European Journal of Epidemiology (2018) 33:847–858
https://doi.org/10.1007/s10654-018-0375-y

METHODS

Estimation of causal effect measures with the R-package stdReg

Arvid Sjölander[1]

Received: 20 September 2017 / Accepted: 5 March 2018 / Published online: 14 March 2018

Abstract

Measures of causal effects play a central role in epidemiology. A wide range of measures exist, which are designed to give relevant answers to substantive epidemiological research questions. However, due to mathematical convenience and software limitations most studies only report odds ratios for binary outcomes and hazard ratios for time-to-event outcomes. In this paper we show how logistic regression models and Cox proportional hazards regression models can be used to estimate a wide range of causal effect measures, with the R-package stdReg. For illustration we focus on the attributable fraction, the number needed to treat and the relative excess risk due to interaction. We use two publicly available data sets, so that the reader can easily replicate and elaborate on the analyses. The first dataset includes information on 487 births among 188 women, and the second dataset includes information on 2982 women diagnosed with primary breast cancer.

Keywords Attributable fraction · Causal effect · Cox proportional hazards regression · Logistic regression · Number needed to treat · Relative excess risk due to interaction

Introduction

A common aim of epidemiologic research is to estimate the causal effect of an exposure on an outcome. To control for potential confounders it is common to use logistic regression models for binary outcomes and Cox proportional hazards (PH) regression models for time-to-event outcomes. Methods for fitting these models are implemented in all major statistical software, which makes them easily accessible to applied epidemiologists.

Logistic regression models and Cox PH regression models are parametrized in terms of log odds ratios and log hazard ratios, respectively. These parameters are mathematically convenient since they are unrestricted, i.e. they can have values anywhere in the range $(-\infty, \infty)$. Thus, logistic regression models and Cox PH regression models will never produce parameter estimates that are outside the supported range. Arguably though, the log odds ratio and the hazard ratio are usually not the most intuitive or relevant measures of the exposure effect. Both are often misinterpreted, in particular among applied epidemiologists and clinicians without statistical training [1–3], and neither is directly informative about the public health impact of the exposure, since they do not take the exposure prevalence into account. Furthermore, when assessing interactions between two exposures, it has been argued that the risk differences are more appropriate than odds ratios or hazard ratios [4].

There exist many other suggestions for causal effect measures in the literature, which are supposed give more relevant answers to substantive epidemiological research questions [5]. For instance, the risk difference and the survival difference are relatively easy to interpret and communicate to non-statisticians. The attributable fraction (AF) and the number needed to treat (NNT) are directly informative about the public health impact of the exposure/treatment. The relative excess risk due to interaction (RERI), the synergy index (S) and the attributable proportion due to interaction (AP) measure the amount of interaction between two exposures on the additive scale. Methods have been developed to estimate these (and related) measures from logistic regression models and Cox PH regression models (see Rothman et al. [5] and the references therein), and several of these methods have been

✉ Arvid Sjölander
arvid.sjolander@ki.se

[1] Karolinska Institute, Nobels väg 12 A, 171 77 Stockholm, Sweden

implemented in statistical software. However, these implementations are typically scattered across various packages and commands, with diverse syntax and functionality.

The aim of this paper is to show how one single R-package, `stdReg` [6], can be used to estimate a wide range of causal effect measures, including all those mentioned above. Briefly, the package uses ‘regression standardization’ to estimate standardized probabilities from a fitted regression model. We described this procedure in a recent paper [7]. In the current paper we show how the standardized probabilities can subsequently be contrasted to form various measures of the exposure effect. A few simple effect measures are already implemented in the `stdReg` package, such as the risk difference and the risk ratio. However, due to the wide range of existing measures, and the creativity among epidemiologists to invent new measures, it would be virtually impossible to implement them all. Rather, we show in this paper how an analyst may obtain, with a minimal amount of programming, a desired effect measure from the standardized probabilities estimated by `stdReg`. We also show how the delta method can be used to estimate the variance and construct confidence intervals for the desired effect measure.

To our knowledge, there are currently only two packages in `R` that carry out regression standardization; `stdReg` and `margins`. However, `margins` is restricted to linear effects (e.g. differences) and cannot be used to compute other measures of causal effects. Furthermore, `margins` is restricted to generalized linear models and does not support models for time-to-event data.

The paper is organized as follows. In “Regression standardization” section we briefly review the method of regression standardization; we refer to Sjölander [7] for a more detailed account. In the subsequent sections we show how the `stdReg` package can be used to estimate various effect measures. For illustration we focus on the AF (“The AF” section), the NNT (“The NNT” section) and the RERI (“The RERI” section). We use two publically available datasets, so that the reader can easily replicate and elaborate on the analyses. The first dataset includes information on 487 births among 188 women, and the second dataset includes information on 2982 women diagnosed with primary breast cancer. These datasets are borrowed from the `AF` package [8]; the help files for this package provide a thorough description of the data. We assume that the reader has some experience with `R` programming, and with the `glm` function from the `stats` package and the `coxph` function from the `survival` package.

Regression standardization

Let X and Y be the exposure and outcome of interest, respectively. For the moment we assume that the outcome is binary (0/1), but we do not make any particular assumption about the exposure. Let Y_x be the potential outcome [9, 10] for a given subject, if that subject would be exposed to the fixed level $X = x$. Finally, let $p(Y_x = 1)$ be the counterfactual probability of the outcome if all subjects in the population would hypothetically be exposed to $X = x$. We here use the term ‘population’ in the usual epidemiological sense, i.e. as referring to a hypothetical, infinitely large superpopulation, from which the observed sample was drawn [5].

