



---

Education corner

## Estimating predicted probabilities from logistic regression: different methods correspond to different target populations

Clemma J Muller\* and Richard F MacLehose

Division of Epidemiology, University of Minnesota, Minneapolis, MN, USA

\*Corresponding author. Division of Epidemiology, University of Minnesota, West Bank Office Building, 1300 S. Second St., STE 300, Minneapolis, MN 55454-1015. E-mail: jaco0484@umn.edu

Accepted 20 January 2014

### Abstract

**Background:** We review three common methods to estimate predicted probabilities following confounder-adjusted logistic regression: marginal standardization (predicted probabilities summed to a weighted average reflecting the confounder distribution in the target population); prediction at the modes (conditional predicted probabilities calculated by setting each confounder to its modal value); and prediction at the means (predicted probabilities calculated by setting each confounder to its mean value). That each method corresponds to a different target population is underappreciated in practice. Specifically, prediction at the means is often incorrectly interpreted as estimating average probabilities for the overall study population, and furthermore yields nonsensical estimates in the presence of dichotomous confounders. Default commands in popular statistical software packages often lead to inadvertent misapplication of prediction at the means.

**Methods:** Using an applied example, we demonstrate discrepancies in predicted probabilities across these methods, discuss implications for interpretation and provide syntax for SAS and Stata.

**Results:** Marginal standardization allows inference to the total population from which data are drawn. Prediction at the modes or means allows inference only to the relevant stratum of observations. With dichotomous confounders, prediction at the means corresponds to a stratum that does not include any real-life observations.

**Conclusions:** Marginal standardization is the appropriate method when making inference to the overall population. Other methods should be used with caution, and prediction at the means should not be used with binary confounders. Stata, but not SAS, incorporates simple methods for marginal standardization.

**Key words:** Bias, logistic regression, risk, predicted probabilities, standardization, target population

---

### Key Messages

- Common methods for predicting probabilities from multivariable logistic regression result in estimates for different target populations; this distinction is often unappreciated by researchers and analysts.
- When making inference to the overall population with categorical confounders and common outcomes in one or more covariate strata, marginal standardization is preferred over conditional prediction methods that set all confounders to a specified value (e.g. the mean).
- With categorical confounders, the stratum defined by setting each confounder to its mean value is typically not relevant to any real-life group of individuals. Conditional predicted probabilities estimated by this method can be severely biased when the goal is to estimate marginal effects.
- Default methods in statistical software packages (e.g. 'lsmeans' in SAS) may lead to inadvertent estimation of conditional probabilities.
- Marginal methods analogous to standardization are easily implemented with the 'margins' command in Stata, and a macro has been previously published for SAS.

## Introduction

Epidemiologists often wish to estimate the risk of an outcome in one group of people compared with a referent group. In observational analyses, these comparisons are typically adjusted for one or more confounding factors. In the simplest scenario, with binary exposure, binary outcome and a small number of categorical covariates, standardization is an easy and intuitive approach for covariate adjustment and requires transparent specification of a clearly defined target population.<sup>1–3</sup> Risk estimates for the exposed and unexposed groups can be standardized to confounder distributions in the exposed, unexposed, overall or some external population,<sup>4</sup> and can then be compared using a risk ratio or risk difference.

Standardized estimates become unstable or impossible to compute as more covariates are considered for adjustment, or when any covariates are continuous. In this scenario, regression modelling is an attractive alternative to standardization; the estimate obtained from these models is often closest to the estimate one would compute by standardizing to the total population.<sup>5</sup> For dichotomous outcomes, logistic regression is the overwhelming choice for analysis of observational and experimental data. Odds ratios are easily obtained from logistic models, but the relative risk is a more intuitive multiplicative measure of effect and is collapsible over covariate strata.<sup>6</sup> Furthermore, the odds ratio overestimates the relative risk for common outcomes, though they are often misinterpreted as being equivalent.<sup>7</sup> On the additive scale, the risk difference is more causally relevant and more readily justified as a measure of average causal effect, particularly in the presence of interactions.<sup>8,9</sup>

The most straightforward approach to estimate risk-based associations is to apply statistical methods that

directly model the relative risk or risk difference.<sup>10</sup> For example, a binomial family generalized linear model with a log link will allow direct estimation of the relative risk.<sup>5,11</sup> Alternatively, by specifying the identity link, one can directly estimate the risk difference. These approaches are limited, however, by the possibility of obtaining predicted probabilities that are greater than 100% using the log link, and less than 0% or greater than 100% using the identity link. Perhaps more troubling, these models frequently do not converge in practice.<sup>12</sup> Solutions include restricting the space of possible parameter values,<sup>13</sup> using Poisson regression to model the relative risk<sup>14,15</sup> and using ordinary least squares regression to model the risk difference.<sup>16,17</sup> These solutions avoid concerns of non-convergence, but have also been criticized for reduced efficiency and the potential for impossible predicted probability values.<sup>18</sup>

