Johan Zetterqvist and Arvid Sjölander*

Doubly Robust Estimation with the R Package drgee

DOI 10.1515/em-2014-0021

Abstract: A common goal of epidemiologic research is to study the association between a certain exposure and a certain outcome, while controlling for important covariates. This is often done by fitting a restricted mean model for the outcome, as in generalized linear models (GLMs) and in generalized estimating equations (GEEs). If the covariates are high-dimensional, then it may be difficult to well specify the model. This is an important concern, since model misspecification may lead to biased estimates. Doubly robust estimation is an estimation technique that offers some protection against model misspecification. It utilizes two models, one for the outcome and one for the exposure, and produces unbiased estimates of the exposure-outcome association if either model is correct, not necessarily both. Despite its obvious appeal, doubly robust estimation is not used on a regular basis in applied epidemiologic research. One reason for this could be the lack of up-to-date software. In this paper we describe a new `R` package, `drgee`, which carries out doubly robust estimation in restricted mean models. The package is constructed to be user-friendly and fast, to facilitate routine use of doubly robust estimation. The paper is structured into theory sections and example sections. The former are intended to serve as a brief but self-consistent tutorial in doubly robust estimation. The latter illustrate the use of the `drgee` package through practical examples. We have used publically available data throughout the paper, so that the reader can easily replicate all examples.

Keywords: generalized estimating equation, generalized linear model, doubly robust estimation, confounding, mediation

1 Introduction

A common goal of epidemiologic research is to study the association between a certain exposure A and a certain outcome Y. Typically, it is desirable to control for covariates L in the analysis. This, for instance, is the case when the covariates are potential confounders for the exposure-outcome association, or when the covariates are potential mediators and the aim is to study the direct exposure effect. A common tool for covariate control is the restricted mean model

$$g\{E(Y|A,L)\} = \beta A + \gamma_0 + \gamma_1 L, \quad [1]$$

where Y is a scalar outcome and g is a suitable link function. Special cases include the linear, log-linear, and logistic models, which are typically used for continuous, "count", and binary outcomes, respectively. The restricted mean model can be fitted by solving a set of GEEs. Alternatively, one may assume that Y has an exponential family distribution with canonical link function g, and fit the model by solving the GLM maximum likelihood score equations.

The model in eq. [1] can be decomposed into two parts:

$$g\{E(Y|A,L)\} - g\{E(Y|A=0,L)\} = \beta A \quad [2]$$

and

$$g\{E(Y|A=0,L)\} = \gamma_0 + \gamma L. \quad [3]$$

***Corresponding author: Arvid Sjölander,** Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden, E-mail: arvid.sjolander@ki.se

Johan Zetterqvist, Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

The part in eq. [2] quantifies the conditional association between A and Y, given L. This part is usually of main interest; we thus refer to it as the "main model". The parameter in the main model, β, is the target parameter. The part in eq. [3] includes the covariates and is usually of secondary interest; we refer to it as the "outcome nuisance model". The two parts are variation independent, which means that the first part may be correct even though the second part is misspecified, or vice versa. However, if the outcome nuisance model is misspecified, then the GEE/GLM estimator of β is generally biased. In this sense, misspecification of the outcome nuisance model "spills over" to the main model.

To protect the main model against bias bue to misspecification of the outcome nuisance model, doubly robust (DR) estimators have been proposed (e.g. Robins et al., 1992; Robins and Rotnitzky, 2001; Bang and Robins, 2005; Tchetgen Tchetgen et al., 2010). These estimators combine an outcome nuisance model with an exposure nuisance model, and are unbiased for the parameters in the main model if either nuisance model is correctly specified, not necessarily both. Thus, DR estimators give the researcher two chances instead of one to make valid inference on the parameters of main interest.

Despite their obvious appeal, DR estimators are not used on a regular basis in applied epidemiologic research. One reason for this could be the lack of up-to-date software. To remedy this deficiency we have implemented an `R` package, `drgee`, which carries out DR-estimation in linear, log-linear, and logistic restricted mean models (Zetterqvist and Sjölander, 2015). To facilitate routine use we have made an effort to make the `R` package as user-friendly and fast as possible. In particular, we have made sure that the package scales well for large data, and that it has an input/output interface which is similar to the standard model interface in `R`.

Three broad classes of estimation methods are implemented by the `drgee` package. The first class is DR-estimation, which requires both a nuisance model for the outcome and a nuisance model for the exposure. The second class only requires a nuisance model for the exposure; we refer to this class as "E-estimation". Our definition of E-estimation covers both the original E-estimation in linear and log-linear models proposed by Robins et al. (1992) (often referred to as "G-estimation" in the causal infererence literature), and "retrospective maximum likelihood" in logistic models proposed by Tchetgen Tchetgen and Rotnitzky (2011). The third class is standard GEE estimation, which only requires a nuisance model for the outcome; to be consistent in jargon we refer to this class as "O-estimation".

In this paper we describe the `R` package `drgee`. The paper is organized as follows. In Sections 2–Section 4 we review the theory of O-, E-, and DR-estimation, respectively, and illustrate through practical examples how these can be carried out with the `drgee` package. These sections are divided into "theory" and "example" subsections. The theory subsections are intended to serve as a brief but self-consistent tutorial in O-, E-, and DR-estimation; readers who want to get an immediate idea of how the `drgee` package works may skip the more difficult theory subsections at a first reading. An appeal of standard GEEs is their ability to handle clustered data. The `drgee` package can handle clustered data as well, which we illustrate in Section 5. In Section 6 we compare the three estimation methods, and illustrate the doubly robustness property of the DR estimator through a simulation study. In Section 7 we provide concluding remarks.

2 O-estimation

2.1 Theory

We assume that data consist of iid observations of (Y, A, L). The simple model [1] contains only main effects. We here consider a more general model which allows for interactions between covariates, and for interactions between the exposure and (some of) the covariates. Towards this end we let $X(L)$ be a $p \times 1$-dimensional function of L, and $V(L)$ be a $q \times 1$-dimensional function of L. We consider the model

$$g\{E(Y|A,L)\} = \beta^T(AX(L)) + \gamma^T V(L) \qquad [4]$$

where g is a link function, β is a $p \times 1$-dimensional parameter, $AX(L)$ is the element-wise multiplication of A with $X(L)$, and γ is a q-dimensional parameter. Typically, both $V(L)$ and $X(L)$ will contain the element "1", so that there is an intercept in the model, and a main effect of the exposure on the outcome. Arguing as in Section 1 we can decompose the model in eq. [4] into the main model

$$g\{E(Y|A,L)\} - g\{E(Y|A=0,L)\} = \beta^T(AX(L)) \quad [5]$$

and the outcome nuisance model

$$g\{E(Y|A=0,L)\} = \gamma^T V(L). \quad [6]$$

Our main interest lies in the main model and β is our target parameter. The outcome nuisance model is of secondary interest. However, in O-estimation both models are fitted simultaneously, in such a way that misspecification of the outcome nuisance model may cause bias in the estimate of β. In O-estimation we use the estimating function

$$\psi_O(Y,A,L;\beta,\gamma) = \begin{Bmatrix} AX(L) \\ V(L) \end{Bmatrix}[Y - g^{-1}\{\beta^T(AX(L)) + \gamma^T V(L)\}].$$

