Logistic regression: When can we do what we think we can do?[1]

May 29th 2017

Maarten L. Buis,

University of Konstanz,

Department of History and Sociology

maarten.buis@uni.konstanz.de

Abstract

In a much cited overview article Mood (2010) criticized many of the ways in which the raw coefficients and odds ratios from logistic regression have been used. However, one interpretation of logistic regression is that it has an unusual dependent variable: a probability, which measures how certain we are that an event of interest happens. This degree of certainty is a function of how much information we have, which in case of logistic regression is captured by the variables we add to the model. If the dependent variable is interpreted in that way many of the problems with logistic regression pointed out by Mood (2010) turn out to be desirable properties of the logistic regression model.

Introduction

Since the appearance of the overview article by Mood (2010) there has been a growing concern within sociology that the odds ratio can no longer be used in research. In particular Mood (2010) concluded that:

---

[1] I thank Michelle V. Jackson, Sebastian E. Wenz, Frank Popham, and Richard T. Campbell for useful comments.

1. "It is problematic to interpret log-odds ratios or odds ratios as substantive effects, because they also reflect unobserved heterogeneity

2. It is problematic to compare log-odds ratios or odds ratios across models with different independent variables, because the unobserved heterogeneity is likely to vary across models.

3. It is problematic to compare log-odds ratios and odds ratios across samples, across groups within samples, or over time – even when we use models with the same independent variables – because the unobserved heterogeneity can vary across the compared samples, groups, or points in time."

In this article I will argue that the first and the last of these problems actually represent desirable properties of the logistic regression model if one interprets the dependent variable, the probability, as an assessment of how likely it is that the event of interest happens.

An example of a set of topics where this interpretation makes sense is inequality of opportunity. In such studies we want to quantify how likely we think it is that someone with a given background attains a desirable position, and the extent to which these assessments differ across backgrounds measures inequality of opportunity. However, there are also problems where this interpretation is not helpful. For example, if one has a (rational choice) theory involving a latent propensity that guides individual choices, and one is interested in structural effects on this latent propensity, then it makes more sense to interpret the results in terms of that latent propensity rather than in terms of odds or probabilities. So one cannot arbitrarily choose the most convenient interpretation of the dependent variable; it is a choice that is predetermined by the research question.

I will start with a general description of the logistic regression model, followed by a description of the problems discussed by Mood (2010) and others (for example: Allison 1999; Auspurg and Hinz 2011; Gail, Wieand, and Piantadosi 1984; Karlson, Holm, and Breen 2012; Lee 1982; Neuhaus and Jewell 1993; Norton 2012; Pang, Kaufman, and Platt 2013; Williams 2009; Wooldridge 2010). I will then

discuss why many of these "problems" are not problems at all if one interprets the dependent variable as a degree of plausibility[2], and when such an interpretation could be useful.

Logistic regression

There are two ways in which one can think about the logistic regression model (for example: Long 1997; Maddala 1986). The first way starts with the observation that a probability is a number between zero and one, and if we just used a linear regression on a probability we could easily end up with predictions outside that range. With logistic regression we apply the logit transformation to the probabilities, meaning that we have a linear model for the log-odds of success instead of the probability of success. This is shown in equation (1) where y is the binary dependent variable, $P(y=1)$ is the probability that the dependent variable takes the value 1, the xs are the explanatory variables and the βs their effects.

$$\ln\left(\frac{P(y = 1)}{1 - P(y = 1)}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \tag{1}$$

The odds can be interpreted as the expected number of "successes" per "failure", and the log-odds is the logarithm of that. The log-odds has no upper or lower bound, so a linear model cannot give invalid predictions. Moreover, the exponentiated coefficients can be interpreted directly as odds ratios, that is, the ratio by which the odds change per unit increase in the explanatory variable.

This can also be seen as a way of fitting an "S-shaped" curve denoting the effect of the explanatory variables on the probability of success, as equation (1) can be rewritten as equation (2). This way the predicted probabilities from a logistic regression model are guaranteed to remain within the allowable range for probabilities.