Counterfactual probabilities are cornerstones in the modern theory of causal inference, and can be used to define a wide range of effect measures. For instance, when the exposure is binary the causal risk difference and risk ratio are defined as $p(Y_1 = 1) - p(Y_0 = 1)$ and $p(Y_1 = 1)/p(Y_0 = 1)$, respectively. We will use counterfactual probabilities to define the AF (“The AF” section), the NNT (“The NNT” section) and the RERI (“The RERI” section).

To estimate $p(Y_x = 1)$ without bias from observational (i.e. non-randomized) data, it is necessary to control for confounding. Let Z be a set of measured confounders and let $p(Y|X, Z)$ be the conditional distribution of Y, given X and Z. If Z is sufficient for confounding control, then

$$p(Y_x = 1) = E\{p(Y = 1|X = x, Z)\}, \tag{1}$$

where the expectation on the right-hand side is taken over the population distribution of Z [10]. Regression standardization attempts to estimate $p(Y_x = 1)$ by estimating the right-hand side of (1), as follows. In a first step, a regression model for $p(Y|X, Z)$ is fitted to the observed data, e.g. a logistic regression model. In a second step, the fitted model is used to estimate $p(Y = 1|X = x, Z)$ for the fixed level $X = x$, and for each observed level of Z in the dataset. In a third step, these estimates are averaged. In concise notation we thus have that

$$\hat{p}(Y_x = 1) = \frac{\sum_{i=1}^{n} \hat{p}(Y = 1|X = x, Z_i)}{n}, \tag{2}$$

where Z_i is the observed level of Z for subject i, $i = 1, \ldots, n$, and $\hat{p}(Y = 1|X = x, Z_i)$ is the estimate of $p(Y = 1|X = x, Z_i)$ obtained from the fitted regression model.

Using the same fitted model from the first step, the second and third steps above are repeated for different values of x. Once the counterfactual probabilities $p(Y_x = 1)$ have been estimated for different values of x, we may contrast these to obtain desired measures of the exposure effect. When X is categorical with few levels (e.g. binary), it is possible to estimate $p(Y_x = 1)$ for all possible values of x. When X is multilevel categorical or continuous, one would typically have to focus on a few selected values of interest.

In many scenarios, the outcome is a time-to-event, e.g. time to death or time to relapse for cancer patients. To have a simple and uniform notation we then let $Y(t)$ be the indicator of having the event before a fixed time-point t, e.g. 5 years from birth or from age at cancer diagnosis. Thus, $p\{Y_x(t) = 1\}$ is the counterfactual probability of having the event before time t if all subjects in the population would hypothetically be exposed to $X = x$. With these definitions, regressions standardization proceeds as outlined above, for any fixed time-point t. However, when the underlying outcome is a time-to-event, it is more appropriate to use a Cox PH regression model than a logistic regression model. First, because the Cox PH regression model deals more naturally with censoring than the logistic regression model. Second, because in the analysis one may want to consider a range of different time-points. If a logistic regression model is used, then a new model has to be fitted for each time-point. In contrast, one single Cox PH regression model may be used to estimate $p\{Y(t) = 1|X = x, Z\}$ for arbitrary values of t. For details on this estimation procedure we refer to Sjölander [7].

Typically, we want to have the variance of the estimated effect, so that we can, for instance, construct confidence intervals and hypothesis tests. The asymptotic variance can be obtained with the delta method [11], as follows. Let $\mathbf{p}$ be the vector of counterfactual probabilities and let $g(\mathbf{p})$ be the desired effect measure. For instance, when X is binary and we wish to estimate the causal risk difference we have that $\mathbf{p} = \{p(Y_1 = 1), p(Y_0 = 1)\}$ and $g(\mathbf{p}) = p(Y_1 = 1) - p(Y_0 = 1)$. Let $\hat{\mathbf{p}}$ be the estimate of $\mathbf{p}$ and let $\mathrm{var}(\hat{\mathbf{p}})$ be the variance-covariance matrix for $\hat{\mathbf{p}}$. Let $\widehat{\mathrm{var}}(\hat{\mathbf{p}})$ be an estimate of $\mathrm{var}(\hat{\mathbf{p}})$. Using the delta method it can be shown that the estimated effect $g(\hat{\mathbf{p}})$ has an asymptotic normal distribution, with variance equal to

$$\mathrm{var}\{g(\hat{\mathbf{p}})\} = \frac{\partial g(\mathbf{p})}{\partial \mathbf{p}} \mathrm{var}(\hat{\mathbf{p}}) \frac{\partial g(\mathbf{p})}{\partial \mathbf{p}^T}. \tag{3}$$

An estimate of the variance, $\widehat{\mathrm{var}}\{g(\hat{\mathbf{p}})\}$, is obtained by replacing $\mathbf{p}$ and $\mathrm{var}(\hat{\mathbf{p}})$ in (3) with $\hat{\mathbf{p}}$ and $\widehat{\mathrm{var}}(\hat{\mathbf{p}})$, respectively. The estimated variance can, for instance, be used to construct a standard Wald-type 95% confidence interval for $g(\mathbf{p})$ on the form $g(\hat{\mathbf{p}}) \pm 1.96\sqrt{\widehat{\mathrm{var}}\{g(\hat{\mathbf{p}})\}}$. For parameters that are restricted to positive values, such as the NNT ("The NNT" section), it is desirable to ensure that the confidence interval is restricted to positive values as well. This may be accomplished by first computing the standard Wald confidence interval for the logarithm of the parameter, then back-transforming to the original scale, which gives the confidence interval $\exp[\log\{g(\hat{\mathbf{p}})\} \pm 1.96\sqrt{\widehat{\mathrm{var}}\{g(\hat{\mathbf{p}})\}}/g(\hat{\mathbf{p}})]$. This transformed confidence interval typically has a coverage probability closer to the nominal level than the untransformed confidence interval.