When both models [5] and [6] are correct, we have that

$$E\{\psi_O(Y,A,L;\beta,\gamma)\} = E[E\{\psi_O(Y,A,L;\beta,\gamma)\}|A,L] = 0.$$

It then follows from standard theory for estimating equations (see Appendix) that a consistent and asymptotically normal estimator $\widehat{\beta}_O$ can be obtained by solving the estimating equation

$$\hat{E}\{\psi_O(Y,A,L;\beta,\gamma)\} = 0$$

for β and γ, where we have used $\hat{E}(\cdot)$ for sample mean. This estimating equation is identical to the ML score equation for a GLM with canonical link function. It is also identical to the estimating equation for a GEE with "independent working correlation matrix". The standard error of $\widehat{\beta}_O$ can be obtained through the "sandwich formula" (see Appendix).

2.2 Example 1

To demonstrate how the `drgee` package can be used for O-estimation, we use the dataset `SLID` from the `car` package (Fox and Weisberg, 2011). This publically available dataset contains data from the 1994 wave of Canadian Survey of Labour and Income dynamics, with variables `wages`, `sex`, `education`, `age` and `language` for composite hourly wage (Canadian dollar), sex, years of education, age and native language respectively. The variables `sex` and `language` are factors with levels (`"Female"`, `"Male"`) and (`"English"`, `"French"`, `"Other"`), respectively. We refer to the `car` package for a more thorough documentation of the dataset.

Suppose that we wish to use these data study if there is a direct effect of sex on wages, not mediated through education level. Such a direct effect is clearly of substantive interest, since it would be an indication of sex discrimination. To eliminate the mediated effect we wish to control for education level. However, to avoid bias we must then additionally control for covariates that can be confounders for the mediator and the outcome (see Valeri and VanderWeele, 2013 and references therein). It is obvious that age can be such a mediator-outcome confounder, since age is likely to be associated with both education level and wages. Native language may also be a mediator-outcome confounder, by being associated with education level and wages through ethnicity and socio-economic status. We thus use a model for the mean of `wages` conditional on `sex`, `education`, `age` and `language`:

$$\begin{aligned} &E(\texttt{wages}|\texttt{sexMale},\texttt{education},\texttt{age},\texttt{languageFrench},\texttt{languageOther};\beta,\gamma_0,\gamma_1,\gamma_2,\gamma_3,\gamma_4) \\ &= \beta\cdot\texttt{sexMale} + \gamma_0 + \gamma_1\cdot\texttt{education} + \gamma_2\cdot\texttt{age} + \gamma_3\cdot\texttt{languageFrench} + \gamma_4\cdot\texttt{languageOther} \end{aligned}$$

where the factors `sex` and `language` are recoded as dummy variables, with reference levels "`Female`" and "`English`", respectively. In this model, $Y = \texttt{wage}$, $A = \texttt{sexMale}$, $L = (\texttt{education}, \texttt{age}, \texttt{languageFrench}, \texttt{languageOther})$, $X(L) = 1$ and $V(L) = (1, L)$. The target parameter is β and the main model is

$$\begin{aligned} &E(\texttt{wages}|\texttt{sexMale}, \texttt{education}, \texttt{age}, \texttt{languageFrench}, \texttt{languageOther}) \\ &- E(\texttt{wages}|\texttt{sexMale} = 0, \texttt{education}, \texttt{age}, \texttt{languageFrench}, \texttt{languageOther}) \\ &= \beta \cdot \texttt{sexMale} \end{aligned} \quad [7]$$

The nuisance parameter is $\gamma = (\gamma_0, \gamma_1, \gamma_2, \gamma_3, \gamma_4)$ and the outcome nuisance model is

$$\begin{aligned} &E(\texttt{wages}|\texttt{sexMale} = 0, \texttt{education}, \texttt{age}, \texttt{languageFrench}, \texttt{languageOther}) \\ &= \gamma_0 + \gamma_1 \cdot \texttt{education} + \gamma_2 \cdot \texttt{age} \\ &\quad + \gamma_3 \cdot \texttt{languageFrench} + \gamma_4 \cdot \texttt{languageOther} \end{aligned} \quad [8]$$

To use O-estimation based on these models we type:

```
> library(car)
> fit <- drgee(oformula=wages~education + age + language,
+ exposure="sex", estimation.method="o", data=SLID)
```

By setting the `estimation.method` argument to "`o`" we tell the `drgee` function to use O-estimation. The `oformula` argument specifies the outcome nuisance model and the `exposure` argument specifies the exposure. There is no need to explicitly specify the outcome, since this is identified by the `drgee` function as the response variabel in `oformula`. To summarize the results we type:

```
> summary(fit)
```

which gives us the output

```
Call: drgee(exposure = "sex", oformula = wages ~ education + age + language, data
      = SLID, estimation.method = "o")

Outcome: wages

Exposure: sexMale

Covariates: education,age,languageFrench,languageOther

Main model: wages ~ sexMale

Outcome nuisance model: wages ~ education + age + languageFrench + languageOther

Outcome link function: identity

      Estimate Std.  Error z value Pr(>|z|)
sexMale    3.4554    0.2091   16.53   <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Note: The estimated parameters quantify the conditional exposure-outcome associa-
tion, given the covariates included in the nuisance models)

3987    complete observations used
```

Only the result for the main parameters are shown in the output. Since `drgee` only uses complete observations, 3428 incomplete observations in the original dataset `SLID` were not used in the calculations.

The estimate of β indicates that men earned 3.46 dollar more per hour than women in the target population, even when controlling for education level, age and native language. This observed sex difference is highly significant, with a p-value much smaller than the nominal 0.05-level.

2.3 Example 2

The linear and interaction-free model in Example 1 is simple, but may not be entirely realistic. Wage distributions are often right skewed, and therefore a linear model may not fit data very well. Furthermore, inference for direct effects may be misleading if significant exposure-mediator interactions are omitted from the model (see Valeri and VanderWeele, 2013 and references therein).