An often overlooked approach for estimating the relative risk or risk difference is to use a logistic model but avoid directly interpreting the coefficients. Instead, inference may be based on predicted probabilities calculated from the model, which are appropriately constrained to fall between 0% and 100%.<sup>12,19</sup> Calculating measures of effect from predicted probabilities following logistic regression is a straightforward generalization of common standardization techniques,<sup>18</sup> and allows incorporation of multiple categorical and continuous confounders. Three methods for combining the predicted probabilities are common in the literature: (i) marginal standardization,<sup>12,20–22</sup> in which predicted probabilities of the outcome are calculated for every observed confounder value and then combined as a weighted average separately for each exposure level;<sup>23,24</sup> (ii) prediction at the modes,<sup>2,20</sup> in which conditional predicted probabilities are calculated for each exposure level with every confounder fixed at its

most common value,<sup>20,22,23</sup> and (iii) prediction at the means,<sup>2,20</sup> in which conditional predicted probabilities are calculated for each exposure level with every confounder fixed at its mean value.<sup>25–29</sup> Predicted probabilities from any of these methods can be contrasted to estimate the relative risk or risk difference.<sup>30,31</sup> The estimated effect measure will typically differ according to the chosen method—sometimes dramatically—and will correspond to effect measures estimated in different target populations.<sup>32</sup> A key but often underappreciated distinction is the marginal interpretation for method 1 vs the conditional interpretations for methods 2 and 3.<sup>12,32</sup> User-error from misapplication of a method that is inappropriate for the intended population of inference can yield misleading effect measure estimates. In this article we discuss the basis for and interpretation of these three methods for estimating predicted probabilities and adjusted risk comparisons, using data from an observational study of the association between physical activity and body mass index.

## Logistic regression and predicted probabilities

Logistic regression uses the logit link to model the log-odds of an event occurring. We consider a simple logistic regression with a dichotomous exposure (E) and a single dichotomous confounder (Z), but the model and results obtained below can easily be expanded to include multiple categorical or continuous confounders. Following estimation of a logistic regression model by maximum likelihood, it is straightforward to predict the probability of the outcome ( $\hat{p}_{ez}$ ) for any  $E = e$  and  $Z = z$  as follows:

$$\hat{p}_{ez} = \frac{\exp[\hat{\alpha} + \hat{\beta}_1 \times e + \hat{\beta}_2 \times z]}{1 + \exp[\hat{\alpha} + \hat{\beta}_1 \times e + \hat{\beta}_2 \times z]} \quad (1)$$

where  $\hat{\alpha}$ ,  $\hat{\beta}_1$  and  $\hat{\beta}_2$  are the estimated regression coefficients. The three methods described below contrast these predicted probabilities to estimate the relative risk or risk difference.

### Method 1: marginal standardization

Method 1 is a regression-based equivalent of the common epidemiological technique of standardization.<sup>2,20,21,33,34</sup> in which the estimate of interest (e.g. rate, prevalence or odds) is proportionally adjusted according to a weight for each level of the confounding factor(s). Assuming proper model specification sufficient to estimate the marginal effect, no uncontrolled confounding and no measurement error, this quantity is the proportion of observations with the outcome that we would have observed had we been

able to force all of the study population to exposure level  $E = e$ :

$$\Pr(Y = 1 | \text{Set}[E = e]) = \sum_z \Pr(Y = 1 | \text{Set}[E = e], Z = z) \Pr(Z = z) \quad (2)$$

where  $\text{Set}[E = e]$  reflects forcing all observations to a single exposure level  $e$ , and  $Z = z$  refers to a given set of observed values for the confounder vector  $\mathbf{Z}$ . This operationalization yields estimates standardized to the total population, but different specifications are common (e.g. the exposed population).<sup>4,35</sup> In observational research, exposures are not randomized and the quantity  $\Pr(Y = 1 | \text{Set}[E = e], Z = z)$  will not be observed, but it can be directly estimated from the data under assumptions of no uncontrolled confounding and  $\Pr(Z = z, E = e, Y = y) > 0$  for all combinations of  $z$ ,  $e$  and  $y$ . To improve estimation, we use the regression estimate from equation 1,  $\hat{p}_{ez}$ , in place of  $\Pr(Y = 1 | \text{Set}[E = e], Z = z)$  to yield:

$$\Pr(Y = 1 | \text{Set}[E = e]) = \sum_z \hat{p}_{ez} \Pr(Z = z). \quad (3)$$

Marginal probabilities obtained from equation 3 reflect a weighted average over the distribution of the confounders and are equivalent to estimates obtained by standardizing to the total population.<sup>4</sup> In the decades since its introduction to the public health literature, method 1 has subsequently been explained as a special case of g-computation methods.<sup>36–38</sup>