We end this section by emphasizing that, in real observational studies, it would rarely be possible to measure all confounders for the exposure-outcome association. This means that it is rarely possible to estimate the counterfactual probabilities such as $p(Y_x = 1)$ without bias. However, if the study is well designed, and potential confounders have been carefully selected, then the bias may be relatively small.

The AF

Definition

The AF measures the proportion of outcome events that would be prevented if the exposure was hypothetically eliminated from the population. For binary outcomes the AF is defined as

$$\mathrm{AF} = 1 - \frac{p(Y_0 = 1)}{p(Y = 1)}, \tag{4}$$

see, for instance, Sjölander [12]. Here, $p(Y = 1)$ is the factual probability (prevalence) of the outcome in the population of interest, and $p(Y_0 = 1)$ is the counterfactual probability of the outcome if the exposure was eliminated (set to 0). For instance, if $p(Y = 1) = 0.05$ and $p(Y_0 = 1) = 0.01$, then an elimination of the exposure would prevent $1 - 0.01/0.05 = 80\%$ of all outcomes. We note that the definition in (4) does not assume that the exposure is binary per se, but it does assume that there is a natural 'zero-level', at which the exposure is completely absent.

For time-to-event outcomes, the AF is defined as in (4), but with Y and Y_0 replaced by $Y(t)$ and $Y_0(t)$, respectively, for a given t [13, 14]. Thus, the AF measures the proportion of outcome events that would be prevented before time t if the exposure was eliminated at baseline ($t = 0$). For many time-to-event outcomes, the AF is a decreasing function of t. For instance, if the outcome is death and t is 200 years from birth, then the AF is 0, since no realistic exposure intervention can prevent a subject from dying within 200 years from birth.

For details on model-based estimation of the AF we refer to Sturmans et al. [15], Deubner et al. [16], Greenland and Drescher [17], Chen et al. [13, 14], Sjölander and Vansteelandt [18, 19].

Estimation with logistic regression models

We illustrate the methods with the dataset `clslowbwt` from the `AF` package. This dataset includes information on 487 births among 188 women. We will use the variables `lbw` (a binary indicator of whether the newborn child has low birthweight, defined as a birthweight smaller or equal to 2500 g), `smoker` (a binary indicator of whether the mother smoked during pregnancy), `race` (race of the mother, coded as `1. White`, `2. Black` or `3. Other`), `age` (age of the mother), and `id` (a unique identification number for each mother). Our aim is to estimate the proportion of low birthweights that would be prevented if nobody would smoke during pregnancy. We will control for mother's race and age in the analysis.

The first step is to fit a logistic regression model that relates the outcome (low birthweight) to the exposure (smoking) and measured confounders (race and age). This is done by

```
> fit <- glm(formula=lbw~smoker+race+age, family="binomial",
  data=clslowbwt)
```

which stores the fitted model into an object called `fit`. The results are summarized, without 'significance stars', by

```
> options(show.signif.stars=FALSE)
> summary(fit)
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.35946    0.53281  -2.551  0.01073
smoker        0.60080    0.21693   2.770  0.00561
race2. Black -0.85852    0.32331  -2.655  0.00792
race3. Other -0.70449    0.24624  -2.861  0.00422
age           0.02399    0.01785   1.344  0.17900
```

We observe that both smoking and race are significantly (at 5% significance level) associated with low birthweight, whereas age is not.

The next step is to estimate standardized probabilities. This is done with the `stdGlm` function in the `stdReg` package, by

```
> fit.std <- stdGlm(fit=fit, data=clslowbwt, X="smoker",
  x=c(NA,0), clusterid="id")
```

which stores the standardization results into an object called `fit.std`. The `fit` argument specifies a fitted generalized linear (e.g. logistic) model and the `data` argument specifies the data frame used to fit the model. The `X` argument specifies the name of the exposure variable and the `x` argument specifies fixed exposure levels for which we wish to estimate the counterfactual probability $p(Y_x = 1)$. We here use a trick; by setting `x` to `NA` each subject retains her own factual exposure level, so that the factual outcome probability $p(Y = 1)$ is estimated. This is useful, since the definition of AF in (4) involves $p(Y = 1)$. By setting `x` to 0, the counterfactual probability $p(Y_0 = 1)$ is estimated. The argument `clusterid` specifies a variable that defines clusters in the data, e.g. mothers with multiple births. This has no effect on the estimates, but makes the `stdGlm` function use the 'sandwich formula' [20] to correct the variance of the estimates for within-cluster dependencies. Summarizing the results gives

```
> summary(fit.std)
     Estimate Std. Error lower 0.95 upper 0.95
<NA>    0.310     0.0301      0.251      0.369
0       0.257     0.0391      0.181      0.334
```

Thus, the factual probability of low birthweight is estimated to be 31.0%, and the counterfactual probability, had nobody smoked during pregnancy, is estimated to be 25.7%.

The `fit.std` object has (among other things) an element called `est`, which is a vector containing the estimated standardized probabilities in the order specified by the `x` argument, and an element called `vcov`, which is the (estimated) variance-covariance matrix of the estimates. We now define a function that uses `est` to estimate the AF:

```
> AF <- function(est){
  p <- est[1]
  p0 <- est[2]
  af <- 1-p0/p
  return(af)
}
```

Using this function gives

```
> AFest <- AF(fit.std$est)
> AFest
[1] 0.1697446
```

Hence, the analysis suggests that around 17% of all low birthweights would be prevented if nobody would smoke during pregnancy. We emphasize that this causal interpretation crucially hinges on race and age being sufficient for confounding control.

The stdReg package has a function confint, which uses the delta method to compute a Wald-type confidence interval for a parameter specified as a function of standardized probabilities. Using this function gives

```
> confint(object=fit.std, fun=AF, level=0.95)
            lower     upper
[1,] -0.008477017 0.3479662
```

The optional argument level controls the coverage probability of the interval, and defaults to 0.95. The 95% confidence interval is quite wide, and suggests that the true AF may be as high as 34.8%. Furthermore, it includes the value 0, so at 5% significance level we cannot reject the null hypothesis that low birthweight is not prevented by eliminating smoking.