Therefore, suppose that we want to use a log-linear model instead, including an interaction between sex and education. We then replace the main model [7] with

$$\begin{aligned}&\log\{E(\texttt{wages}|\texttt{sexMale},\texttt{education},\texttt{age},\texttt{languageFrench},\texttt{languageOther})\}\\&-\log\{E(\texttt{wages}|\texttt{sexMale}=0,\texttt{education},\texttt{age},\texttt{languageFrench},\texttt{languageOther})\}\\&=\beta_1\cdot\texttt{sexMale}+\beta_2\cdot\texttt{sexMale}\cdot\texttt{education},\end{aligned}\qquad[9]$$

in which $X(L)=(1,\texttt{education})$, and the outcome nuisance model [8] with

$$\begin{aligned}&\log\{E(\texttt{wages}|\texttt{sexMale}=0,\texttt{education},\texttt{age},\texttt{languageFrench},\texttt{languageOther})\}\\&=\gamma_0+\gamma_1\cdot\texttt{education}+\gamma_2\cdot\texttt{age}+\gamma_3\cdot\texttt{languageFrench}+\gamma_4\cdot\texttt{languageOther}\end{aligned}\qquad[10]$$

To use O-estimation based on these models we type:

```
> fit <- drgee(oformula = wages~education + age + language,
+ exposure = "sex", iaformula = ~education, olink = "log",
+ estimation.method = "o", data = SLID)
```

The `olink` argument specifies the link function g. It defaults to "`identity`" which gives a linear model. By setting the `olink` argument equal to "`log`" we tell the `drgee` function to fit a log-linear model instead. The `iaformula` argument specifies $X(L)$, i.e. the set of covariates that are assumed to interact with the exposure. Summarizing the results gives:

```
Call: drgee(exposure = "sex", oformula = wages ~ education + age + language,
   iaformula = ~education, olink = "log", data = SLID, estimation.method = "o")

Outcome: wages

Exposure: sexMale

Covariates: education,age,languageFrench,languageOther

Main model: wages ~ sexMale + sexMale:education

Outcome nuisance model: wages ~ education + age + languageFrench + languageOther

Outcome link function: log

                     Estimate Std. Error z value Pr(>|z|)
sexMale              0.581088   0.062848   9.246  < 2e-16 ***
sexMale:education   -0.026175   0.004511  -5.802 6.54e-09 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
(Note: The estimated parameters quantify the conditional exposure-outcome associa-
tion, given the covariates included in the nuisance models)
3987     complete observations used
```

We observe that both the main effect of sex and the sex-education interaction are highly significant.

3 E-estimation

In E-estimation we use the main model [5], as in O-estimation. However, in contrast to O-estimation, E-estimation leaves the covariate part $g\{E(Y|A=0,L)\}$ unspecified. Instead E-estimation combines the main model [5] with an exposure nuisance model. The `drgee` package implements E-estimation when g is the identity, log or logit link function. We discuss the identity and log link functions in Section 3.1. The logit function requires special treatment, and is discussed in Section 3.3.

3.1 Theory: E-estimation when g is the identity or log link function

When g is the identity or log link function, we use a model for the mean of the exposure conditional on the covariates:

$$h\{E(A|L)\} = \alpha^T Z(L). \tag{11}$$

In this model, h the identity, log or logit link function, α an $r \times 1$-dimensional nuisance parameter and $Z(L)$ an $r \times 1$-dimensional function of L. We will refer to eq. [11] as the exposure nuisance model. To find an E-estimator of β, we can proceed as follows. Define

$$S(Y,A,L;\beta) = g^{-1}\{g(Y) - \beta^T(AX(L))\} \tag{12}$$

When the main model [5] is correct with true parameter β, $S(Y,A,L;\beta)$ predicts $E(Y|A=0,L)$. If g is the identity link function, then

$$\begin{aligned} E\{S(Y,A,L;\beta)|A,L\} &= E\{Y - \beta^T(AX(L))|A,L\} \\ &= E(Y|A,L) - \beta^T(AX(L)) \\ &= E(Y|A=0,L), \end{aligned}$$

where the main model [5] was used in the last equality. If g is the log link function, then

$$\begin{aligned} E\{S(Y,A,L;\beta)|A,L\} &= E\left\{e^{\log(Y)-\beta^T(AX(L))}|A,L\right\} \\ &= E(Y|A,L)e^{-\log\{E(Y|A,L)\}+\log\{E(Y|A=0,L)\}} \\ &= E(Y|A=0,L), \end{aligned}$$

where the main model [5] was used in the second equality. Since $A - h^{-1}\{\alpha^T Z(L)\}$ is a constant conditional on A and L and since $X(L)$, $E(Y|A=0,L)$ and $h^{-1}\{\alpha^T Z(L)\}$ are constants conditional on L, we have (by the law of total expectation) that

$$\begin{aligned} &E\left\{X(L)\left[A - h^{-1}\{\alpha^T Z(L)\}\right]S(Y,A,L;\beta)|L\right\} \\ &= X(L)E\left\{E\left(\left[A - h^{-1}\{\alpha^T Z(L)\}\right]S(Y,A,L;\beta)|A,L\right)|L\right\} \\ &= X(L)E\left(\left[A - h^{-1}\{\alpha^T Z(L)\}\right]E\{S(Y,A,L;\beta)|A,L\}|L\right) \\ &= X(L)E\left(\left[A - h^{-1}\{\alpha^T Z(L)\}\right]E(Y|A=0,L)|L\right) \\ &= X(L)\left[E(A|L) - h^{-1}\{\alpha^T Z(L)\}\right]E(Y|A=0,L). \end{aligned}$$

The last expression equals 0 when the exposure nuisance model [11] is correct. This motivates the estimating function

$$\psi_E(Y,A,L;\beta,\alpha) = \begin{bmatrix} X(L)[A - h^{-1}\{\alpha^T Z(L)\}]S(Y,A,L;\beta) \\ Z(L)[A - h^{-1}\{\alpha^T Z(L)\}] \end{bmatrix}.$$

When both the main model [5] and the exposure nuisance model [11] are correct, we have that

$$E\{\psi_E(Y,A,L;\beta,\alpha)\} = E[E\{\psi_E(Y,A,L;\beta,\alpha)|L\}] = 0.$$

It then follows from standard theory for estimating equations (see Appendix) that a consistent and asymptotically normal estimator $\widehat{\beta}_E$ can be obtained by solving the estimating equation

$$\hat{E}\{\psi_E(Y,A,L;\beta,\alpha)\} = 0$$

for β and α. This approach is often referred to as "G-estimation" in the causal inference literature, and was initially proposed by Robins et al. (1992). The standard error of $\widehat{\beta}_E$ can be obtained through the "sandwich formula" (see Appendix).