To apply method 1 in practice after performing a logistic regression, the exposure E is set to the (possibly counterfactual) level  $e$  for everyone in the dataset, and the logistic regression coefficients are used to calculate predicted probabilities for everyone at their observed confounder pattern and newly assigned exposure value. Following estimation of  $\Pr(Y = 1 | \text{Set}[E = e])$  for both levels of a binary exposure, it is straightforward to calculate the causal effect for exposure vs no exposure as the relative risk:

$$\text{RR} = \Pr(Y = 1 | \text{Set}[E = 1]) / \Pr(Y = 1 | \text{Set}[E = 0]) \quad (4)$$

or as the risk difference:

$$\text{RD} = \Pr(Y = 1 | \text{Set}[E = 1]) - \Pr(Y = 1 | \text{Set}[E = 0]). \quad (5)$$

Because predicted probabilities are computed under the same distribution of  $\mathbf{Z}$ , there is no confounding of the corresponding effect measure estimates. Confidence intervals are typically calculated using the delta method or bootstrap method,<sup>12,18</sup> though other approaches to variance estimation have been explored.<sup>21</sup>

### Method 2: prediction at the modes

Method 2 calculates the predicted probability of the outcome for each exposure level assuming everyone in the population had the most common values of the confounders:

$$\Pr(Y = 1 | \text{Set}[E = e], Z = z^m) \quad (6)$$

where  $z^m$  reflects the modal value(s) of confounder vector  $Z$ . In practice, the quantity in equation 6 is estimated by  $\hat{p}_{ez^m}$  following a logistic regression model. In a simple scenario with just one dichotomous confounder, we would calculate predicted probabilities only for the stratum corresponding to the most frequently observed value (0 or 1) in the population. By definition, this corresponds to the most common confounder stratum in the analysis. However, as additional variables are added to the model and each is set to its own modal value, the number of individuals with the specified covariate pattern will decrease accordingly. In the absence of extremely large samples, this can quickly result in a sparsely populated (or empty) stratum. This might not be problematic if the modal population is of substantive interest, but is ill-advised when the goal is to make inference to the overall population. Alternatively, method 2 could be modified to estimate predicted probabilities in the stratum defined by the most common joint pattern of all confounder values in the population. With this approach, the predicted probability will relate to a stratum guaranteed to contain at least some observations. In either case, effect estimates computed under method 2 control for confounding by standardizing both the exposed and unexposed populations to the same target: in this case the population of those with the modal distribution of  $Z$ .

### Method 3: prediction at the means

Method 3 calculates the predicted probability of the outcome by exposure status assuming that every person in the dataset has the mean value of each confounder. This is written as:

$$\Pr(Y = 1 | \text{Set}[E = e], Z = \bar{z}) \quad (7)$$

where  $\bar{z}$  is the observed average of each confounder in vector  $Z$ . As with methods 1 and 2, logistic regression is typically used to estimate  $\hat{p}_{e\bar{z}}$  for equation 7. This approach may appear to differ from method 2 only in the specification of the stratum used for standardization. However, for a continuous or ordinal confounder, the mean value may or may not be observed with  $\Pr(Z = \bar{z}) > 0$  in the data. For a dichotomous confounder coded 0 or 1, the observed

average corresponds to the proportion of the population with values of 1, and prediction at the means will never correspond to any observation in the data as long as  $\Pr(Z = z) > 0$  for both levels  $z$  of the confounder.<sup>32</sup> Prediction at the means may be implemented under the erroneous assumption that it estimates the average, or marginal, predicted probability for the entire population (i.e. method 1).

### Marginal standardization vs prediction at the means

In linear regression, methods 1 and 3 will yield identical results, but this equality does not hold for nonlinear models such as logistic regression.<sup>2,8</sup> Mathematical properties of this phenomenon are similar to those of the more widely appreciated non-collapsibility of odds ratios.<sup>39-41</sup> Similarly, when the outcome is rare in all confounder strata or when the linear combination of confounders in the logit model is only weakly correlated with the outcome, method 3 can approximate method 1. However, when these conditions are assumed in error, the methods can diverge dramatically.