Estimation with Cox PH regression models

We illustrate the methods with the dataset rott2 from the AF package. This dataset includes information on 2982 women diagnosed with primary breast cancer from the Rotterdam tumor bank in the Netherlands. Follow-up is measured in months since diagnosis, and ranges from 1 to 231 months. We will use the variables rf (follow-up time, measured in months, since diagnosis), rfi (an indicator of whether the patient died or had a relapse before censoring), chemo (an indicator of whether the patient received chemotherapy, coded as “yes” or “no”), age (patient’s age at surgery), meno (menopausal status, coded as 0 for pre and 1 for post), size (tumor size in three categories: “ < = 20mm”, “ > 20 − 50mmm” and “ > 50mm”), grade (tumor grade; 2 or 3), nodes (the number of positive lymph nodes, ranging from 0 to 34), pr (progesterone receptors, fmol/l), and er (oestrogen receptors, fmol/l). Chemotherapy is supposed to give the patients a better prognosis, e.g. to prevent deaths and relapses. Our aim is to estimate the proportion of deaths and relapses that would be prevented if all patients received chemotherapy. We will control for age, menopausal status, tumor size, tumor grade, lymph nodes, progesterone and oestrogen receptors in the analysis.

To be consistent with the notation in “Definition” section, where we used values 0 and 1 for ‘unexposed’ and ‘exposed’, respectively, we first define the binary exposure variable

```
> rott2$nochemo <- as.numeric(rott2$chemo=="no")
```

We fit a Cox PH regression model that relates the outcome (time to death/relapse) to the exposure (absence of chemotherapy) and measured confounders (age, menopausal status, tumor size, tumor grade, lymph nodes, progesterone and oestrogen receptors). This is done by

```
> fit <- coxph(Surv(rf,rfi)~nochemo+age+meno+size+factor(grade)+
  I(exp(-0.12*nodes))+pr+er, data=rott2, method="breslow")
```

We here used the transformation $\exp(-0.12 * \texttt{nodes})$, since previous authors have shown that this gives a better model fit [21]. We obtain the results

```
> summary(fit)
                             coef  exp(coef)   se(coef)       z  Pr(>|z|)
nochemo                 2.814e-01  1.325e+00  7.225e-02   3.895  9.82e-05
age                    -1.597e-02  9.842e-01  3.521e-03  -4.535  5.76e-06
menopre                -1.322e-01  8.762e-01  8.948e-02  -1.478     0.140
size>20-50mmm           2.886e-01  1.335e+00  5.855e-02   4.929  8.25e-07
size>50mm               4.805e-01  1.617e+00  8.910e-02   5.394  6.91e-08
factor(grade)3          3.473e-01  1.415e+00  6.499e-02   5.343  9.13e-08
I(exp(-0.12 * nodes))  -1.853e+00  1.568e-01  9.794e-02 -18.918   < 2e-16
pr                     -9.554e-05  9.999e-01  1.059e-04  -0.902     0.367
er                     -5.079e-05  9.999e-01  1.041e-04  -0.488     0.626
```

We observe that the absence of chemotherapy is indeed associated with a higher rate of death/relapse. We next use the fitted model to estimate standardized probabilities. This is done with the stdCoxph function in the stdReg package, by

```
> fit.std <- stdCoxph(fit=fit, data=rott2, X="nochemo", t=10:60,
  x=c(NA,0))
```

The syntax for stdCoxph is similar to the syntax for stdGlm. However, stdCoxph has an additional argument t, which specifies the time points at which to carry out the standardization; we here consider a sequence of 10 through 60 months after diagnosis. summary(fit.std) produces a long output displaying the results for each of these time points separately (not shown here).

Similar to stdGlm, the stdCoxph function creates an object with elements est and vcov. However, when created by the stdCoxph function, the element est is a matrix containing estimated standardized survival probabilities, $1-\hat{p}\{Y_x(t)=1\}$, for the specified values of t (in rows) and x (in columns). The element vcov is a list containing the variance–covariance matrices at each value of t. We now define a function that uses est to estimate the AF:

```
> AF <- function(est){
    p <- 1-est[, 1]
    p0 <- 1-est[, 2]
    af <- 1-p0/p
    return(af)
}
```

We use this function, and obtain point-wise 95% confidence intervals, by

```
> AFest <- AF(fit.std$est)
> AFci <- confint(object=fit.std, fun=AF)
```

We plot the estimated AF and the confidence intervals by

```
> plot(t, AFest, type="l", ylab="AF",
  xlab="time since diagnosis (months)", ylim=c(0.05,0.3))
> matlines(t, AFci, lty="dashed", col="black")
```

The resulting plot is displayed in Fig. 1. We observe that the AF declines with time, from 18.4% to 14.2%, at 10 and 60 months after diagnosis, respectively. Thus, the analysis suggests that 18.4% of all deaths/relapses that occurred before 10 months after diagnosis would have been prevented if all patients had been given chemotherapy. When considering a time window up to 60 months after diagnosis, only 14.2% would have been prevented. The point-wise 95% confidence intervals exclude the value 0 everywhere in the time range, which means that we have statistically significant (at 5% significance level) evidence for a preventative effect of chemotherapy everywhere in the time range.

The NNT

Definition

The NNT is supposed to measure the average number of subjects that would have to be treated, among those that are factually untreated, to prevent one unfavorable outcome event. In the literature, the NNT is usually defined for binary outcomes and binary treatments as

$$\frac{1}{p(Y=1|X=0)-p(Y=1|X=1)}, \tag{5}$$

where $p(Y=1|X=0)$ and $p(Y=1|X=1)$ are the probabilities of the outcome among the untreated and treated, respectively [22]. However, this definition implicitly assumes that there is no confounding, and is thus in practice restricted to randomized control trials. In the presence of confounding, $p(Y=1|X=0)$ may very well be much larger than $p(Y=1|X=1)$ even in the absence of a causal treatment effect, thus falsely implying a small NNT.