3.2 Example 3

Continuing Example 2 (Section 2.3), suppose that we want to combine the main model [9] with the exposure nuisance model

$$\begin{aligned} &\text{logit}\{E(\texttt{sexMale}|\texttt{education},\texttt{age},\texttt{languageFrench},\texttt{languageOther})\} \\ &= \alpha_0 + \alpha_1 \cdot \texttt{education} + \alpha_2 \cdot \texttt{age} + \alpha_3 \cdot \texttt{languageFrench} + \alpha_4 \cdot \texttt{languageOther} \end{aligned} \quad [13]$$

To use E-estimation based on these models we would type:

```
> fit <- drgee(outcome="wages",
+ eformula=sex~education + age + language,
+ iaformula=~education, olink="log", elink="logit",
+ estimation.method="e", data=SLID)
```

By setting the `estimation.method` argument to "`e`" we tell the `drgee` function to use E-estimation. We then use an `eformula` argument to specify the exposure nuisance model. Now there is no need to explicitly specify the exposure, since this is identified by the `drgee` function as the response variable in `eformula`. The elink argument specifies the link function h for the exposure nuisance model. Since no outcome nuisance model is used in E-estimation, the `oformula` argument may be omitted. However, we then need to specify the `outcome` through the outcome argument. Summarizing the results gives:

```
Call: drgee(outcome = "wages", eformula = sex ~ education + age
   + language, iaformula = ~education, olink = "log", elink =
   "logit", data = SLID, estimation.method = "e")

Outcome: wages

Exposure: sexMale

Covariates: education,age,languageFrench,languageOther

Main model: wages ~ sexMale + sexMale:education

Outcome link function: log

Exposure nuisance model: sexMale ~ education + age + languageFrench + languageOther

Exposure link function: logit
```

```
                    Estimate Std. Error z value Pr(>|z|)
sexMale              0.370139   0.064773   5.714   1.1e-08 ***
sexMale:education   -0.010613   0.004738  -2.240   0.0251 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Note: The estimated parameters quantify the conditional exposure-outcome associa-
tion, given the covariates included in the nuisance models)

3987  complete observations used
```

We observe that the obtained E-estimates are quite different from the O-estimates (see Section 2.3). This indicates that at least one of the nuisance models [10] and [13] is misspecified.

3.3 Theory: E-estimation when *g* is the logit link function

When g is the logit link function and main model [5] holds, $E\{S(Y,A,L;\beta)|A,L\}$ is not generally equal to $E(Y|A=0,L)$. This implies that G-estimation (see Section 3.1) does not work. However, when both A and Y are binary we can instead use the symmetry of the odds ratio to perform E-estimation. In this case, the main model [5] is equivalent with the model

$$\text{logit}\{E(A|Y=1,L)\} - \text{logit}\{E(A|Y=0,L)\} = \beta^T X(L).$$

Combining this with the exposure nuisance model

$$\text{logit}\{E(A|Y=0,L)\} = \delta^T W(L), \qquad [14]$$

where δ is an $r \times 1$-dimensional parameter and $W(L)$ is an $r \times 1$-dimensional function of L, we arrive at the model

$$\text{logit}\{E(A|Y,L)\} = \beta^T (YX(L)) + \delta^T W(L),$$

which has the same structural form as the model [4] used in O-estimation when g is the logit link function. Therefore, we can use the estimating function ψ_O to obtain perform E-estimation if we let A and Y "switch place" in the estimating function. It then follows from standard theory for estimating equations (see Appendix) that a consistent and asymptotically normal estimator $\widehat{\beta}_E$ can be obtained by solving the estimating equation

$$\hat{E}\left(\begin{Bmatrix} YX(L) \\ W(L) \end{Bmatrix} [A - \text{expit}\{\beta^T (YX(L)) + \delta^T W(L)\}] \right) = 0$$

for β and δ. The standard error of $\widehat{\beta}_E$ can be obtained through the "sandwich formula" (see Appendix).

3.4 Example 4

To perform E-estimation with logit outcome link function, we need both outcome and exposure to be binary. Suppose that we recode `wages` in the `SLID` dataset as a binary variable:

```
> SLID$highWage<-ifelse(SLID$wages<=14,0,1)
```

We can then use the logistic main model

$$\begin{aligned}&\text{logit}\{E(\texttt{highWage}|\texttt{sexMale},\texttt{education},\texttt{age},\texttt{languageFrench},\texttt{languageOther})\}\\&-\text{logit}\{E(\texttt{highWage}|\texttt{sexMale}=0,\texttt{education},\texttt{age},\texttt{languageFrench},\texttt{languageOther})\}\\&=\beta_1\cdot\texttt{sexMale}+\beta_2\cdot\texttt{sexMale}\cdot\texttt{education}\end{aligned} \quad [15]$$

and the logistic exposure nuisance model

$$\begin{aligned}&\text{logit}\{E(\texttt{sexMale}|\texttt{highWage}=0,\texttt{education},\texttt{age},\texttt{languageFrench},\texttt{languageOther})\}\\&=\delta_0+\delta_1\cdot\texttt{education}+\delta_2\cdot\texttt{age}+\delta_3\cdot\texttt{languageFrench}+\delta_4\cdot\texttt{languageOther}\end{aligned} \quad [16]$$

To use E-estimation based on these models we type:

```
> fit <-
drgee(outcome = "highWage", eformula = sex~education + age + language,
+ iaformula = ~education, olink = "logit", elink = "logit",
+ estimation.method = "e", data = SLID)
> summary(fit)

Call: drgee(outcome = "highWage", eformula = sex ~ education + age +
   language, iaformula = ~education, olink = "logit", elink = "logit",
   data = SLID, estimation.method = "e")

Outcome: highWage

Exposure: sexMale

Covariates: education,age,languageFrench,languageOther

Main model: highWage ~ sexMale + sexMale:education

Outcome link function: logit

Exposure nuisance model: sexMale ~ education + age + languageFrench + languageOther

Exposure link function: logit

                    Estimate Std. Error z value Pr(>|z|)
sexMale              1.93520   0.30980   6.247   4.2e-10 ***
sexMale:education   -0.06361   0.02287  -2.782   0.00541 **
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Note: The estimated parameters quantify the conditional exposure-outcome associa-
tion, given the covariates included in the nuisance models)

3987   complete observations used
```

Again, both main effect and interaction are highly significant.

4 DR-estimation

In DR-estimation, we combine a main model with an outcome nuisance model and an exposure nuisance model to construct an estimator $\widehat{\beta}_{DR}$ such that when the main model is correct, $\widehat{\beta}_{DR}$ is consistent and asymptotically normal if at least one of the nuisance models is correct. As with E-estimation, we treat the case when g is the logit link function separately.

