{18}$$

Furthermore, $N/|S|$ is a positive value, thus Bolzano's theorem[16] guarantees a unique solution $k = \hat{k}$. When a suitable $\hat{k}$ is found, observe that the following integral is identically 1:

$$\frac{|S|}{N}\int_{x_0}^{x_1} p\left(x;\hat{k},t\right)dF(x;\mu) \tag{19}$$

Moreover the integrand is nonnegative on its domain of $[x_0, x_1]$. Letting $F' = f$, this allows us to interpret the following expression as a probability density function in x, describing the distribution of disease due to alcohol exposure:

$$\frac{|S|}{N}\, p\left(x;\hat{k},t\right)f(x;\mu) \tag{20}$$

A change in the total exposure of alcohol among a given population also changes the rate of alcohol-attributable disease. For our formulation of attributable disease estimation, observe

that such a change is equivalent to changing the population mean daily exposure parameter μ in $f(x;\mu)$ to a new value μ^*. Swapping the resulting unconditional probability distribution $f(x;\mu^*)$ into the expectation integral with the calibrated conditional probability function produces an estimate for disease incidence under the new exposure scenario.

To further illustrate the methodology, we present a worked example of calibrating and using an absolute risk function corresponding to the risk of hospitalization due to alcoholic psychoses by daily alcohol exposure among men aged 35-64 in Canada in 2014. The population of this group was 7,033,524, of which 81.26% were classified as current drinkers: persons who had consumed at least one 12g alcoholic drink in the past year. This population consumed 29.54 grams-ethanol per day on average and suffered 5,865 recorded hospitalizations due to alcoholic psychoses.[10] Because classification as a current drinker one to have consumed at least one 12g alcoholic drink in the past year, we set the lower bound of mean daily exposure to 0.03 grams/day. We set the upper bound of mean daily exposure to 250 grams/day, which corresponds to the mean exposure levels observed among Canadian street-involved persons living with alcohol dependence.[17] It is typical in estimates of alcohol attributable burden of disease to set an upper bound on alcohol exposure that is lower than the maximum observed in a given population[5, 10], in part due to the need to extrapolate dose-response risk relationships for partially attributable disease

beyond a mean daily exposure of 150 grams/day. This, in turn, is due to the relatively small population and low follow-up rate of heavy drinkers[18-20]. The upper bound of 250 grams/day was therefore chosen to provide estimates that can be reasonably compared with those found in the Canadian Substance Use Costs and Harms[10] study.

Following Kehoe et al.,[9] we use a scaled gamma distribution to model the distribution of alcohol exposure. Under this gamma model, we assume a linear relationship between the mean and standard deviation with $\mu = 1.171\sigma$. We determine the gamma distribution shape κ and scale θ parameters from mean daily exposure by the relations $\mu = \kappa\theta$ and $\sigma = \sqrt{\kappa\theta^2}$, so the distribution itself is determined entirely from the mean daily exposure μ. We require that the exposure distribution $F(x;\mu)$ satisfies

$$\int_{x_0}^{x_1} dF(x;\mu) = \int_{0.03}^{250} dF(x;29.54) \approx 0.8126 \equiv P_{CD} \tag{21}$$

To that end, we rescale the density function of the gamma distribution with mean $\mu = 29.54$ and standard deviation $\sigma = \frac{29.54}{1.171}$, denoted by $\gamma(x;\mu = 29.54, \sigma = \frac{\mu}{1.171})$, by a factor of

$$r = \frac{0.8126}{\int_{0.03}^{250} \gamma\left(x;\mu = 29.54, \sigma = \frac{\mu}{1.171}\right) dx} \tag{22}$$

This yields a continuous exposure distribution of

$$f(x; 29.54) = r \cdot \gamma\left(x; \mu = 29.54, \sigma = \frac{\mu}{1.171}\right) \tag{23}$$

The associated cumulative distribution function is defined by

$$F(x; 29.54) = \int_{0.03}^{x} f(x; 29.54)dx \tag{24}$$

Our project team sets the threshold parameter t to be the minimum value for which all persons with mean daily exposure at least t are classified as binge drinkers. This value is 67.25 grams-ethanol/day, as binge drinking for Canadian men is defined as exposure of five Canadian standard drinks (at 13.45 grams-ethanol per drink) within a single drinking event.