To derive a causal definition of the NNT, let N be a fixed number of untreated subjects. Among these, $Np(Y=1|X=0)$ subjects will on average have the outcome. Suppose now that we would treat all N subjects. Under this counterfactual scenario, the probability of the outcome is $p(Y_1=1|X=0)$; that is, the probability of the outcome if everybody would be treated among those that are factually untreated. Thus, among those N subjects that are factually untreated, $Np(Y_1=1|X=0)$ subjects would on average have the outcome if all were treated. Setting $Np(Y=1|X=0)-Np(Y_1=1|X=0)=1$ and solving for N gives

$$\text{NNT}=\frac{1}{p(Y=1|X=0)-p(Y_1=1|X=0)}. \tag{6}$$

In the absence of confounding, the potential outcome Y_1 is equally distributed among treated and untreated, so that $p(Y_1=1|X=0)=p(Y_1=1|X=1)=p(Y=1|X=1)$,

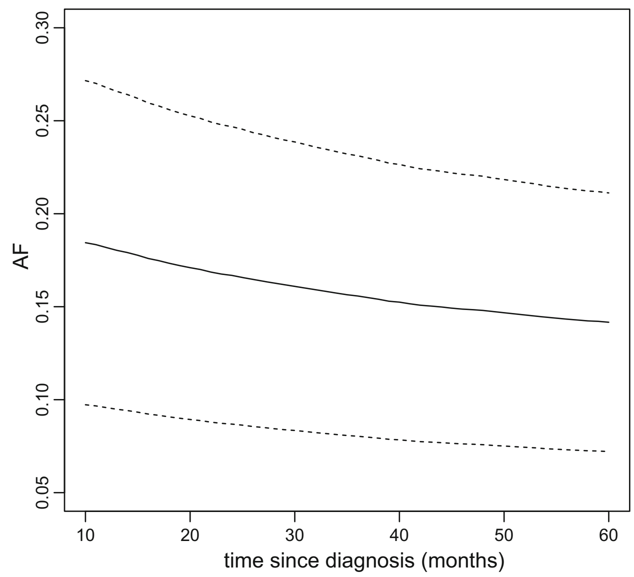

Fig. 1 The estimated AF (solid line) as a function of time since diagnosis, together with point-wise 95% confidence intervals (dashed lines)

and the causal definition in (6) simplifies to the usual definition (6) in the literature.

Estimation of the NNT requires a minor deviation from the recipe outlined in "Regression standardization" section. Because the counterfactual probability $p(Y_1 = 1|X = 0)$ only applies to the subset of the population with $X = 0$, we have to replace the averages on the right-hand sides of (1) and (2) with the averages among this subset. We thus have that

$$p(Y_1 = 1|X = 0) = E\{p(Y = 1|X = 1, Z)|X = 0\}, \tag{7}$$

and

$$\hat{p}(Y_1 = 1|X = 0) = \frac{\sum_{i=1}^{n}(1 - X_i)\hat{p}(Y = 1|X = 1, Z_i)}{\sum_{i=1}^{n}(1 - X_i)}. \tag{8}$$

For time-to-event outcomes, we define the NNT as in (6), but with Y and Y_1 replaced with $Y(t)$ and $Y_1(t)$, respectively. Thus, the NNT measures the average number of subjects that would have had to be treated at baseline ($t = 0$), among those that were factually untreated, in order to prevent one unfavorable outcome event before time t.

For details on model-based estimation of the NNT we refer to Bender et al. [23] and Laubender and Bender [24].

Estimation with logistic regression models

We illustrate the methods with the clslowbwt dataset. We define the 'treatment' as absence of the smoking during pregnancy. With this definition, the NNT is interpreted as the average number of smokers that would have to refrain from smoking during pregnancy, in order to prevent one low birthweight.

To be consistent with the notation in "Definition" section, where we used values 0 and 1 for untreated and treated, respectively, we first define the treatment variable

```
> clslowbwt$nosmoke <- 1-clslowbwt$smoker
```

We fit the logistic regression model

```
> fit <- glm(formula=lbw~nosmoke+race+age, family="binomial",
  data=clslowbwt)
```

We use the fitted model to estimate standardized probabilities:

```
> fit.std <- stdGlm(fit=fit, data=clslowbwt, X="nosmoke",
  x=c(NA,1), clusterid="id", subsetnew=nosmoke==0)
```

The subsetnew argument specifies a subset of observations to be used when estimating the standardized probabilities. We note that, although we have not used it here, the glm function has a subset argument that allows for subsetting when fitting the regression model. This is different from subsetting when standardizing; thus we have used the term 'subsetnew' for the latter. As argued in "Definition" section we only wish to standardize over the untreated (i.e. the smokers) when estimating the NNT; this is achieved by setting subsetnew=nosmoke==0. We note that, by setting x to NA within this subset, we estimate the factual outcome probability among the untreated; $p(Y = 1|X = 0)$. Summarizing the results gives

```
> summary(fit.std)
     Estimate Std. Error lower 0.95 upper 0.95
<NA>    0.415     0.0486       0.32      0.511
1       0.284     0.0480       0.19      0.378
```

Thus, the factual probability of low birthweight is estimated to be 41.5%, and the counterfactual probability, had nobody smoked, is estimated to be 28.4%. We emphasize that these figures only apply to those who factually did smoke. In contrast, the figures obtained in "Estimation with logistic regression models" section by `summary(fit.std)` apply to the whole population (i.e. both smokers and non-smokers).