4.1 Theory: DR-estimation when *g* is the identity or log link function

Let $\gamma^T V(L)$ be a model for $g\{E(Y|A=0,L)\}$ and $\alpha^T Z(L)$ a model for $h\{E(A|L)\}$ as in eqs [6] and [11], respectively. Also, let $S(Y,A,L;\beta)$ be as in eq. [12]. When the main model [5] is correct and when g is the identity or log link function, the results in Section 3 imply that

$$E\big[S(Y,A,L;\beta) - g^{-1}\{\gamma^T V(L)\}|A,L\big] = E(Y|A=0,L) - g^{-1}\{\gamma^T V(L)\}.$$

We then have that

$$\begin{aligned}
&E\big[X(L)[A - h^{-1}\{\alpha^T Z(L)\}]S(Y,A,L;\beta) - g^{-1}\{\gamma^T V(L)\}|L\big] \\
&= X(L)E\big\{E\big([A - h^{-1}\{\alpha^T Z(L)\}][S(Y,A,L;\beta) - g^{-1}\{\gamma^T V(L)\}]|A,L\big)|L\big\} \\
&= X(L)E\big([A - h^{-1}\{\alpha^T Z(L)\}]E\big[S(Y,A,L;\beta_0) - g^{-1}\{\gamma^T V(L)\}|A,L\big]|L\big) \\
&= X(L)E\big([A - h^{-1}\{\alpha^T Z(L)\}][E(Y|A=0,L) - g^{-1}\{\gamma^T V(L)\}]|L\big) \\
&= X(L)[E(A|L) - h^{-1}\{\alpha^T Z(L)\}][E(Y|A=0,L) - g^{-1}\{\gamma^T V(L)\}].
\end{aligned}$$

The last expression equals 0 when at least one of the nuisance models [6] and [11] is correct. This motivates the estimating function

$$\psi_{DR}(Y,A,L;\beta,\beta^\dagger,\gamma,\alpha) = \begin{pmatrix} X(L)\big[A - h^{-1}\{\alpha^T Z(L)\}\big]\big[S(Y,A,L;\beta) - g^{-1}\{\gamma^T V(L)\}\big] \\ \left\{\begin{matrix} AX(L) \\ V(L) \end{matrix}\right\}\big[S(Y,A,L;\beta^\dagger) - g^{-1}\{\gamma^T V(L)\}\big] \\ Z(L)\big[A - h^{-1}\{\alpha^T Z(L)\}\big] \end{pmatrix}.$$

When the main model [5] is correct and either the outcome nuisance model [6] or the exposure nuisance model [14] is correct, we have that

$$E\{\psi_{DR}(Y,A,L;\beta,\beta^\dagger,\gamma,\alpha)\} = E[E\{\psi_{DR}(Y,A,L;\beta,\beta^\dagger,\gamma,\alpha)|L\}] = 0.$$

It then follows from standard theory for estimating equations (see Appendix) that a consistent and asymptotically normal estimator $\hat{\beta}_{DR}$ can be obtained by solving the equation

$$\hat{E}\{\psi_{DR}(Y,A,L;\beta,\beta^\dagger,\gamma,\alpha)\} = 0 \qquad [17]$$

for $(\beta^T,\beta^{\dagger T},\gamma^T,\alpha^T)^T$. The standard error of $\widehat{\beta}_{DR}$ can be obtained through the "sandwich formula" (see Appendix).

4.2 Example 5

Continuing Example 2 (Section 2.3) and Example 3 (Section 3.2), suppose that we want to combine the main model [9] with the outcome nuisance model [10] and the exposure nuisance model [13]. To use DR-estimation based on these models we type:

```
> fit <- drgee(oformula=wages~education+age+language,
+ eformula=sex~education+age+language,
+ iaformula=~education, olink="log", elink="logit",
+ estimation.method="dr", data=SLID)
```

By setting the `estimation.method` argument to "`dr`" we tell the `drgee` function to use DR-estimation. The `outcome` and `exposure` arguments may be omitted, since the `drgee` function identifies the outcome and exposure as the response variables in `oformula` and `eformula`, respectively. Summarizing the results gives:

```
> summary(fit)

Call: drgee(oformula = wages ~ education + age + language,
   eformula = sex ~ education + age + language,
   iaformula = ~education, olink = "log",
   elink = "logit", data = SLID, estimation.method = "dr")

Outcome: wages

Exposure: sexMale

Covariates: education,age,languageFrench,languageOther

Main model: wages ~ sexMale + sexMale:education

Outcome nuisance model: wages ~ education + age + languageFrench + languageOther

Outcome link function: log

Exposure nuisance model: sexMale ~ education + age + languageFrench + languageOther

Exposure link function: logit

                    Estimate Std. Error z value Pr(>|z|)
sexMale              0.57752   0.06333   9.119    < 2e-16 ***
sexMale:education   -0.02591   0.00455  -5.696   1.23e-08 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Note: The estimated parameters quantify the conditional exposure-outcome
association, given the covariates included in the nuisance models)

3987 complete observations used
```

We observe that the DR estimates are very similar to the O-estimates in Example 2 (Section 2.3), but quite different from the E-estimates in Example 3 (Section 3.2). This indicates that the exposure nuisance model may not be well specified. However, it does not prove that the outcome nuisance model is correct; we demonstrate in Section 6.1 that all methods may agree well even though both nuisance models are misspecified.

4.3 Theory: DR-estimation when *g* is the logit link function

When g is the logit link function, DR-estimation can be performed by the `drgee` package if the exposure A is binary and h is the logit link function. In this case the main model [5] is equivalent to the odds ratio model

$$\text{logit}\{E(Y|A=1,L)\} - \text{logit}\{E(Y|A=0,L)\} = \beta^T X(L). \quad [18]$$

By combining this model with the outcome nuisance model

$$\text{logit}\{E(Y|A=0,L)\} = \gamma^T V(L) \quad [19]$$

and the exposure nuisance model [14] we can construct the estimating function

$$\psi_{DR}(Y,A,L;\beta,\beta^\dagger,\beta^\ddagger,\gamma,\delta) = \begin{pmatrix} X(L)\{A - E^*(L;\beta,\gamma,\delta)\}[Y - \text{expit}\{\beta^T(AX(L)) + \gamma^T V(L)\}] \\ \left\{ \begin{matrix} AX(L) \\ V(L) \end{matrix} \right\}[Y - \text{expit}\{\beta^{\dagger T} AX(L) + \gamma^T V(L)\}] \\ \left\{ \begin{matrix} YX(L) \\ W(L) \end{matrix} \right\}[A - \text{expit}\{\beta^{\ddagger T} YX(L) + \delta^T W(L)\}] \end{pmatrix}$$

where

$$E^*(L;\beta,\gamma,\delta) = \left[1 + \frac{[1-\text{expit}\{\delta^T W(L)\}]\text{expit}\{\gamma^T V(L)\}}{\text{expit}\{\delta^T W(L)\}\text{expit}\{\beta^T X(L) + \gamma^T V(L)\}}\right]^{-1}.$$