Consider a hypothetical example in which the confounder of interest ( $Z$ ) is sex and half of the study population are men ( $Z = 1$ ) and half are women ( $Z = 0$ ). For simplicity, we only consider the probability of the outcome among the unexposed. Suppose 50% of unexposed women and 99% of unexposed men in the study population experience the outcome, with  $\hat{\alpha} = 0, \hat{\beta}_2 = 4.60$  from a logistic model fit to these data. Figure 1 depicts the corresponding

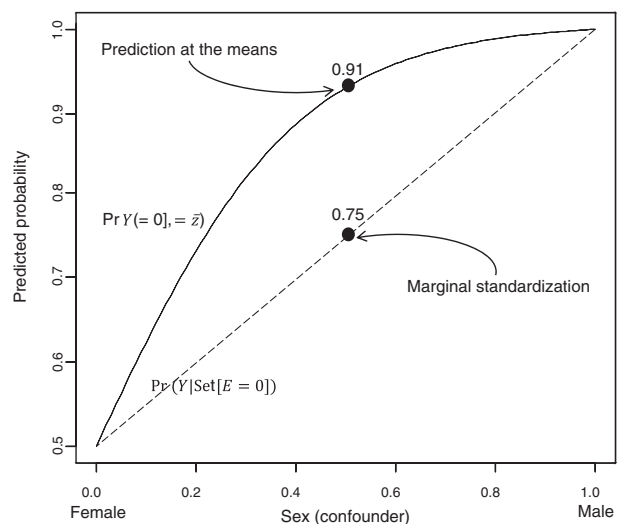


Figure 1. Half of a sigmoid curve depicting calculation of predicted probabilities following logistic regression using marginal standardization (dashed straight line) and prediction at the means (solid curved line) in unexposed people from a hypothetical population

predicted probabilities of the outcome with confounder values ranging continuously from 0 to 1. The curve demonstrates how the probability of the outcome among the unexposed increases as the dichotomous confounder, sex, varies from 0 (female) to 1 (male). Prediction at the means amounts to finding the proportion of men in the dataset, and then calculating the predicted probability of the outcome for that value as if sex were a continuous variable.

In this example, 50% of the study population are men, so  $\bar{z} = 0.5$  and we would calculate  $\Pr(Y = 1 | \text{Set}[E = 0], Z = 0.5) = \exp[4.60 \times 0.5] / (1 + \exp[4.60 \times 0.5]) = 0.91$ . Many analysts motivate this approach by assuming that the predicted probability at the mean value of sex is equivalent to the mean of the predicted probabilities for males and females. To the contrary, 0.91 reflects the predicted probability for the nonexistent group of people who are 50% male. In actuality, any mean of the male and female predicted probabilities will be a linear combination of the two (as depicted by the straight dashed line in Figure 1). Again, assuming half of the observations are female and half are male, we find that  $\sum \Pr(Y = 1 | \text{Set}[E = 0], Z = z) \Pr(Z = z) = (0.50 \times 0.5 + 0.99 \times 0.5) = 0.75$ . In this scenario, application of method 3 results in a predicted probability for a non-existent target population, and overestimates the marginal mean value (method 1) by 21%. In the following section, we present a substantive example to demonstrate this phenomenon in more detail.

## Example

Data for this example were drawn from Project EAT (Eating and Activity in Teens and Young Adults)-II and Project EAT-III, the second and third waves of a longitudinal study designed to examine dietary intake, physical activity, weight control behaviours, weight status and factors associated with these outcomes in young people. In Project EAT-I (1998–99), 2516 junior and senior high school students at 31 public schools in the Minneapolis/St. Paul metropolitan area of Minnesota completed surveys and anthropometric measures.<sup>42,43</sup> In Project-EAT-II (2003–04) and -III (2008–09) investigators followed up on 2437 and 1902 participants, respectively, to evaluate longitudinal change in behaviours and weight status.

We investigated the cross-sectional association between body mass index and a dichotomous measure of physical activity in the Project EAT-III survey. Our target population of interest is the total population of young people 16–23 years old who were enrolled in this study. The outcome, overweight/obesity, was defined as body mass index  $\geq 25 \text{ kg/m}^2$ . The exposure, low physical activity, was defined as engaging in fewer than 10 h of moderate or vigorous physical activity per week. Both the outcome and the

exposure were common: prevalence of overweight/obesity was 46.1%, and prevalence of low physical activity was 90.5%. In logistic regression models, we adjusted for confounding due to dichotomous indicators of age  $\geq 18$  years at wave 3, female sex, White race, whether the participant had dieted in the year preceding wave 2 data collection and overweight/obese body mass index as measured at wave 2. Our final sample size was 1678, after excluding observations with missing data for any relevant variable.

Table 1 gives descriptive statistics by overweight/obesity status at wave 3, as well as the mean and mode of each covariate for the combined sample. To prepare for implementing method 2, we set each confounder equal to its respective mode ( $\geq 18$  years old, female, White, did not diet in the year before wave 2, not overweight or obese at wave 2). This stratum comprised 148 observations (9% of the data) and was the third most common joint confounder pattern (Table 2). Even the most prevalent joint distribution of the dichotomous confounders (age  $\geq 18$  years, male, White race, did not diet in the year preceding wave 2 data collection, not overweight or obese at wave 2) comprised just 266 observations (16% of the data).