Recall that our assumed form for absolute risk of alcoholic psychoses hospitalization is

$$p(x; k, t) = Pr(D|X = x) = \begin{cases} 0, & x < t \\ \exp\big(k(x - t)\big) - 1, & x \geq t \end{cases} \tag{25}$$

We have already set $t = 67.25$, so we need to determine a value of k. This is done by ensuring that the following relation holds:

$$\int_{0.03}^{250} p(x; k, 67.25)\, dF(x; 29.54) = \frac{5{,}865}{7{,}033{,}524} \tag{26}$$

Solutions are found in the InterMAHP package[4] by local, derivative free constrained optimization by linear approximations[21] and is implemented via the `nloptr` R package.[22] In our example, we calibrate and plot absolute risk, disease density, and cumulative disease functions. Our exposure distribution estimates that approximately 25% of the population in question consumes more than six Canadian standard drinks per day ($6 \times 13.45 = 80.7$ grams-ethanol), and we provide an estimate of the proportion of disease due to alcoholic psychosis this population subgroup suffers. The `ggplot2` R package[23] is used to produce all plots.

First, let us examine the loglinear slope. Solving Equation (26) yields a loglinear slope of $\hat{k} \approx 2.96 \times 10^{-4}$, and the corresponding absolute risk curve $p(x; 2.96 \times 10^{-4}, 67.25)$ is presented in Figure 1. Note that a loglinear slope of this scale produces nearly linear behaviour on the interval $[0.03,250]$.

The disease probability density function and cumulative distribution functions displayed in Figures 2 and 3 show a fairly significant skew towards lower levels of alcohol exposure in the density of disease. Near-linearity of the absolute risk function indicates that this shape is largely influenced by the gamma distribution describing exposure.

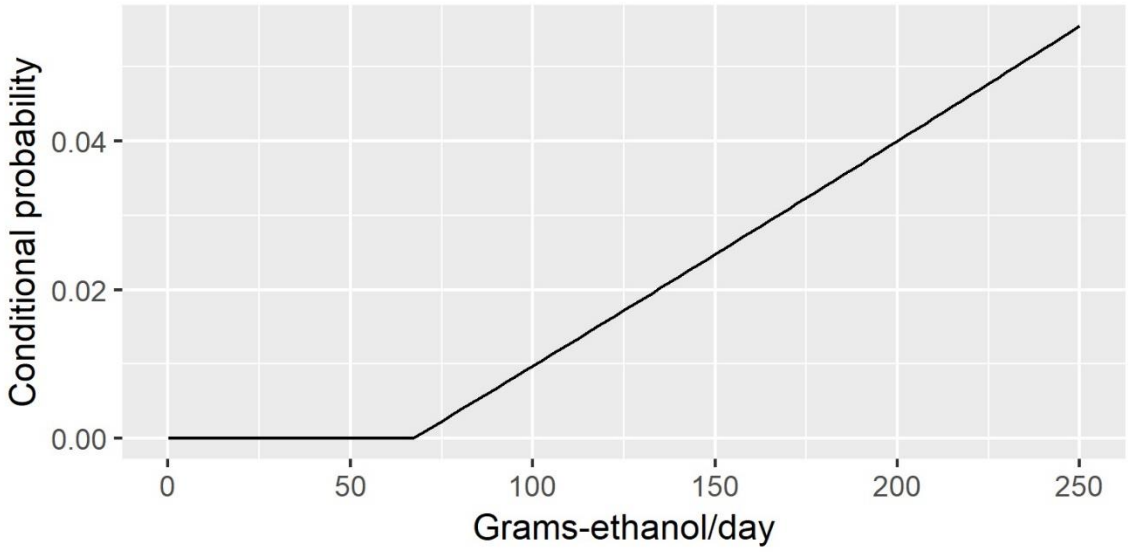

Figure 1: The absolute risk function of the form in Equation (16) with threshold 67.25 and calibrated k of 2.96e-4.

Finally, we estimate the proportion of hospitalizations due to alcohol psychoses within a particular population subgroup. The accumulation of disease up to a mean daily exposure of 80.7 grams-ethanol is approximately 5.4%, or 318 hospitalizations. This leaves the vast majority of disease (94.6%, or 3,547 hospitalizations), which is suffered by only 25% of this drinking population.