Tchetgen Tchetgen et al. (2010) showed that

$$E\{\psi_{DR}(Y,A,L;\beta,\beta^\dagger,\beta^\ddagger,\gamma,\delta\} = E[E\{\psi_{DR}(Y,A,L;\beta,\beta^\dagger,\beta^\ddagger,\gamma,\delta)|L\}] = 0.$$

when the main model [18] is correct and either the outcome nuisance model [19] or the exposure nuisance model [14] is correct. It then follows from standard theory for estimating equations (see Appendix) that a consistent and asymptotically normal estimator $\hat{\beta}_{DR}$ can be obtained by solving the equation

$$\hat{E}\{\psi_{DR}(Y,A,L;\beta,\beta^\dagger,\beta^\ddagger,\gamma,\delta)\} = 0$$

for $(\beta^T,\beta^{\dagger T},\beta^{\ddagger T},\gamma^T,\delta^T)^T$. Tchetgen Tchetgen et al. (2010) showed that this estimator attains the semiparametric efficiency bound when all three models are correct, i.e. under the intersection of models [14], [18] and [19]. We note that this intersection model is not fully parametric, since it does not parametrize the distribution of L.

4.4 Example 6

Continuing Example 4 (Section 3.4), suppose that we want to combine main model [15], the exposure nuisance model [16] and the logistic outcome nuisance model

$$\begin{aligned}&\text{logit}\{E(\texttt{highWage}|\texttt{sexMale}=0,\texttt{education},\texttt{age},\texttt{languageFrench},\texttt{languageOther})\}\\&\quad=\gamma_0+\gamma_1\cdot\texttt{education}+\gamma_2\cdot\texttt{age}+\gamma_3\cdot\texttt{languageFrench}+\gamma_4\cdot\texttt{languageOther}\end{aligned}$$

To use DR-estimation based on these models we type:

```
> fit <- drgee(oformula=highWage~education+age+language,
+ eformula=sex~education+age+language, iaformula=~education,
+ olink="logit", elink="logit", estimation.method="dr", data=SLID)
> summary(fit)

Call: drgee(oformula = highWage ~ education + age + language,
   eformula = sex ~ education + age + language,
   iaformula = ~education, olink = "logit", elink = "logit",
   data = SLID, estimation.method = "dr")

Outcome: highWage

Exposure: sexMale

Covariates: education,age,languageFrench,languageOther

Main model: highWage ~ sexMale + sexMale:education

Outcome nuisance model: highWage ~ education + age + languageFrench + languageOther

Outcome link function: logit

Exposure nuisance model: sexMale ~ education + age + languageFrench +
                         languageOther

Exposure link function: logit
```

```
                 Estimate Std. Error z value Pr(>|z|)
sexMale             2.9050    0.4015   7.236   4.62e-13 ***
sexMale:education  -0.1341    0.0295  -4.547   5.44e-06 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Note: The estimated parameters quantify the conditional exposure-outcome associa-
tion, given the covariates included in the nuisance models)

3987    complete observations used
```

We observe that the DR estimates are quite different from the E-estimates in Example 4 (Section 3.4). This indicates that the exposure nuisance model [16] may not be well specified.

5 Estimation with clustered data

5.1 Theory

When data are clustered, the parameter estimates obtained in Sections 2–4 are still consistent. However, their standard errors must be corrected for within-cluster correlations. Suppose that data contain m independent clusters with n_i observations $O_{i1}, ..., O_{in_i}$ in cluster i for $i = 1, ..., m$. Corrected standard errors can then be obtained by replacing the generic estimating function $\psi(O_{ij}, \theta)$ in the sandwich formula with $\sum_{j=1}^{n_i} \psi(O_{ij}, \theta)$ for $i = 1, ..., n_i$. This modification corrects the standard errors, but does not affect the parameter estimates.

5.2 Example 7

To demonstrate estimation with clustered data, we use the dataset `ohio` in `geepack` package (Højsgaard et al., 2006). This dataset contains data to study the health effects of air pollution on children. The children were examined annually at ages 7–10. Four variables are included in the dataset: `resp`, `smoke`, `age` and `id`. The variables `resp` (binary), `smoke` (binary) and age (continuous) indicate wheezing status, maternal smoking and age, respectively, at each examination. `id` is the child's identification number. Suppose that we are interested in the association between maternal smoking and wheezing status conditional on age. Since all subjects are from the same city, we assume that the effect of air pollution is similar for all subjects in the sample. Assuming the main model

$$\text{logit}\{E(\texttt{resp}|\texttt{smoking}, \texttt{age})\} - \text{logit}\{\text{E}(\texttt{resp}|\texttt{smoking} = 0, \texttt{age})\} = \beta \cdot \texttt{smoking},$$

the outcome nuisance model

$$\text{logit}\{E(\texttt{resp}|\texttt{smoking} = 0, \texttt{age})\} = \gamma_0 + \gamma_1 \cdot \texttt{age},$$

and the exposure nuisance model

$$\text{logit}\{E(\texttt{smoking}|\texttt{resp} = 0, \texttt{age})\} = \alpha_0 + \alpha_1 \cdot \texttt{age},$$

we can perform DR-estimation with cluster-corrected standard errors by setting the argument `clusterid` to "`id`":

```
> library(geepack)
> data(ohio)
> fit <- drgee(oformula=resp~age,eformula=smoke~age,
+ olink="logit",elink="logit", estimation.method="dr",
+ data=ohio, clusterid="id")
> summary(fit)
```

```
Call: drgee(oformula = resp ~ age, eformula = smoke ~ age,
   olink = "logit", elink ="logit", data = ohio,
   estimation.method = "dr", clusterid="id")

Outcome: resp

Exposure: smoke

Covariates: age

Main model: resp ~ smoke

Outcome nuisance model: resp ~ age

Outcome link function: logit

Exposure nuisance model: smoke ~ age

Exposure link function: logit

        Estimate  Std. Error  z value  Pr(>|z|)

smoke    0.2721      0.1781    1.528   0.127

(Note: The estimated parameters quantify the conditional exposure-outcome
association, given the covariates included in the nuisance models)

 2148  complete observations used

Cluster-robust Std. errors
using  537 clusters defined by levels of id
```

6 Simulation studies

To further demonstrate the capabilities of the `drgee` package and to compare the estimation methods, we carried out two simulation studies.