Table 3 shows results from the unadjusted tabular analysis, and from the three methods of calculating predicted probabilities. Stata syntax for obtaining these results is given in the Appendix (available as Supplementary data at *IJE* online). Effect measures are presented as prevalence ratios and prevalence differences. Marginal

**Table 1.** Distribution of exposure and dichotomous confounders by outcome (body mass index) among wave 3 participants of Eating and Activity in Teens and Young Adults, Minneapolis/St. Paul, MN, 2008–09

	Wave 3 body mass index <sup>a</sup>			Mean	Mode
	Overweight/ obese	Normal	Combined		
	( <i>n</i> = 773)	( <i>n</i> = 905)	( <i>n</i> = 1678)		
	%	%			
Exposure					
Low physical activity	93.3	88.1	0.9	1	
Dichotomous confounders					
Age $\geq 18$ years	81.4	75.1	0.8	1	
Female	48.0	58.5	0.5	1	
White race	66.4	70.4	0.7	1	
Dieted in the year preceding wave 2 <sup>b</sup>	50.6	34.8	0.4	0	
Overweight/obese at wave 2	60.8	5.9	0.3	0	

<sup>a</sup>Overweight/obese = body mass index  $\geq 25 \text{ kg/m}^2$ , normal = body mass index  $< 25 \text{ kg/m}^2$ .

<sup>b</sup>Wave 2 conducted 2003–04.

**Table 2.** Five most common strata defined by the joint distributions of dichotomous confounder values among wave 3 participants ( $n = 1678$ ) of Eating and Activity in Teens and Young Adults, Minneapolis/St. Paul, MN, 2008–09

Rank	Dichotomous confounder values					Distribution	
	Age $\geq 18$ years	Female	White race	Dieted in the year preceding wave 2 <sup>a</sup>	Overweight/obese <sup>b</sup> at wave 2	<i>n</i>	%
1	1	0	1	0	0	266	15.9
2	1	1	1	1	0	183	10.9
3 <sup>c</sup>	1	1	1	0	0	148	8.8
4	1	1	1	1	1	99	5.9
5	1	0	1	0	1	86	5.1

<sup>a</sup>Wave 2 conducted 2003–04.<sup>b</sup>Body mass index  $\geq 25$  kg/m<sup>2</sup>.<sup>c</sup>Boldface stratum defined by independently setting each confounder equal to its mode.**Table 3.** Predicted probabilities and effect measure estimates comparing prevalence of overweight/obesity by physical activity status among wave 3 participants ( $n = 1678$ ) of Eating and Activity in Teens and Young Adults, Minneapolis/St. Paul, MN, 2008–09

	Wave 3 overweight/obesity <sup>a</sup> predicted probability		Low physical activity vs high physical activity			
	Low physical activity %	High physical activity %	Prevalence ratio	95% CI	Prevalence difference	95% CI
Unadjusted	47.5	32.5	1.5	1.2, 1.8	15.0	7.3, 22.7
Marginal standardization <sup>b</sup>	47.3	35.1	1.3	1.1, 1.6	12.2	6.2, 18.2
Prediction at the modes <sup>b</sup>	19.3	9.4	2.1	1.2, 2.9	9.9	5.6, 14.2
Prediction at the means <sup>b</sup>	51.1	31.2	1.6	1.2, 2.1	19.9	10.2, 29.7

CI, confidence interval.

<sup>a</sup>Body mass index  $\geq 25$  kg/m<sup>2</sup>.<sup>b</sup>Adjusted for dichotomous indicators of  $\geq 18$  years old, female sex, White race, dieting in year preceding wave 2 (2003–04) and being overweight/obese (body mass index  $\geq 25$  kg/m<sup>2</sup>) at wave 2.

standardization allows inference to the total population, and we observe only slight differences of the prevalence ratio (1.3 vs 1.5) and prevalence difference (12.2% vs 15.0%) compared with the unadjusted analysis. Prediction at the modes yields a prevalence ratio (2.1) and prevalence difference (9.9%) that only pertain to people who are  $\geq 18$  years old, female, White, did not diet in the year before wave 2 and were not previously overweight or obese. Prediction at the means yields a prevalence ratio (1.6) and prevalence difference (19.9%) relevant to the non-existent target population of people who are 78% aged 18 years or older, 54% female, 69% White, 42% dieted in the year prior to wave 2 and 31% overweight/obese at wave 2.

## Discussion

For an effect measure estimate to be interpretable, the target population of interest must be clearly stated.<sup>44</sup> When estimating risk (or prevalence) ratios or differences following logistic regression, insufficient attention to this detail has likely resulted in estimation of associations for

unintended, and perhaps nonexistent, target populations. In the presence of dichotomous confounders, method 1 is the appropriate choice when the goal is to model the average association in the overall study population. In our example, method 2 estimates predicted probabilities for a stratum comprising just 9% of the sample, whereas method 3 estimates predicted probabilities for a non-existent target population. Methods 2 and 3 are therefore poor choices as surrogate estimates of the marginal association in the overall population of interest. As previously noted, method 3 can approximate marginal averages when the outcome is rare in all covariate strata or when the linear combination of confounders is uninformative, conditions which were not met in our example. Method 1 can also be combined with methods that model the exposure as a function of covariates (e.g. propensity scores) to generate doubly robust effect measure estimates, as previously described for regression models in general,<sup>45</sup> and specifically for logistic regression<sup>46,47</sup> and marginal effects estimation.<sup>48</sup> This may be especially desirable when regressing an outcome on a large number of confounders,

which intensifies concerns about bias due to model misspecification.