---

[2] I will use degree of plausibility or chance as general terms for the assessment of how likely it is that an event occurs, and consider probabilities, odds, and log odds as specific ways of quantifying the general concept of degree of plausibility or chance.

$$P(y = 1) = \frac{\exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}{1 + \exp(\beta_0 + \beta_1 x_1 + \beta_2 x_2)} \tag{2}$$

The second way of thinking about logistic regression is to assume that there is an unobserved latent propensity to experience an event. If that latent propensity passes a threshold (typically 0) then the event will occur. What is observed is whether or not the event occurs and not the propensity. The latent variable is influenced by the explanatory variables and an error term, as in equation (3), where y^* is the latent propensity and ε is the error term.

$$y^* = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{3}$$

The probability of experiencing the event is in this model the probability that the latent propensity is larger than zero, which according to equation (3) can be written as

$$P(y = 1) = P(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon > 0) \tag{4}$$

Since both the βs and the xs are fixed, the only uncertain part is the error term. As a consequence equation (4) can be rewritten as equation (5):

$$P(y = 1) = P(\varepsilon > -[\beta_0 + \beta_1 x_1 + \beta_2 x_2]) \tag{5}$$

The probability that a random variable, in this case ε, is larger than some number can be derived from the cumulative distribution function. If we assume that the error term follows a logistic distribution with mean 0 and a variance equal to $\pi^2/3$, then equation (5) will become equation (2), thus leading to logistic regression.

The fact that both derivations lead to the exact same model, has led to an attitude that it is pretty much irrelevant which derivations one chooses. The point of this article is that the difference in interpretation has a big impact.

The problem

One characteristic of logistic regression that has troubled many authors (for example: Allison 1999; Auspurg and Hinz 2011; Cramer 2007; Gail, Wieand, and Piantadosi 1984; Hauck, Neuhaus, Kalbfleisch, and Anderson 1991; Lee 1982; Mood 2010) is that if we add a variable to a logistic regression model, the effects of the other variables is very likely to change, *even if* this additional variable is uncorrelated with the other explanatory variables. All that is required is that this new variable influences the outcome, it does not need to be correlated with the other explanatory variables. This phenomenon is also referred to as non-collapsability. (Asmussen and Edwards 1983; Greenland and Pearl 2011; Greenland, Robins, and Pearl 1999; Whittemore 1978) This is problematic because control variables are added to a model to adjust for possible confounders. If a variable is not correlated with another explanatory variable, then it cannot be a confounder, and we would like the effects of our explanatory variables of interest to not change when adding these "non-confounders". The property of logistic regression that the effects do change when adding a "non-confounder" is typically explained with the presence of unobserved heterogeneity.

The simplest version of this explanation starts with the latent variable representation of the logistic regression model. As discussed above, the dependent variable in that representation of logistic regression is a latent propensity to experience the event. The fact that the dependent variable is latent means that the unit is unknown; it does not have a known scale like meters, euros or minutes. Instead the unit of that latent variable is fixed by fixing the variance of the error term. So if something changes the residual variance, then that also changes the scale of the dependent variable. One way to change the residual variance is to add a new explanatory variable to the model. That new variable was previously part of the residual. So adding that variable to the regression reduces the residual variance, under the usual assumption that the explanatory variable is uncorrelated with the error term. The scale of the dependent variable thus changes by adding a variable to our model.

This is pretty damning; if the scale of the dependent variable changes then a comparison of coefficients becomes meaningless. If the scale of the dependent variable is so dependent on which control variables you add, even if those control variables are uncorrelated with the explanatory variable of interest, then how can the coefficient of a logistic regression represent the effect of a variable? This is the basis for Mood's first claim that it is problematic to interpret log-odds ratios or odds ratios as substantive effects and second claim that it is problematic to compare across models with different independent variables.

This problem also affects comparisons of logistic regression coefficients across groups. To borrow an example from Allison (1999): Say, we are estimating one logistic regression to explain some labor market outcome for men and another for women. He makes a plausible argument that women's career paths and labor market experiences tend to be more variable than those of men, leading to a residual variance that is larger for women than for men. As a consequence the scale of the dependent variable in the sample of men is different from the scale of the dependent variable in the sample of women. This is the basis for Mood's third claim that it is problematic to compare log-odds ratios or odds ratios across groups.