6.1 Simulation study 1

In the first simulation we generated data from the following model:

$$\left.\begin{array}{rcl} L_1, L_2 \; i.i.d & \sim & N(0,1) \\ L & = & (L_1, L_2) \\ A|L & \sim & Be\{E(A|L)\} \\ \text{logit}\{E(A|L)\} & = & \alpha_0 + \alpha_1 L_1 + \alpha_2 L_2 \\ Y|A,L & \sim & N\{E(Y|A,L),1\} \\ E(Y|A,L) & = & \beta A + \gamma_0 + \gamma_1 L_1 + \gamma_2 L_2 + \gamma_3 L_1 L_2 \\ \beta & = & 1.5 \\ (\gamma_0, \gamma_1, \gamma_2, \gamma_3) & = & (-1, -1, -1, 1.5) \\ (\alpha_0, \alpha_1, \alpha_2) & = & (0.5, 1, 1) \end{array}\right\}$$

Under this model, we generated 1,000 samples with 500 observations each. Each sample was analyzed with O-estimation, E-estimation and DR-estimation. For each estimation method we carried out four analyses. In the first analysis, β was estimated with both nuisance models correctly specified. In the second analysis we used the misspecified outcome nuisance model

$$E(Y|A=0,L)=\gamma_0+\gamma_1 L_1. \quad [20]$$

In the third analysis we used the misspecified exposure nuisance model

$$\text{logit}\{E(A|L)\}=\alpha_0+\alpha_1 L_1. \quad [21]$$

In the fourth analysis we used both misspecified nuisance models [20] and [21]. For each estimation method, we calculated the mean (over the 1,000 samples) estimate of β, the mean standard error, and the empirical coverage probability of a 95% Wald confidence interval for β. For comparison we also calculated the empirical standard error, i.e. the standard deviation of the estimate over the 1,000 samples. The results are shown in Table 1.

Table 1: Comparison of estimation methods. I: both nuisance models correctly specified, II: outcome nuisance model misspecified, III: exposure nuisance model misspecified, IV: both nuisance models misspecified.

		Mean estimate	Mean standard error	Empirical standard error	Empirical coverage probability of 95% CI
I	$\hat{\beta}_O$	1.497	0.106	0.107	0.941
	$\hat{\beta}_E$	1.503	0.167	0.173	0.946
	$\hat{\beta}_{DR}$	1.497	0.107	0.109	0.939
II	$\hat{\beta}_O$	0.522	0.200	0.205	0.004
	$\hat{\beta}_E$	1.503	0.167	0.173	0.946
	$\hat{\beta}_{DR}$	1.503	0.167	0.173	0.946
III	$\hat{\beta}_O$	1.497	0.106	0.107	0.941
	$\hat{\beta}_E$	0.519	0.200	0.205	0.003
	$\hat{\beta}_{DR}$	1.497	0.106	0.108	0.940
IV	$\hat{\beta}_O$	0.522	0.200	0.205	0.004
	$\hat{\beta}_E$	0.519	0.200	0.205	0.003
	$\hat{\beta}_{DR}$	0.519	0.200	0.205	0.003

In the first analysis, all three mean estimates are close to the true value 1.5. This demonstrates that all three methods give unbiased estimates of β when all modelling assumptions are correct. The standard error is larger for the E-estimator than for the O-estimator. This is not a coincidence; Robins et al. (1992) showed that the O-estimator is always at least as efficient as the E-estimator. Note that the standard error of the DR estimator is close to the standard error of the O-estimator. In the second analysis, the assumed outcome nuisance model is misspecified, leading to bias in the O-estimator. The E-estimator does not use the outcome nuisance model and is therefore unaffected by this misspecification. Despite using a misspecified model for the outcome nuisance, the DR estimator is unbiased. Furthermore, its standard error is nearly identical to the standard error of the E-estimator. In the third analysis, the assumed exposure nuisance model is misspecified, leading to bias in the E-estimator. The DR estimator is unbiased for β and its standard error is nearly identical to the standard error of the O-estimator. In the fourth analysis, both nuisance models are misspecified, leading to bias for all three methods. In this analysis, both the mean estimates and standard errors are very similar for all three methods. This demonstrates that agreement between the three methods does not imply that both nuisance models are correctly specified. We finally observe that the mean standard errors agree well with the empirical standard errors for all methods and analyzes, and that the 95% confidence interval has close to 95% coverage probability as long as the corresponding estimator is unbiased.

6.2 Simulation study 2

In the second simulation we generated data from the following model:

$$\left.\begin{aligned} L_1, L_2 \ i.i.d &\sim N(0,1) \\ L &= (L_1, L_2) \\ Y|A=0,L &\sim Be\{E(Y|A=0,L)\} \\ \text{logit}\{E(Y|A=0,L)\} &= \gamma_0 + \gamma_1 L_1 + \gamma_2 L_2 + \gamma_3 L_1 L_2 \\ A|Y=0,L &\sim Be\{E(A|Y=0,L)\} \\ \text{logit}\{E(A|Y=0,L)\} &= \alpha_0 + \alpha_1 L_1 + \alpha_2 L_2 + \alpha_3 L_1 L_2 \\ \text{logit}\{E(Y|A,L)\} - \text{logit}\{E(Y|A=0,L)\} &= \beta_1 A + \beta_2 A L_1 \\ (\beta_1, \beta_2) &= (1.5, 1) \\ (\gamma_0, \gamma_1, \gamma_2, \gamma_3) &= (-1, -1, -1, 1.5) \\ (\alpha_0, \alpha_1, \alpha_2, \alpha_3) &= (-1, 1, 1, -1.5) \end{aligned}\right\}$$

Under this model, we generated 1,000 samples with 500 observations each. Each sample was analyzed with O-estimation, E-estimation and DR-estimation. For each estimation method we carried out four analyses. In the first analysis, both outcome and exposure nuisance models were correct. In the second analysis, we used the misspecified outcome nuisance model

$$\text{logit}\{E(Y|A=0,L)\} = \gamma_0 + \gamma_1 L_1 + \gamma_2 L_2. \quad [22]$$

In the third analysis, we used the misspecified exposure nuisance model

$$\text{logit}\{E(A|Y=0,L)\} = \alpha_0 + \alpha_1 L_1 + \alpha_2 L_2. \quad [23]$$

In the fourth analysis we used both misspecified nuisance models [22] and [23]. We calculated the same summary statistics as in the first simulation. The results are shown in Table 2.

The results in Table 2 again demonstrate that the DR estimator is indeed doubly robust, i.e. that it is unbiased as long as at least one of the nuisance models is correctly specified. For the main effect β_1, the standard error of the DR estimator is very similar to the standard errors of the O- and E-estimators. However, for the interaction β_2, the standard errors of the DR estimator is generally larger than the standard errors of the O- and E-estimators.

7 Discussion

In this paper we have summarized the theory behind O-estimation, E-estimation and DR estimation in restricted mean models. We have have also described, through practical examples, how the `drgee` package in `R` can be used to perform these three types of estimations. Finally, we have carried out a simulation study to compare the estimation methods, and to demonstrate the doubly robustness property of the DR estimator.