Whereas the marginal standardization approach of method 1 can be used to estimate probabilities for any specified target population, the other methods are less universal. Method 2 yields valid results if the goal is to restrict inference to the stratum of individuals defined by the most common confounder values. However, when even a moderate number of confounders are included in the regression, the modal stratum may reflect only a small proportion of the population. In our example the effect in the modal stratum contrasted activity levels among White women aged 18 years and older who did not diet and were not overweight or obese at wave 2. This is certainly a legitimate target population, but it may not be of particular interest to anyone, certainly not to the exclusion of all other confounder strata. However, method 2 may be appealing to epidemiologists when evidence of effect measure modification is present.

Method 3 estimates associations at the mean of each confounder in the regression model. In the presence of binary covariates, prediction at the means yields results that are not meaningful to any real-world group of individuals. Put simply, no one can be 54% female or 31% overweight/obese at wave 2. The justification for this choice of measure is often not clearly stated, but may reflect an assumption that the association computed at the means of the covariates is equivalent to the overall mean association. Unfortunately, in regression models that transform the linear predictor—such as the inverse logit, or expit, transformation in logistic regression—this is not generally true.<sup>18</sup> When calculating predicted probabilities, the inverse logit of the averages (method 3) is not equal to the average of the inverse logits (method 1). In practice, the error introduced by method 3 may be minor when the confounder has a weak association with the outcome. As we have shown, however, misapplication of prediction at the means for more informative binary confounders can result in substantially different effect measure estimates. Method 3 can produce meaningful results for continuous confounders if the mean values are possible in the population, but in these analyses inference should be restricted to the stratum defined by the covariate means.

It is important to note that although marginal effect estimation has many benefits, it should not be used to predict conditional effects for individuals or specific subgroups of the overall population. This is an especially important distinction for analyses with clinical applications, such as selecting the best treatment for a specific patient. Instead, investigators in these situations may prefer conditional prediction models. However, even when models are

conditioned on some covariates, it may be desirable to marginalize over others.

### Statistical software

For implementing method 1, Stata (StataCorp, College Station, TX) is the most user-friendly.<sup>31</sup> In versions 11 and later the ‘margins’ command defaults to the population-averaged estimates, as shown in the Appendix for the total and the exposed populations (available as Supplementary data at *IJE* online). Confidence intervals can be calculated by bootstrap or the delta method, with the latter generated automatically by the ‘margins’ command.<sup>12,31</sup> For method 2, users must individually specify the modal value for each confounder. The ‘atmeans’ option easily generates results for method 3. In Stata versions 10 and earlier, the ‘adjust’ command can be used to estimate predicted probabilities for all three methods, but the default is prediction at the means and more complicated programming is necessary to implement method 1.<sup>12</sup>

To our knowledge, there is no simple way to obtain predicted probabilities corresponding to method 1 in SAS (SAS Institute Inc., Cary, NC). The LSMEANS option appended to PROC GENMOD seems most intuitive, but does not easily produce anything other than results reflecting prediction at the means. Even the OM (observed marginals) option simply changes the default assumption of perfect balance between confounder levels to the observed proportions (i.e. prediction at the means) for confounders identified as categorical in the CLASS statement.<sup>30</sup> We refer readers who wish to apply marginal standardization using SAS to a macro described elsewhere.<sup>49</sup> The PROC RLOGIST command in SAS-callable SUDAAN (Research Triangle Institute, Research Triangle Park, NC) can implement method 1 (syntax available from the authors), and SUDAAN methods have been published that describe marginal prediction to estimate risk ratios and risk differences following logistic regression with complex survey data.<sup>50</sup>

### Conclusion

To estimate an unbiased risk ratio or risk difference from confounder-adjusted logistic regression, the method of calculating predicted probabilities must align with the target population of interest. We have reviewed the applications of and inferential target populations for three methods commonly used in the literature. The most intuitive and meaningful approach, accomplished by weighted averaging of the predicted probabilities for each observed covariate profile, is analogous to standardization to the total population. Prediction at the modes yields valid results, but inference is restricted to the single stratum defined by the most

common covariate values. Prediction at the means is not meaningful in the presence of dichotomous covariates, and can create substantial errors when confounders are strongly associated with the outcome and when the effect measure estimate is interpreted as representing the population average. Unfortunately, in statistical software packages that include commands to extract predicted probabilities from logistic models, the default approach and corresponding target population are not always clearly defined in the software documentation. Specifically, default commands in SAS and Stata can lead to inadvertent misapplication of prediction at the means. When adjusting for dichotomous confounders, marginal standardization is the appropriate method for making inference to the overall source population from which the study sample was drawn, and we hope the syntax provided in the Appendix will facilitate its use (available as Supplementary data at *IJE* online).