We define a function that estimates the NNT

```
> NNT <- function(est){
  p <- est[1]
  p1 <- est[2]
  nnt <- 1/(p-p1)
  return(nnt)
}
```

Using the function gives

```
> NNTest <- NNT(fit.std$est)
> NNTest
[1] 7.607845
```

Hence, the analysis suggests that smoking during pregnancy must be prevented for around 7.6 women, in order to prevent one low birthweight. A 95% confidence interval is obtained by

```
> confint(object=fit.std, fun=NNT, type="log")
        lower    upper
[1,] 2.724625 21.24303
```

Setting `type=‘‘log’’` forces `confint` to first compute a confidence interval for the logarithm of the NNT, then backtransforming to the original scale. This transformed confidence interval only includes positive values, as it should, but it is quite wide and suggests that the true NNT may be as high as 21.2.

Estimation with Cox PH regression models

We illustrate the methods with the `rott2` dataset. We aim to estimate the average number of patients that would have had to be treated with chemotherapy, among those that were factually untreated, in order to prevent one death/relapse before a specific time-point *t*.

We first fit the Cox PH regression model

```
> fit <- coxph(Surv(rf,rfi)~chemo+age+meno+size+factor(grade)+
  I(exp(-0.12*nodes))+pr+er, data=rott2, method="breslow")
```

We use the fitted model to estimate standardized probabilities among those that are untreated:

```
> fit.std <- stdCoxph(fit=fit, data=rott2, X="chemo", t=10:60,
  x=c(NA,"yes"), subsetnew=chemo=="no")
```

We define a function that estimates the NNT:

```
> NNT <- function(est){
  p <- 1-est[, 1]
  p1 <- 1-est[, 2]
  nnt <- 1/(p-p1)
  return(nnt)
}
```

We plot the estimated NNT together with point-wise 95% confidence intervals

```
> NNTest <- NNT(fit.std$est)
> NNTci <- confint(object=fit.std, fun=NNT)
> plot(t, NNTest, type="l", ylab="NNT (patients)",
  xlab="time since diagnosis (months)", ylim=c(5,100))
> matlines(t, NNTci, lty="dashed", col="black")
```

The resulting plot is displayed in Fig. 2. We observe that the NNT declines with time, from 68.6 patients to 13.8 patients at 10 and 60 months after diagnosis, respectively. Thus, the analysis suggests that 68.6 patients would have had to be treated, among those that were factually untreated, in order to prevent one death/relapse before 10 months. When considering a time window up to 60 months after diagnosis, it would have been enough to treat 13.8 patients.

The RERI

Definition

The RERI is usually defined for two binary exposures. To follow the notation in "Regression standardization" section it is convenient to recode the two exposures into one categorical exposure with levels 00 (both exposures equal to 0), 01 (first exposure equal to 0, second equal to 1), 10 (first exposure equal to 1, second equal to 0) and 11 (both exposures equal to 1). The (causal) RERI is defined as

$$\mathrm{RERI} = \frac{p(Y_{11} = 1) - p(Y_{10} = 1) - p(Y_{01} = 1) + p(Y_{00} = 1)}{p(Y_{00} = 1)}, \tag{9}$$

where, as before, $p(Y_x = 1)$ is the counterfactual probability of the outcome if everybody would be exposed to level $X = x$.

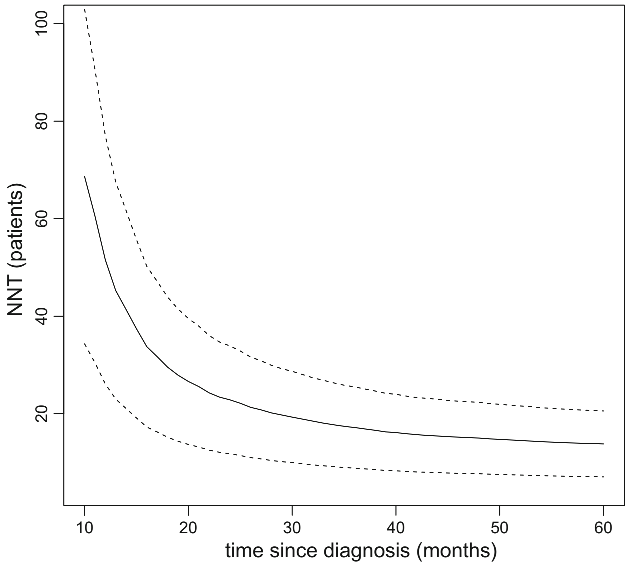

Fig. 2 The estimated NNT (solid line) as a function of time since diagnosis, together with point-wise 95% confidence intervals (dashed lines)

The numerator in (9) is the additive interaction between the two exposures. It has been argued that additive interaction is more useful for assessing the public health importance of interventions than interactions on other (e.g. multiplicative) scales. Furthermore, additive interactions can sometimes be used to infer the presence of certain 'mechanistic/biologic' interactions (see [4] and the references therein). Because the RERI is defined as the additive interaction divided with the positive constant $p(Y_{00} = 1)$, the RERI will always have the same sign (positive, negative or zero) as the additive interaction. We note that there is a wide variety of interaction measures in the epidemiologic literature, and we refer to VanderWeele and Knol [4] for a discussion of their interpretations and relative merits.

For time-to-event outcomes, the RERI is defined as in (9), but with Y_x replaced with $Y_x(t)$ for all x. Thus, the RERI becomes a function of time t.

Estimation with logistic regression models

We again illustrate the methods with the clslowbwt dataset, and we let the two exposures of interest be smoker and race. Estimating the causal effect of race poses two important problems. First, it can be argued that the underlying counterfactual query (e.g. 'what would the probability of the outcome be if everyone was black/white?') is vague, and that the causal effect of race is thus ill-defined [25]. Second, for any given outcome there is arguably a huge number of risk factors that also correlate with race, and thus the potential for unmeasured confounding is enormous. We ignore these problems here, since our analysis merely serves as an illustration.