The biggest problem is that the scale of latent dependent variable changes when the residual variance changes. There is an easy solution to that, as there is another way of looking at logistic regression: logistic regression models linear effects of explanatory variables on the log-odds of experiencing the event. In that interpretation the dependent variable is no longer a latent variable but the log-odds. That scale is known: the logarithm of the odds, which is the logarithm of the expected number of "successes" per "failure". This scale does not change when we add or remove variables from our model and it is the same across groups.

However, this does not solve everything: the effects, the odds ratios, will still change if you add an unconfounding variable. Adapting a numerical example by Buis (2011): Say we have two countries, A and B, and within each country we look at the effect of a binary variable *x* on a binary variable *y*, as

described in Table 1. These cross tabulations make for convenient examples, as the parameters of logistic regression model can be easily computed by hand: The odds of getting a high y is larger in country A compared to country B, but in both countries the odds of getting a high y is three times larger for high x individuals compared to low x individuals. This factor three is the odds ratio one would get in a logistic regression model. Another important feature of this example is that the distribution of x is the same in each country, so country is not a confounding variable.

--- Insert Table 1 here ----

If one were interested in the effect of x and had used a linear regression then it would not matter whether one controlled for country or not, as country and x are in this example completely uncorrelated. However, this is not true with logistic regression. Not controlling for country means collapsing Table 1 to Table 2: If we do not want to control for country then it is sufficient to know that there are 300 + 200 =500 observation with low x and low y, 200+100=300 observations with low x and high y, etc. When we do that the odds ratio drops from 3 to 2.778. The reason for that change in effect of x is that collapsing a table means averaging the probabilities. For example the probability of being a high y given a low x in countries A and B is respectively 200/400=.5 and 100/100=.25. After collapsing Table 1 to Table 2 the probability of being a high y given a low x is 300/800 = (.5 + .25)/2 = .375. However, with logistic regression we work with odds and odds ratios, and these are non-linear transformations of probabilities. So averaging probabilities does not correspond to averaging odds. To continue the example: (1+.333)/2=0.667≠0.6.

--- Insert Table 2 here ---

So even though the scale of the dependent variable is comparable across models and groups, the size of the effect is still dependent on which variables were added to the model even if those added variables are non-confounding variables. In particular, in the model without the extra variable the probabilities tend to be less extreme (more in the middle and further away from either one or zero),

7

and the effects are thus smaller compared to the model with the extra variable (Neuhaus and Jewell 1993).

A solution

Much of the effort in this area has focused thus far on removing the effect of non-confounding variables, by for example using average marginal effects or linear probability models (Auspurg and Hinz 2011; Mood 2010), using standardized regression coefficients (Karlson 2015; Winship and Mare 1984), or directly estimating the degree of heterogeneity and controlling for it (Allison 1999; Williams 2009). However, what does this effect of unconfounding variables mean and is it always desirable to remove it?

Consider the meaning of the dependent variable in a logistic regression: In a pure mechanical sense logistic regression models the proportion of successes. But often this proportion is given the additional meaning of chance, that is, the researcher's assessment of how likely it is that an individual will experience an event. In general a chance involves a person who has a statement about which he or she has a given set of evidence that entitles him or her to be absolutely sure that the statement is true, or absolutely sure that the statement is false, or somewhere in between. A probability, or an odds, or a log odds measures her or his "degree of rational believe" (Keynes 1921 [2004]), or degree of "partial assent" (Cox 1946), or "degrees of plausibility" (Jaynes 2003). So, a chance tries to quantify uncertainty.

It is useful to think about where that uncertainty comes from by comparing two models. The first model can be illustrated using the example of the chance that a child enters university. This chance is influenced by many things and some of these are known to the researcher (e.g. family background) and others unknown (e.g. whether a teacher's teaching style fits well to the learning style of the student in question). The uncertainties we want to capture with chance in this case are due to all the unknown influences. As a consequence, learning more about the student should change our assessment of the chance that that student will enter university. In this case the chance is conditional on what we know about that student, and not an absolute property of the student (Keynes 1921

[2004]). The second model can be illustrated using the example of a (thought) experiment involving flipping a coin or rolling a dice. In this case the (thought) experiment is constructed in such a way that there are no unknown influences that we in theory could know in advance (at least, not without cheating). Instead we are uncertain because this is a process with an *a priori* unknowable outcome. In such an experiment the chance could be seen as an absolute property of the coin or die because the experiment was created in such a way that there is no information to add. However, phenomena where there is no possible additional information are exceedingly rare in the social sciences, so the first model describes most chances as estimated by logistic regression in the social sciences better than the second model.