## Supplementary Data

Supplementary data are available at *IJE* online.

## Funding

This work was supported by a grant from the National Institutes of Health (grant number 1U01-HD061940).

## Acknowledgements

Our grateful thanks to Jay Kaufman for providing feedback and advice on early drafts. We also thank anonymous reviewers for their helpful critiques and suggestions for improving the manuscript.

Guarantor: Both authors will serve as guarantors for the content of this manuscript.

**Conflict of interest:** None declared.

## References

- Rothman KJG, Greenland S, Lash TL. Chapter 3, Chapter 4. *Modern Epidemiology*. 3rd edn. Philadelphia, PA: Lippincott-Raven, 2008.
- Wilcosky TC, Chambless LE. A comparison of direct adjustment and regression adjustment of epidemiologic measures. *J Chronic Dis* 1985;**38**:849–56.
- Szklo M, Nieto FJ. Chapter 7. *Epidemiology: Beyond the Basics*. 2nd edn. Sudbury, MA: Jones & Bartlett, 2007.
- Sato T, Matsuyama Y. Marginal structural models as a tool for standardization. *Epidemiology* 2003;**14**:680–86.
- Greenland S, Maldonado G. The interpretation of multiplicative-model parameters as standardized parameters. *Stat Med* 1994;**13**:989–99.
- Greenland S. Absence of confounding does not correspond to collapsibility of the rate ratio or rate difference. *Epidemiology* 1996;**7**:498–501.
- Cummings P. The relative merits of risk ratios and odds ratios. *Arch Pediatr Adolesc Med* 2009;**163**:438–45.
- Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol* 1987;**125**: 761–68.
- Greenland S, Poole C. Invariants and noninvariants in the concept of interdependent effects. *Scand J Work, Environ Health* 1988;**14**:125–29.
- Austin PC, Laupacis A. A tutorial on methods to estimating clinically and policy-meaningful measures of treatment effects in prospective observational studies: a review. *Int J Biostat* 2011;**7**:6.
- Lumley T, Kronmal RA, Ma S. *Relative Risk Regression in Medical Research: Models, Contrasts, Estimators, and Algorithms*. University of Washington Biostatistics Working Paper. Seattle, WA: University of Washington, 2006.
- Localio AR, Margolis DJ, Berlin JA. Relative risks and confidence intervals were easily computed indirectly from multivariable logistic regression. *J Clin Epidemiol* 2007;**60**:874–82.
- Wacholder S. Binomial regression in GLIM: estimating risk ratios and risk differences. *Am J Epidemiol* 1986;**123**:174–84.
- Spiegelman D, Hertzmark E, Wand HC. Point and interval estimates of partial population attributable risks in cohort studies: examples and software. *Cancer Causes Control*;2007;**18**:571–79.
- Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. *Am J Epidemiol* 2005;**162**: 199–200.
- Woolridge JM. *Introductory Econometrics: A Modern Approach*. 4th edn. Stamford, CT: South Western Cengage learning, 2009.
- Cheung YB. A modified least-squares regression approach to the estimation of risk difference. *Am J Epidemiol* 2007;**166**: 1337–44.
- Greenland S. Model-based estimation of relative risks and other epidemiologic measures in studies of common outcomes and in case-control studies. *Am J Epidemiol* 2004;**160**:301–05.
- Cummings P. Methods for estimating adjusted risk ratios. *Stata J* 2009;**9**:175–96.
- Lane PW, Nelder JA. Analysis of covariance and standardization as instances of prediction. *Biometrics* 1982;**38**:613–21.
- Joffe MM, Greenland S. Standardized estimates from categorical regression models. *Stat Med* 1995;**14**:2131–41.
- Rothman KJ, Greenland S, Lash TL. Chapter 21. *Modern Epidemiology*. 3rd edn. Philadelphia, PA: Lippincott-Raven, 2008.
- Berge JM, MacLehose R, Eisenberg ME, Laska MN, Neumark-Sztainer D. How significant is the ‘significant other’? Associations between significant others’ health behaviors and attitudes and young adults’ health outcomes. *Int J Behav Nutr Phys Activity* 2012;**9**:35.
- Gaieski DF, Mikkelsen ME, Band RA *et al*. Impact of time to antibiotics on survival in patients with severe sepsis or septic shock in whom early goal-directed therapy was initiated in the emergency department. *Crit Care Med* 2010;**38**:1045–53.
- Bolling SF, Li S, O’Brien SM, Brennan JM, Prager RL, Gammie JS. Predictors of mitral valve repair: clinical and surgeon factors. *Ann Thorac Surg* 2010;**90**:1904–11; discussion 12.
- Denham M, Schell LM, Deane G, Gallo MV, Ravenscroft J, DeCaprio AP. Relationship of lead, mercury, mirex, dichlorodiphenyldichloroethylene, hexachlorobenzene, and