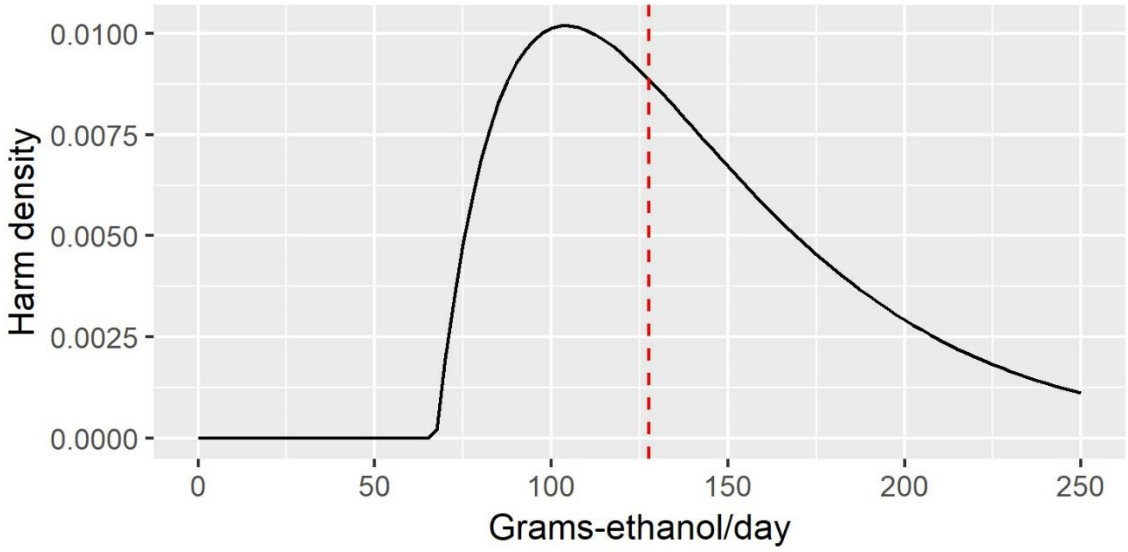

Figure 2: The probability density function of the form in Equation (20). Median value of approximately 127:7g/day indicated by dashed line.

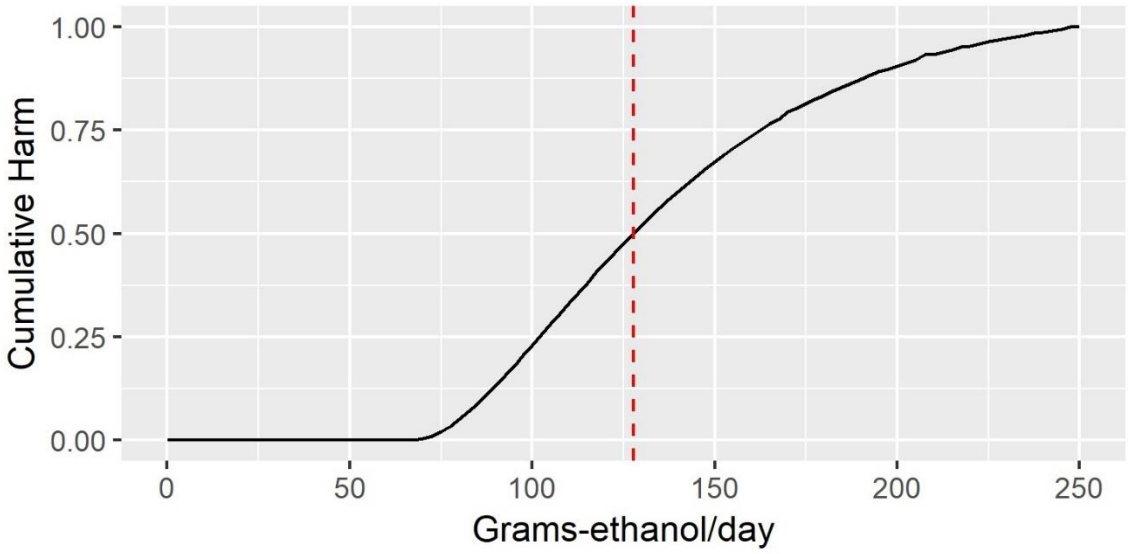

Figure 3: Cumulative distribution function of the density function in Figure 2. Median value of approximately 127.7g/day indicated by dashed line.