Another question that needs to be cleared is that if we use logistic regression as a model for a chance, then what is the statement we are uncertain about? We are not uncertain about the persons in our dataset, as we have already observed whether or not they experienced the event we are interested in. Instead, we imagine a person with the same characteristics as a person in the dataset but for whom it is unknown whether or not he or she experienced the event[3]. We use the predicted proportion of persons with those characteristics that experienced the event as our measurement of the degree of plausibility that this new person experiences the event. So the chances are not a property of the units that are being studied, but a property of the researcher: it is her or his assessment of how likely it is that some with similar characteristics will experience the event.

If we look at chance in this way, then it is necessarily conditional on the available information. In logistic regression the "available knowledge" translates to the set explanatory variables included in the model. So the degree of plausibility is defined by the choice of control variables rather than something that exists outside the model. This links back to the scale of the latent variable being dependent on which control variables are added to the model. However, instead of this dependence being a problem, it is

---

[3] If we have continuous characteristic in our model finding such a similar person may involve interpolation or extrapolation, but the basic argument remains the same.

now a desirable property of the logistic model in the sense that the chances we get out of a logistic regression react to new information (new variables) as one would expect. In particular, as one adds new variables, and thus becomes surer about the outcome, the chances will become more extreme. This property was illustrated in the example from tables 1 and 2.

This way of thinking about probability not only involves a change in chances due to adding extra information, but also a change in the effects of variables on chances due to adding extra information. With effect I mean the extent to which two chances are not equal. This could be measured in many ways. The most common ones are the difference between two probabilities or the ratio of two odds. If we choose to measure the effect as the ratio of odds (the aptly named odds ratio), then the effect will have desirable properties in terms of the chance interpretation of logistic regression. Before adding extra information, we will be less sure, and thus the odds should take less extreme values than after adding extra information. This means that in the chance interpretation of logistic regression we want the odds to be closer together before adding variables than after adding variables. If the chances are forced to be closer together due to less information being available, then the effects have to be smaller; there is less room for variables to have an effect before adding extra information compared to after adding extra information. So not only should one expect the chances to change in reaction to extra information, but also effects on those chances. This means that adding unconfounding variables should increase the effect of other variables, which is exactly what happens in logistic regression. Moreover, if the rolling dice or coin flipping model of chance applies to the chance we are studying, then logistic regression model will still act as expected: any additional variable will in that case not influence the outcome, and a variable that adds nothing will not change the logistic regression coefficients as the resulting table is collapsible along that dimension.

A comparison of effects across groups is also influenced by this interpretation of the dependent variable. If one group is more predictable than another, then that should lead to larger effects in the more predictable group. The logic is the same as with adding an unconfounding variable: In a group where one is less sure the range of possible predicted odds should be more restrictive compared to a

group where one is surer. As a consequence there is more room for a variable to have an effect in a group where one is surer. Consider again the example by Allison (1999) where we investigate the difference between men and women in effect of a variable on the chance of some labor market outcome. The labor market experiences tend to more predictable for men than for women. This should make the effects of variables for men larger than those effects for women. As was discussed above the logistic regression model has exactly this characteristic. So rather than that being a problem, this is again a desirable characteristic and a comparison of odds ratios across groups will thus give an accurate description of differences in effects across these groups.

Mood (2010) also considered the comparison of coefficients across models with different explanatory variables problematic. Such a comparison is typically done to quantify the indirect effect of a variable. For example, we may want to know how much of the effect of parental background on the probability of entering university can be explained by the fact that children with a privileged parental background tend to perform better at school prior to entering university, and those children that performed better at school are more likely to enter university. In a linear regression model we could just estimate two models: one with just parental background and one with parental background and prior academic performance. The difference in effect of parental background between these models quantifies the indirect effect of parental background on entering university via prior academic performance. However, this trick does not work with logistic regression: If adding a variable should change the effect of parental background regardless of whether it is correlated with family background, then the comparison of coefficients between models with and without the intervening variables is not a good way of quantifying the indirect effect. Instead one can scale the effects in both models such that they refer to the uncertainty present without adding the intervening variables (Buis 2010; Erikson, Goldthorpe, Jackson, Yaish, and Cox 2005) or scale the effects in both models such that they refer to the uncertainty present when also adding the intervening variables (Karlson, Holm, and Breen 2012). Either can be reasonable. The logic behind the former is that the explanandum is the total effect, which is the effect of the variable of interest without adding the intervening variables. The logic behind the