- polychlorinated biphenyls to timing of menarche among Akwesasne Mohawk girls. *Pediatrics* 2005; **115**: e127-34.
27. Mohanty SK. Multiple deprivations and maternal care in India. *Int J Perspect Sex Reprod Health* 2012; **38**:6-14.
  28. Vadiveloo MK, Dixon LB, Elbel B. Consumer purchasing patterns in response to calorie labeling legislation in New York City. *Int J Behav Nutr Phys Activity* 2011; **8**:51.
  29. Wall MM, Larson NI, Forsyth A *et al*. Patterns of obesogenic neighborhood features and adolescent weight: a comparison of statistical approaches. *Am J Prev Med* 2012; **42**:e65-75.
  30. The GENMOD procedure. In: *SAS/STAT® 922 User's Guide*. Cary, NC: SAS Institute, 2010.
  31. Williams R. Using the margins command to estimate and interpret adjusted predictions and marginal effects. *Stata J* 2012; **12**:308-31.
  32. Hanmer MJ, Kalkan KO. Behind the curve: clarifying the best approach to calculating predicted probabilities and marginal effects from limited dependent variable models. *Am J Polit Sci* 2013; **57**:263-77.
  33. Kleinman LC, Norton EC. What's the risk? A simple approach for estimating adjusted risk measures from nonlinear models including logistic regression. *Health Serv Res* 2009; **44**:288-302.
  34. Hernan MA, Robins JM. *Causal Inference, Section 13.3*. 2013. <http://www.hsph.harvard.edu/miguel-hernan/causal-inference-book/> (9 January 2014, date last accessed)
  35. Greenland S, Holland P. Estimating standardized risk differences from odds ratios. *Biometrics* 1991; **47**:319-22.
  36. Ahern J, Hubbard A, Galea S. Estimating the effects of potential public health interventions on population disease burden: a step-by-step illustration of causal inference methods. *Am J Epidemiol* 2009; **169**:1140-47.
  37. Fleischer NL, Fernald LC, Hubbard AE. Estimating the potential impacts of intervention from observational data: methods for estimating causal attributable risk in a cross-sectional analysis of depressive symptoms in Latin America. *J Epidemiol Community Health* 2010; **64**:16-21.
  38. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *Am J Epidemiol* 2011; **173**:731-38.
  39. Newman SC. Commonalities in the classical, collapsibility and counterfactual concepts of confounding. *J Clin Epidemiol* 2004; **57**:325-29.
  40. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Stat Sci* 1999; **14**:29-46.
  41. Miettinen OS, Cook EF. Confounding: essence and detection. *Am J Epidemiol* 1981; **114**:593-603.
  42. Neumark-Sztainer D, Croll J, Story M, Hannan PJ, French SA, Perry C. Ethnic/racial differences in weight-related concerns and behaviors among adolescent girls and boys: findings from Project EAT. *J Psychosom Res* 2002; **53**:963-74.
  43. Neumark-Sztainer D, Story M, Hannan PJ, Croll J. Overweight status and eating patterns among adolescents: where do youths stand in comparison with the healthy people 2010 objectives? *Am J Public Health* 2002; **92**:844-51.
  44. Maldonado G, Greenland S. Estimating causal effects. *Int J Epidemiol* 2002; **31**:422-29.
  45. Lunceford JK, Davidian M. Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. *Stat Med* 2004; **23**:2937-60.
  46. Tchetgen Tchetgen EJ, Rotnitzky A. Double-robust estimation of an exposure-outcome odds ratio adjusting for confounding in cohort and case-control studies. *Stat Med* 2011; **30**:335-47.
  47. Tchetgen Tchetgen EJ. On a closed-form doubly robust estimator of the adjusted odds ratio for a binary exposure. *Am J Epidemiol* 2013; **177**:1314-16.
  48. Funk MJ, Westreich D, Wiesen C, Sturmer T, Brookhart MA, Davidian M. Doubly robust estimation of causal effects. *Am J Epidemiol* 2011; **173**:761-67.
  49. Zou GY. Assessment of risks by predicting counterfactuals. *Stat Med* 2009; **28**:3761-81.
  50. Bieler GS, Brown GG, Williams RL, Brogan DJ. Estimating model-adjusted risks, risk differences, and risk ratios from complex survey data. *Am J Epidemiol* 2010; **171**:618-23.