To make race binary we restrict the analysis to women who are either black or white, and we define a new four-level exposure as

```
> clslowbwt$smokerrace <- factor(paste0(clslowbwt$smoker,
  as.numeric(clslowbwt$race=="2. Black")))
```

We fit the logistic regression model

```
> fit <- glm(formula=lbw~smokerrace+age, family="binomial",
  data=clslowbwt, subset=race!="3. Other")
```

The subsetting on race!=``3. Other'' restricts the analysis to women who are either black or white. We use the fitted model to estimate standardized probabilities:

```
> fit.std <- stdGlm(fit=fit, X="smokerrace", data=clslowbwt,
  clusterid="id")
```

When the exposure is a factor variable, the stdGlm function by default standardize at all exposure levels, which makes it unnecessary to specify the x argument. Summarizing gives

```
   Estimate Std. Error lower 0.95 upper 0.95
00   0.3739     0.0694     0.2378      0.510
01   0.0957     0.0564    -0.0147      0.206
10   0.4369     0.0579     0.3235      0.550
11   0.3885     0.1214     0.1506      0.626
```

There appears to be a strong heterogeneity in the risk of low birthweight between the four exposure groups. Among black non-smokers ($x = 01$) the risk of low birthweight is 9.6%, whereas among white smokers ($x = 10$) the risk is 43.7%.

We define a function that estimates the RERI

```
> RERI <- function(est){
  p00 <- est[1]
  p01 <- est[2]
  p10 <- est[3]
  p11 <- est[4]
  reri <- (p11-p10-p01+p00)/p00
  return(reri)
}
```

and we use this function to estimate the RERI together with a 95% confidence interval

```
> RERIest <- RERI(fit.std$est)
> RERIest
[1] 0.6145113
> confint(object=fit.std, fun=RERI)
          lower    upper
[1,] -0.1574166 1.386439
```

The estimated RERI is equal to 0.61 and the 95% confidence interval includes 0. Thus, we cannot rule out the null hypothesis of no additive interaction between smoking and race.

Estimation with Cox PH regression models

We again illustrate the methods with the rott2 dataset, and we let the two exposures of interest be nochemo and grade.

We first define a new four-level exposure:

```
> rott2$chemograde <- factor(paste0(rott2$nochemo,
  rott2$grade-2))
```

We fit a Cox PH model and use the fitted model to estimate standardized probabilities

```
> fit <- coxph(Surv(rf,rfi)~chemograde+age+meno+size+
  I(exp(-0.12*nodes))+pr+er, data=rott2, method="breslow")
> fit.std <- stdCoxph(fit=fit, data=rott2, X="chemograde", t=10:60)
```

We define a function that estimates the RERI:

```
> RERI <- function(est){
  p00 <- 1-est[, 1]
  p01 <- 1-est[, 2]
  p10 <- 1-est[, 3]
  p11 <- 1-est[, 4]
  reri <- (p11-p10-p01+p00)/p00
  return(reri)
}
```

and we use this function to plot the estimated RERI together with point-wise 95% confidence intervals

```
> RERIest <- RERI(fit.std$est)
> RERIci <- confint(object=fit.std, fun=RERI)
> plot(t, RERIest, type="l", ylab="RERI",
  xlab="time since diagnosis (months)", ylim=c(-0.1,0.6))
> matlines(t, RERIci, lty="dashed", col="black")
```

The resulting plot is displayed in Fig. 3. We observe that the RERI decreases slightly with time, from 0.28 at 10 months after diagnosis, to 0.16 at 60 months after diagnosis.

Discussion

Measures of causal effects play a central role in epidemiology. Using appropriate measures when summarizing the results is crucial to make the analysis relevant from a public health perspective. In this paper we have shown how a wide range of effect measures can be estimated with the R-package stdReg, with a minimal effort of programming from the analyst. We have specifically focused on the AF, the NNT and the RERI, but in principle any effect measure can be estimated along the same lines as these, provided that the measure can be written as come contrast between standardized probabilities.

If the confounders included in the regression model are sufficient for confounding control, then standardization estimates the counterfactual probability of the outcome, had everybody in the population attained a fixed level of the exposure. In this sense, standardization estimates population (or marginal) causal effects. An alternative is to use the fitted regression model to estimate causal effects at specific levels of the confounders, i.e. subpopulation (or conditional) causal effects. In the standard use of logistic regression and Cox PH regression it is assumed that the odds ratio and hazard ratio, respectively, are constant across levels of the confounders. However, these models generally imply that other measures, such as the AF and the NNT, vary across confounder levels. To present conditional causal effects, other than the odds ratio of hazard ratio, the analyst would then typically have to restrict attention to a few selected confounder levels, which makes the results less general than when presenting marginal causal effects.

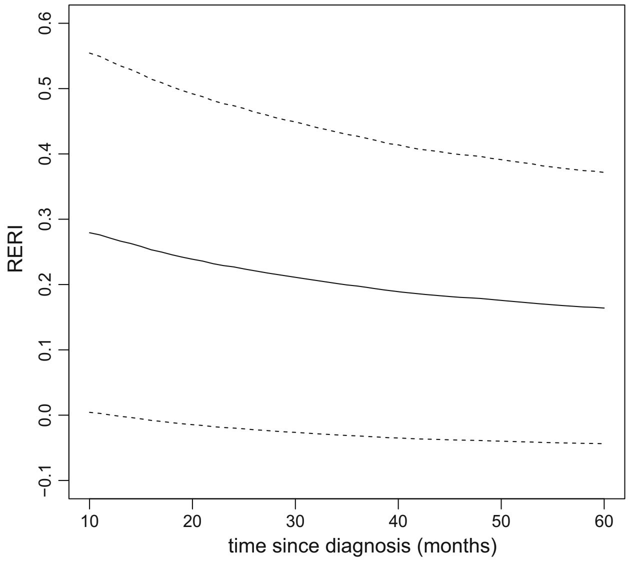

Fig. 3 The estimated RERI (solid line) as a function of time since diagnosis, together with point-wise 95% confidence intervals (dashed lines)