latter is that the intervening variable is part of the model, so the uncertainty that remains after adding the intervening variable is the relevant amount of uncertainty. A consequence of that is that the total effect using the method by Karlson, Holm, and Breen (2012) will differ depending on which intervening variables one wants to model. So if one wants to compare indirect effects for different intervening variables then that would be easier using the method by Erikson et al. (2005) or Buis (2010).

So, even though the coefficients are exactly the same regardless of whether we consider the logistic regression as a model of a latent variable or as a model of a chance, the interpretation of the coefficients are very different. So much so, that changing the interpretation of the dependent variable to a degree of plausibility solves some of the problems with logistic regression. If the solution is that easy, then why wouldn't we always choose the chance interpretation of logistic regression? To be more precise, when would a chance interpretation of logistic regression be appropriate? Let's start with the alternative: the latent variable interpretation. Its main strength is that it can easily incorporate a theory of individual behavior involving rational choice and utility maximization. If one is interested in testing such a theory, then the latent variable interpretation is the most sensible choice. In a latent variable interpretation one thus tries to reconstruct the process through which the individuals in the data came to make the decision they made. There are many research question, where this is exactly what one wants to do.

If one interprets the dependent variable as a chance, then one no longer talks about the persons in the data, as we already observed whether or not they experienced the event of interest, and there is thus no uncertainty. Instead with the chance interpretation of logistic regression we talk about similar persons for whom we don't (yet) know whether they will experience the event of interest or not. A logical application of this way of thinking would be the evaluation of some treatment or intervention, as there the interest is often in whether or not this treatment or intervention should be applied more often. In those cases we are interested in a reasonable assessment of how likely it is that a positive outcome occurs for those new cases, and in particular to what extent that chance improves (or deteriorates) when a person receives a given treatment.

Within sociology another set of questions that would fit well in this framework would be questions involving "inequality of opportunity". Here we look at a society and consider the extent to which they offer all its members the same chances of attaining favorable positions regardless of their background. These favorable positions could for example be a "high" occupation, more education, a desirable partner, or health. Inequality of opportunity is about persons who have not attained that favorable position yet, but have a chance to do so. In particular, the difference in these chances across groups is what defines inequality of opportunity. Moreover, inequality of opportunity is a property of a society not an individual. The chances can also be viewed as properties of groups rather than individuals; all persons with the same characteristics share the same chance. Seen from this perspective, the logistic regression model interpreted in terms of chances is a "population averaged model" (Agresti 2013; Fitzmaurice, Laird, and Ware 2004).

Summary and conclusions

There has been a long discussion on whether the influence of unobserved heterogeneity makes the interpretation of logistic regression coefficients problematic. Mood (2010) summarized that discussion by naming three problems:

1. "It is problematic to interpret log-odds ratios or odds ratios as substantive effects, because they also reflect unobserved heterogeneity

2. It is problematic to compare log-odds ratios or odds ratios across models with different independent variables, because the unobserved heterogeneity is likely to vary across models.

3. It is problematic to compare log-odds ratios and odds ratios across samples, across groups within samples, or over time – even when we use models with the same independent variables – because the unobserved heterogeneity can vary across the compared samples, groups, or points in time."

This article adds to this debate by stating that unobserved heterogeneity does have the effects as summarized by for example Mood (2010) and Auspurg and Hinz (2011), but that that does not have to be a problem and can actually be desirable. The key difference with previous contributions is that this article focusses on finding research problems where these characteristics of the odds ratio are desirable rather than problematic. These non-problematic research problems have in common that their dependent variable is the chance or the "degree of plausibility" – that is, an assessment of how likely it is that an event occurs – instead of a latent propensity. A degree of plausibility should be dependent on the information available for making such an assessment; the more (mutually consistent) information we have the surer we are. In logistic regression the set of explanatory variables represents the information used for making such an assessment. So adding variables to a model that predicts a degree of plausibility should result in different coefficients.