Discussion

The approach described in this paper provides a clear, reproducible method for researchers to estimate absolute risk curves of conditions that are wholly-attributable to exposure to a particular epidemiological risk factor. Further, the method can be used to estimate the portion of disease suffered by particular exposure groups and the change in the burden of disease, which would result from a change in population exposure. These methods can be used in isolation or incorporated into more extensive models, similar to InterMAHP or SAPM, to produce more comprehensive estimates of the burden of disease of the chosen risk factor and/or the extent to which this may be modified by changes in exposure. We've presented a form for an absolute risk function in the field of alcohol epidemiology, but how well does this generalize? Outside of alcohol epidemiology, at minimum the following relation still holds:

$$\int_{\mathbb{R}} p(x)dF(x) = \frac{N}{|S|} \tag{27}$$

Assuming one knows the incidence $N/|S|$ and the exposure distribution $F(x)$, the main obstruction to producing an absolute risk function $p(x)$ is assuming an appropriate form for

the function. For example, the main considerations in setting the form for the alcohol psychoses absolute risk function above were:

1. **Plausibility.** The relationship prescribed by the function $p(x; k, t)$ is plausible due to its similarity to relative risk functions for partially alcohol attributable conditions[4] and due to the restrictions imposed on diagnosis of alcohol psychoses.[12]
2. **Existence.** Set

$$I(k) = \int_{x_0}^{x_1} p(x; k, t) dF(x; \mu) \tag{28}$$

Observed values of N tend to be far smaller than $|S|$.[24] Existence of a positive solution to $I(k) = \frac{N}{|S|}$ is guaranteed because the form chosen for $p(x; k, t)$ guarantees $I(0) = 0$, $I(k) > 0$ for $k > 0$, and $\lim_{k \to \infty} I(k) \to \infty$.

3. **Uniqueness.** The function $I(k)$ is strictly increasing in k, i.e. $\frac{dI}{dk} > 0$ for all k, so a solution is guaranteed to be unique if it exists.

Many problems and considerations in choosing an epidemiological model are discussed in Greenland[25] and will not be rehashed here. Some concerns are independent of methodology but essential to application, and we address these now.

There are two temporality concerns. The first regards the period of observation used to collect event data, and the second regards application of this methodology to change-in-exposure scenarios.

1. The event count in the worked example was aggregated over the calendar year 2014, so the calibrated absolute risk function as presented must be interpreted as the yearly conditional probability of an event given average daily exposure over the course of a year. If aggregation of events is over a different period or represented as a rate, the methodology still applies, but the result must be interpreted in kind.
2. A change in exposure due to a change in policy occurs over time. Moreover, there may be a fundamental lag between the change in exposure and the change in risk. Modeling that attempts to estimate future changes in disease incidence due to changes in policy must take these lag effects into account, cf. Holmes et al.[26]

In the general mathematical derivation, we made simplifying assumptions regarding the form of the absolute risk function, and we optimized over a single parameter. These assumptions serve the context of alcohol epidemiology where (i) there are several exposure models to choose from and (ii) it is reasonable to assume that risk of disease increases exponentially with respect to exposure. In practice, weaker assumptions are possible and

more parameters may be optimized over as long as existence and uniqueness of the solution can still be demonstrated.

This article was motivated by asking 'How does disease caused by wholly-attributable conditions distribute over levels of exposure?' and 'How does total exposure affect rates of wholly-attributable conditions?' Whilst we provide a worked example to demonstrate the methodology with alcohol-related applications in mind, this method is readily generalizable to other contexts and may be of use to those working in the fields of other health behaviors such as tobacco or diet.