There are other `R` packages available for doubly robust estimation, e.g. `tmle`, `ltmle`, `iWeigReg`, and `multiPIM`. However, these packages target a different parameter, namely the marginal (over covariates) exposure effect, whereas our package targets the conditional (on covariates) exposure effect.

Epidemiology is a rapidly evolving field, and it is desirable that applied epidemiologists use the best methods available when analyzing data. However, in our experience epidemiologists often resort to suboptimal standard methods, due to the lack of up-to-date software. We believe that the `drgee` package fills an important gap between theory and practice, and that it will facilitate the use of DR-estimation in the epidemiologic field.

Table 2: Comparison of estimation methods. I: both nuisance models correctly specified, II: outcome nuisance model misspecified, III: exposure nuisance model misspecified, IV: both nuisance models misspecified.

		Mean estimate	Mean estimated standard error	Empirical standard error	Empirical coverage probability of 95% CI
I	$\widehat{\beta}_{1\,O}$	1.528	0.269	0.266	0.961
	$\widehat{\beta}_{2\,O}$	1.020	0.283	0.283	0.940
	$\widehat{\beta}_{1\,E}$	1.534	0.275	0.272	0.952
	$\widehat{\beta}_{2\,E}$	1.034	0.330	0.337	0.948
	$\widehat{\beta}_{1\,DR}$	1.538	0.280	0.280	0.958
	$\widehat{\beta}_{2\,DR}$	1.039	0.386	0.394	0.950
II	$\widehat{\beta}_{1\,O}$	0.777	0.241	0.240	0.158
	$\widehat{\beta}_{2\,O}$	1.278	0.248	0.251	0.819
	$\widehat{\beta}_{1\,E}$	1.534	0.275	0.272	0.952
	$\widehat{\beta}_{2\,E}$	1.034	0.330	0.337	0.948
	$\widehat{\beta}_{1\,DR}$	1.538	0.281	0.277	0.960
	$\widehat{\beta}_{2\,DR}$	1.038	0.372	0.379	0.951
III	$\widehat{\beta}_{1\,O}$	1.528	0.269	0.266	0.961
	$\widehat{\beta}_{2\,O}$	1.020	0.283	0.283	0.940
	$\widehat{\beta}_{1\,E}$	0.720	0.245	0.243	0.119
	$\widehat{\beta}_{2\,E}$	1.502	0.338	0.358	0.711
	$\widehat{\beta}_{1\,DR}$	1.531	0.276	0.273	0.961
	$\widehat{\beta}_{2\,DR}$	1.052	0.379	0.395	0.949
IV	$\widehat{\beta}_{1\,O}$	0.777	0.241	0.240	0.158
	$\widehat{\beta}_{2\,O}$	1.278	0.248	0.251	0.819
	$\widehat{\beta}_{1\,E}$	0.720	0.245	0.243	0.119
	$\widehat{\beta}_{2\,E}$	1.502	0.338	0.358	0.711
	$\widehat{\beta}_{1\,DR}$	0.785	0.248	0.245	0.187
	$\widehat{\beta}_{2\,DR}$	1.057	0.333	0.348	0.939

Appendix: Theory for estimating equations

In this Appendix we briefly review some basic theory for estimating equations, we refer to Newey and McFadden (1994) for a more detailed exposition.

Suppose that we are interested in a parameter θ indexing a (semi)parametric model P_θ of the distribution underlying some random vector O and that there is a vector valued function $\psi(O;\theta)$ such that $E\{\psi(O;\theta_0)\} = 0$ for some unique value θ_0. Under suitable regularity conditions, a consistent and asymptotically normal estimator $\hat{\theta}_n$ of θ_0 can be obtained by solving the estimating equation

$$\hat{E}\{\psi(O;\theta)\} = 0$$

for θ, where $O_1, \ldots, O_n$ are independent observations of O. The asymptotic covariance of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ is given by the "sandwich formula" as

$$\Sigma = \left\{E\left(\frac{\partial\psi}{\partial\theta^T}_{|\theta=\theta_0}\right)\right\}^{-1} var\{\psi(\theta_0)\}\left[\left\{E\left(\frac{\partial\psi}{\partial\theta^T}_{|\theta=\theta_0}\right)\right\}^{-1}\right]^T. \quad [24]$$

The inner element on the right-hand side is referred to as the "meat", and the outer elements are refererred to as the "bread". By substituting $\hat{\theta}_n$ for θ_0 and sample mean and covariance for their population counterparts in eq. [24], we can obtain a consistent estimator for Σ.

Note that the asymptotic properties of the estimator $\hat{\theta}_n$ do not depend on the correctness of the model P_θ. If the model is correctly specified, the uniqueness of θ_0 ensures that $\hat{\theta}_n$ is consistent and asymptotically normal for the true value of θ in the model. If the model is misspecified, $\hat{\theta}_n$ is still consistent and asymptotically normal estimator for θ_0, defined as the solution to $E\{\psi(O;\theta_o)\} = 0$, where the expectation is taken with respect to the true underlying distribution.

Funding: This work was funded by the Swedish Research Council (grant/award no.: 340-2012-6007).

References

Bang, H., and Robins, J. (2005). Doubly robust estimation in missing data and causal inference models. Biometrics, 61:962–973.

Fox, J., and Weisberg, S. (2011). An R Companion to Applied Regression. 2nd Edition. Thousand Oaks, CA: Sage. http://socserv.socsci.mcmaster.ca/jfox/Books/Companion.

Højsgaard, S., Halekoh, U., and Yan, J. (2006). The R package geepack for generalized estimating equations. Journal of Statistical Software, 15:1–11.

Newey, W., and McFadden, D. (1994). Large sample estimation and hypothesis testing. Handbook of Econometrics, 4:2111–2245.

Robins, J., Mark, S., and Newey, W. (1992). Estimating exposure effects by modelling the expectation of exposure conditional on confounders. Biometrics, 48:479–495.

Robins, J., and Rotnitzky, A. (2001). Comment on 'inference for semiparametric models: Some questions and an answer,' by Bickel and Kwon. Statistica Sinica, 11:920–936.

Tchetgen Tchetgen, E., Robins, J., and Rotnitzky, A. (2010). On doubly robust estimation in a semiparametric odds ratio model. Biometrika, 97:171–180.

Tchetgen Tchetgen, E., and Rotnitzky, A. (2011). Double-robust estimation of an exposure-outcome odds ratio adjusting for confounding in cohort and case-control studies. Statistics in Medicine, 30:335–347.

Valeri, L., and VanderWeele, T. (2013). Mediation analysis allowing for exposure–mediator interactions and causal interpretation: Theoretical assumptions and implementation with SAS and SPSS macros. Psychological Methods, 18:137.

Zetterqvist, J., and Sjölander, A. (2015). drgee: doubly robust generalized estimating equations. http://CRAN.R-project.org/package = drgee, version 1.1.3.