To be more precise, consider what one would expect if one adds an additional variable z, and that variable is relevant for predicting a probability. In that case one will be more certain after adding it. In other words, the chances should become more extreme. This should also influence the effect of other variables. Say I am interested in the effect of a variable x. The effect is the extent to which the chances are not the same across groups, which can be measured in different ways. For example, as the difference in probabilities or as the ratio of odds. Before adding the additional variable z the predicted chances should be closer together compared to after adding that additional variable. As a consequence, there is less room for the variable x to have an effect before adding z than after adding z. So adding the additional variable z should increase the effect of a variable, even if x and z are uncorrelated. The log-odds ratios and odds ratios from a logistic regression show exactly this behavior. So Mood (2010) was right when she noted in the first problem that these coefficients are dependent on which variables are included in the model even if those additional variables are uncorrelated with the variables of interest, but that property of logistic regression can actually be desirable instead of a problem.

The size of the effect on a degree of plausibility is a function of how certain we are that the event of interest happens. This degree of certainty can change by adding a variable as discussed above, but it can also differ from group to group. Here we would expect stronger effects in groups where we are more certain. Within groups were we are more certain the predicted chances can be more extreme, so there is more room for a variable to have an effect. Logistic regression coefficients have exactly this property, as can be most clearly seen in the latent variable representation of logistic regression. So rather than making odds ratios incomparable across groups, this property of logistic regression ensures that a comparison of odds ratios give an accurate description of the difference in effects across groups.

The comparison coefficients across models is more complicated. The purpose of such a comparison is often to see how much of an effect can be explained by a set of intervening variables. One first estimates a model with the explanatory variable of interest but without the intervening variables and then a model with both the explanatory variable of interest and the intervening variables. The difference in effect of the variable of interest between these models is interpreted as the part of the effect that is explained by the intervening variables. However, this does not work with logistic regression, since adding variables will change the effect even if the added variables are uncorrelated with the explanatory variable of interest. So Mood's second problem is a real problem, even in the chance interpretation of a logistic regression model.

Looking at logistic regression as explaining a degree of plausibility can thus solve some of the problems, but the effect on a degree of plausibility is not the right effect for all research questions. If we want to understand past individual actions then an effect in terms of the latent variable, with its link to rational choice theory, can be more promising. Effects on a degree of plausibility can be useful when the results are to be used for predicting future uncertain events. This could be the case when we want to evaluate whether a specific treatment works. In that case we want to use the results to help choose the treatment for future cases. Another area where effects on degrees of plausibility can be helpful is the study of macro processes like inequality of opportunity. In that case we imagine looking at persons who differ with respect to their background and look at how likely it is that they obtain some favorable

position. The extent to which those chances differ measure how closed that society is. Again we predict future events, but now we use that to say something about a macro phenomenon; the openness of a society.

To summarize, the three problems with logistic regression stated by Mood (2010) do not have to be as bad if we interpret the dependent variable, the chance, as an assessment of how likely it is that the event of interest occurs. In that case one can conclude that:

1. The odds ratio is a meaningful effect-size. The fact that it is dependent on which variables are included in the model is not a problem but actually a requirement for an effect on a chance.

2. It is indeed problematic to compare coefficients across models with different sets of explanatory variables, since effects on chances are supposed to change when variables are added to the model even if they are uncorrelated with the other explanatory variables.

3. A comparison of odds ratios across groups provides an accurate description of the difference in effects across these groups.

References

Agresti, Alan. 2013. *Categorical data analysis*. Hoboken, NJ: Wiley.

Allison, Paul D. 1999. "Comparing logit and probit coefficients across groups." *Sociological Methods & Research* 28:186-208.

Asmussen, Søren and David Edwards. 1983. "Collapsibility and response variables in contingency tables." *Biometrika* 70:567-578.

Auspurg, Katrin and Thomas Hinz. 2011. "Gruppenvergleiche bei Regressionen mit binären abhängigen Variablen—Probleme und Fehleinschätzungen am Beispiel von Bildungschancen im Kohortenverlauf." *Zeitschrift für Soziologie* 40:62-73.