1. Eide GE and Gefeller O. Sequential and average attributable fractions as aids in the selection of preventive strategies. *Journal of clinical epidemiology* 1995; 48: 645-655.
2. Eide GE and Heuch I. Attributable fractions: fundamental concepts and their visualization. *Statistical Methods in Medical Research* 2001; 10: 159-193.
3. Meier P, Brennan A and Purshouse R. Modelling the potential impact of pricing and promotion policies for alcohol in England: results from the Sheffield alcohol policy model version 2008 (1-1). *Liverpool, UK: National Collaborating Centre for Drug Prevention* 2008.
4. Sherk A, Stockwell T, Rehm J, et al. *The International Model of Alcohol Harms and Policies (InterMAHP): A comprehensive guide to the estimation of alcohol-attributable morbidity and mortality. Version 1.0*. 2017. Canadian Institute for Substance Use Research.
5. WHO. *Global status report on alcohol and health*. Report no. 924156475X, 2014. Geneva, Switzerland: World Health Organization.
6. Levin ML. The occurrence of lung cancer in man. *Acta Unio int contra cancrum* 1953; 9: 531-941.
7. Eide GE and Heuch I. Average attributable fractions: a coherent theory for apportioning excess risk to individual risk factors and subpopulations. *Biometrical journal* 2006; 48: 820-837.

8. Griswold MG, Fullman N, Hawley C, et al. Alcohol use and burden for 195 countries and territories, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *The Lancet* 2018; 392: 1015-1035.
9. Kehoe T, Gmel G, Shield KD, et al. Determining the best population-level alcohol consumption model and its impact on estimates of alcohol-attributable harms. *Population health metrics* 2012; 10: 1.
10. Canadian Substance Use Costs and Harms Scientific Working Group. *Canadian substance use costs and harms (2007-2014). (Prepared by the Canadian Institute for Substance Use Research and the Canadian Centre on Substance Use and Addiction.)*. 2018. Ottawa, ON: Canadian Centre on Substance Use and Addiction.
11. Brennan A, Meier P, Purshouse R, et al. The Sheffield alcohol policy model–a mathematical description. *Health economics* 2015; 24: 1368-1388.
12. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders (DSM-5)*. American Psychiatric Pub, 2013.
13. Royston P and Sauerbrei W. *Multivariable model-building: a pragmatic approach to regression anaylsis based on fractional polynomials for modelling continuous variables*. John Wiley and Sons, 2008.
14. Angus C, Henney M, Webster L, et al. Alcohol-attributable Diseases and Dose-response Curves for the Sheffield Alcohol Policy Model Version 4.0, https://doi.org/10.15131/shef.data.6819689.v1 (2018).
15. Rehm J, Sherk A, Shield KD, et al. *Risk relations between alcohol use and non-injury causes of death*. September 2017 2017. Toronto, Canada: Centre for Addiction and Mental Health.
16. Russ S. A translation of Bolzano's paper on the intermediate value theorem. *Historia Mathematica* 1980; 7: 156-185.
17. Stockwell T, Pauly B, Chow C, et al. Does managing the consumption of people with severe alcohol dependence reduce harm? A comparison of participants in six Canadian managed alcohol programs with locally recruited controls. *Drug and alcohol review* 2017.
18. Rehm J. Alcohol consumption and mortality. What do we know and where should we go? *Addiction* 2000; 95: 989.
19. Rehm J, Gmel G, Sempos C, et al. Alcohol-attributable mortality and morbidity. *Alcohol Res Health* 2003; 27: 39-51.
20. Rothman KJ, Greenland S and Lash TL. *Modern epidemiology*. Lippincott Williams and Wilkins, 2008.
21. Powell MJ. A direct search optimization method that models the objective and constraint functions by linear interpolation. *Advances in optimization and numerical analysis*. Springer, 1994, pp.51-67.

22. Johnson S. The NLopt nonlinear-optimization package, http://ab-initio.mit.edu/nlopt (2014).
23. Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York, 2016.
24. Engelhard C, Touquet G, Tansens A, et al. Alcohol-induced psychotic disorder: a systematic literature review. *Tijdschrift voor psychiatrie* 2015; 57: 192.
25. Greenland S. Modeling and variable selection in epidemiologic analysis. *American journal of public health* 1989; 79: 340-349.
26. Holmes J, Meier PS, Booth A, et al. The temporal relationship between per capita alcohol consumption and harm: a systematic review of time lag specifications in aggregate time series analyses. *Drug and alcohol dependence* 2012; 123: 7-14.