We emphasize that, although slightly beyond the scope of our paper, careful model selection is crucial for estimation of causal effects, and rather different than model selection for prediction. When the aim is to make predictions, one usually attempts to include variables that are strongly associated with the outcome, regardless of the underlying mechanism. Such variables can be selected by fairly automatized procedures, such as step-wise regression. When the aim is to estimate causal effects, one should attempt to include variables that are confounders for the exposure–outcome relationship. Such variables are often strongly associated with the outcome. However, the reverse does not hold; a variable may be strongly associated with the outcome, yet it is not a confounder, and may lead to substantial bias if included in the regression model [10].

The stdReg package uses a fitted regression model to carry out standardization. In this paper we have focused on logistic regression models and Cox PH regression models, since these are the most common models in epidemiology.

More generally though, the function `stdGlm` can be used to carry out standardization with any type of generalized linear model fitted by the `glm` function, e.g. linear regression or probit regression, as described by Sjölander [7]. The `stdReg` package also contains a function for standardization with shared frailty gamma-Weibull models, `stdParfrailty`, which is described by Dahlqwist et al. [26]. In the future we plan to extend the package even further, to allow for standardization with semiparametric frailty models and generalized linear mixed models.

All code in this paper is available at the HTML version of `R`'s online documentation, which is accessed by `help.start()`.

Acknowledgements The author gratefully acknowledges financial support from the Swedish Research Council, Grant No. 2016-01267.

References

1. Davies H, Crombie I, Tavakoli M. When can odds ratios mislead? BMJ. 1998;316(7136):989–91.
2. Holcomb W Jr, Chaiworapongsa T, Luke D, Burgdorf K. An odd measure of risk: use and misuse of the odds ratio. Obstet Gynecol. 2001;98(4):685–8.
3. Case L, Kimmick G, Paskett E, Lohman K, Tucker R. Interpreting measures of treatment effect in cancer clinical trials. The Oncologist. 2002;7(3):181–7.
4. VanderWeele T, Knol M. A tutorial on interaction. Epidemiol Methods. 2014;3(1):33–72.
5. Rothman K, Greenland S, Lash T. Modern epidemiology. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
6. Sjölander A, Dahlqwist E. stdReg: regression standardization. R package version 2.2.0; 2017.
7. Sjölander A. Regression standardization with the R package stdReg. Eur J Epidemiol. 2016;31(6):563–74.
8. Dahlqwist E, Sjölander A. AF: model-based estimation of confounder-adjusted attributable fractions. R package version 0.1.4; 2017.
9. Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educ Psychol. 1974;66(5):688–701.
10. Pearl J. Causality: models, reasoning, and inference. 2nd ed. New York: Cambridge University Press; 2009.
11. Casella G, Berger R. Statistical inference. 2nd ed. Pacific Grove, CA: Duxbury; 2002.
12. Sjölander A. Attributable fractions. Wiley StatsRef: Statistics Reference Online; 2014.
13. Chen Y, Hu C, Wang Y. Attributable risk function in the proportional hazards model for censored time-to-event. Biostatistics. 2006;7(4):515–29.
14. Chen L, Lin D, Zeng D. Attributable fraction functions for censored event times. Biometrika. 2010;97(3):713–26.
15. Sturmans F, Mulder P, Valkenburg H. Estimation of the possible effect of interventive measures in the area of ischemic heart diseases by the attributable risk percentage. Am J Epidemiol. 1977;105(3):281–9.
16. Deubner D, Wilkinson W, Helms M, Herman T, Curtis G. Logistic model estimation of death attributable to risk factors for cardiovascular disease in Evans County, Georgia. Am J Epidemiol. 1980;112(1):135–43.
17. Greenland S, Drescher K. Maximum likelihood estimation of the attributable fraction from logistic models. Biometrics. 1993;49(3):865–72.
18. Sjölander A, Vansteelandt S. Doubly robust estimation of attributable fractions. Biostatistics. 2011;12(1):112–21.
19. Sjölander A, Vansteelandt S. Doubly robust estimation of attributable fractions in survival analysis. Stat Methods Med Res. 2014; https://doi.org/10.1177/0962280214564003.
20. Stefanski L, Boos D. The calculus of m-estimation. Am Stat. 2002;56(1):29–38.
21. Sauerbrei W, Royston P, Look M. A new proposal for multivariable modelling of time-varying effects in survival data based on fractional polynomial time-transformation. Biometr J. 2007;49(3):453–73.
22. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. N Engl J Med. 1988;318(26):1728–33.
23. Bender R, Kuss O, Hildebrandt M, Gehrmann U. Estimating adjusted NNT measures in logistic regression analysis. Stat Med. 2007;26(30):5586–95.
24. Laubender RP, Bender R. Estimating adjusted risk difference (RD) and number needed to treat (NNT) measures in the cox regression model. Stat Med. 2010;29(7–8):851–9.
25. VanderWeele T, Hernan M. Causal effects and natural laws: towards a conceptualization of causal counterfactuals for non-manipulable exposures, with application to the effects of race and sex. In: Berzuini P, Dawid P, Bernardinelli L, editors. Causality: statistical perspectives and applications. Chichester: Wiley; 2012. p. 101–13.
26. Dahlqwist E, Pawitan Y, Sjölander A. Regression standardization and attributable fraction estimation with between-within frailty models for clustered survival data. Stat Methods Med Res. https://doi.org/10.1177/0962280217727558.