Buis, Maarten L. 2010. "Direct and indirect effects in a logit model." *Stata Journal* 10:11-29.

—. 2011. "The consequences of unobserved heterogeneity in a sequential logit model." *Research in Social Stratification and Mobility* 29:247-262.

Cox, Richard T. 1946. "Probabiltiy, frequency, and reasonable expectation." *American Journal of Physics* 14:1-13.

Cramer, Jan S. 2007. "Robustness of Logit Analysis: Unobserved Heterogeneity and Mis-specified Disturbances." *Oxford Bulletin of Economics and Statistics* 69:545-555.

Erikson, Robert, John H Goldthorpe, Michelle Jackson, Meir Yaish, and David R Cox. 2005. "On class differentials in educational attainment." *Proceedings of the National Academy of Sciences of the United States of America* 102:9730-9733.

Fitzmaurice, Garret M., Nan M. Laird, and James H. Ware. 2004. *Applied longitudinal analysis*. Hoboken, NJ: Wiley.

Gail, M. H., S. Wieand, and S. Piantadosi. 1984. "Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates." *Biometrika* 71:431-444.

Greenland, Sander and Judea Pearl. 2011. "Adjustments and their consequences—collapsibility analysis using graphical models." *International Statistical Review* 79:401-426.

Greenland, Sander, James M. Robins, and Judea Pearl. 1999. "Confounding and Collapsibility in Causal Inference." *Statistical Science* 14:29-46.

Hauck, Walter W., John M. Neuhaus, John D. Kalbfleisch, and Sharon Anderson. 1991. "A consequence of omitted covariates when estimating odds ratios." *Journal of Clinical Epidemiology* 44:77-81.

Jaynes, Edwin T. 2003. *Probability theory: the logic of science*: Cambridge university press.

Karlson, Kristian Bernt. 2015. "Another look at the method of y-standardization in logit and probit models." *The Journal of Mathematical Sociology* 39:29-38.

Karlson, Kristian Bernt, Anders Holm, and Richard Breen. 2012. "Comparing regression coefficients between same-sample nested models using logit and probit: A new method." *Sociological Methodology* 42:286-313.

Keynes, John Maynard. 1921 [2004]. *A Treatise on Probability*. Mineola, NY: Dover Publications, Inc.

Lee, Lung-Fei. 1982. "Specification error in multinomial logit models: Analysis of the omitted variable bias." *Journal of Econometrics* 20:197-209.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks: Sage.

Maddala, Gangadharrao S. 1986. *Limited-dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge university press.

Mood, Carina. 2010. "Logistic regression: Why we cannot do what we think we can do, and what we can do about it." *European Sociological Review* 26:67-82.

Neuhaus, John M. and Nicholas P. Jewell. 1993. "A geometric approach to assess bias due to omitted covariates in generalized linear models." *Biometrika* 80:807-815.

Norton, Edward C. 2012. "Log odds and ends." in *NBER Working Papers, 18252*. Cambridge, MA: National Bureau of Economic Research.

Pang, Menglan, Jay S Kaufman, and Robert W Platt. 2013. "Studying noncollapsibility of the odds ratio with marginal structural and logistic regression models." *Statistical Methods in Medical Research*.

Whittemore, Alice S. 1978. "Collapsibility of multidimensional contingency tables." *Journal of the Royal Statistical Society. Series B (Methodological)* 40:328-340.

Williams, Richard. 2009. "Using heterogeneous choice models to compare logit and probit coefficients across groups." *Sociological Methods & Research* 37:531-559.

Winship, Christopher and Robert D. Mare. 1984. "Regression models with ordinal variables." *American Sociological Review* 49:512-525.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT press.

Table 1: Hypothetical example while controlling for country

		y		N	probability	odds	Odds ratio
country	*x*	Low	High				
A	Low	200	200	400	.5	1	3
	High	100	300	400	.75	3	
B	Low	300	100	400	.25	.333	3
	High	200	200	400	.5	1	

Table 1: Hypothetical example while controlling for country

Table 2: Hypothetical example without controlling for country

	Y		N	probability	odds	Odds ratio
x	Low	High				
Low	500	300	800	.375	0.6	2.778
High	300	500	800	.625	1.667	