sociological science

Marginal Odds Ratios: What They Are, How to Compute Them, and Why Sociologists Might Want to Use Them

Kristian Bernt Karlson,[a] Ben Jann[b]

a) University of Copenhagen; b) University of Bern

Abstract: As sociologists are increasingly turning away from using odds ratios, reporting average marginal effects is becoming more popular. We aim to restore the use of odds ratios in sociological research by introducing marginal odds ratios. Unlike conventional odds ratios, marginal odds ratios are not affected by omitted covariates in arbitrary ways. Marginal odds ratios thus behave like average marginal effects but retain the relative effect interpretation of the odds ratio. We argue that marginal odds ratios are well suited for much sociological inquiry and should be reported as a complement to the reporting of average marginal effects. We define marginal odds ratios in terms of potential outcomes, show their close relationship to average marginal effects, and discuss their potential advantages over conventional odds ratios. We also briefly discuss how to estimate marginal odds ratios and present examples comparing marginal odds ratios with conventional odds ratios and average marginal effects.

Keywords: odds ratio; logit; regression; marginal effects; confounding; mediation

Citation: Karlson, Kristian Bernt, and Ben Jann. 2023. "Marginal Odds Ratios: What They Are, How to Compute Them, and Why Sociologists Might Want to Use Them." Sociological Science 10: 332-347.

Received: January 31, 2023

Accepted: February 17, 2023

Published: April 27, 2023

Editor(s): Arnout van de Rijt, Stephen Vaisey

DOI: 10.15195/v10.a10

Copyright: © 2023 The Author(s). This open-access article has been published under a Creative Commons Attribution License, which allows unrestricted use, distribution and reproduction, in any form, as long as the original author and source have been credited.

source: https://doi.org/10.48350/182033 | downloaded: 28.4.2024

LOGIT models and their ensuing odds ratios form the backbone of much sociological research. Despite their prominence, recent methodological research has brought to sociologists' attention some serious problems in using and interpreting odds ratios (Allison 1999; Mood 2010; Breen, Karlson, and Holm 2018; Bloome and Ang 2022). These problems are rooted in a peculiar property of the logit model: the magnitude of its coefficients changes even if one controls for a third variable that is uncorrelated with the predictor of interest, a property known as noncollapsibility, rescaling, or sensitivity to unobserved heterogeneity. Although sociologists have responded to this challenge in different ways, the reporting of (average) marginal effects implied by a logit model or obtained from a linear probability model is now recommended in the methods literature (Breen et al. 2018; Mize 2019; Mize, Doan, and Long 2019; Long and Mustillo 2021). Marginal effects are not arbitrarily affected by the error term and yield readily interpretable effects on the probability scale, which to many is more intuitive than a ratio between odds (Cramer 2007; Norton and Dowd 2018).

Marginal effects are also beginning to replace odds ratios as a preferred effect metric in substantive research. This change in practice becomes clear when one considers articles published in the *American Sociological Review* between 2010 and 2021. Upon conducting a search on the *Review*'s website, we found that the term "marginal effect" appeared in 11 articles of which the vast majority (nine) were published between 2016 and 2021. Similarly, "linear probability model" appeared in 16 articles of which the vast majority (13) were published between 2016 and 2021.

In contrast, "odds ratio" appeared in 41 articles of which only a minority (nine) were published between 2016 and 2021. Although marginal effects are gaining popularity over odds ratios, they do not necessarily align with much sociological research in which relative inequality is a key concept (e.g., in stratification research, political sociology, medical sociology, or demography). Indeed, many sociologists still prefer the odds ratio precisely because it is a relative measure and because it is insensitive to the marginal distribution of the dependent variable (Mare 1981; Erikson and Goldthorpe 1992). In contrast, the magnitude of a marginal effect depends on the distribution of the binary outcome (Mare 1981:76; Holm, Ejrnæs, and Karlson 2015), a property that makes it difficult to directly compare effect sizes among, say, populations with different overall outcome rates.

In this article, we aim to restore the odds ratio as a relevant effect metric in sociological research by introducing what we term a "marginal odds ratio." This effect metric has properties similar to the properties of marginal effects, including being unaffected by noncollapsibility, but it retains the relative effect interpretation. It thus presents itself as a viable alternative or complement to the reporting of marginal effects.[1] Drawing on work in statistics and epidemiology on this topic (e.g., Zhang 2008; Pang, Kaufman, and Platt 2016; Daniel, Zhang, and Farewell 2021), we first define the marginal odds ratios in the potential outcomes framework.[2] This framework makes clear the marginal odds ratio estimand and shows its close relationship to the average marginal effect estimand. We then explain how the marginal odds ratio should be interpreted as a population-averaged effect and how this interpretation differs from the conventional odds ratio typically obtained from a logit model that has a conditional interpretation. We then go on to briefly outline how to estimate the marginal odds ratio using counterfactual predictions from a logit model. We also present two examples demonstrating the versatility of the marginal odds ratio. In a companion technical paper, we give a thorough technical introduction to estimation approaches and introduce software that makes the estimation of marginal odds ratios straightforward (Jann and Karlson 2023).

Marginal Odds Ratios

We define the marginal odds ratio using potential outcomes notation (Rubin 1974). This notation makes clear the estimand and its close relationship to the average marginal effect. In our exposition, we exclusively focus on binary treatments and refer readers interested in the extension to continuous treatments to our technical paper (Jann and Karlson 2023). Let Y_t be the potential outcome of an individual receiving either the treatment ($T = 1$) or the control ($T = 0$). Comparing Y_t for the treated (Y_1) or untreated (Y_0) is informative about the effect of T on Y. Scholars are often interested in the average treatment effect, which is defined as $E[Y_1] - E[Y_0]$, that is, the difference in the expectation over each potential outcome. For binary outcomes, the expectation equals the probability of success, meaning that the average treatment effect equals an average marginal effect defined as

$$\text{AME} = \Pr[Y_1 = 1] - \Pr[Y_0 = 1]. \quad (1)$$

The AME is the success probability difference if everyone was treated relative to if everyone was untreated. In a similar vein, we define the *marginal odds ratio* as

$$\text{MOR} = \frac{\text{odds}(\Pr[Y_1 = 1])}{\text{odds}(\Pr[Y_0 = 1])}, \tag{2}$$

where $\text{odds}(p)$ stands for $p/(1-p)$. This odds ratio is the ratio of the odds of success if everyone was treated relative to the odds of success if everyone was untreated.[3] The estimands in Equations (1) and (2) involve the same counterfactual quantities, but the AME is a probability difference, whereas the MOR is a ratio between odds.

The estimands in Equations (1) and (2) can also be expressed as depending on other variables, which we denote $\mathbf{X}$. For example, applied researchers will often be interested in adjusting for a set of additional covariates if they want to control for potential confounding or are interested in effects for different subpopulations. If we assume that $\mathbf{X}$ has a given distribution in the population, then we can express the conditional success probability given $\mathbf{X} = \mathbf{x}$ as $\Pr(Y_t = 1 \mid \mathbf{X} = \mathbf{x}) = E[Y_t \mid \mathbf{X} = \mathbf{x}]$. By the law of iterated expectations, we can write the unconditional success probability as

$$\Pr(Y_t = 1) = E_{\mathbf{X}}[\Pr(Y_t = 1 \mid \mathbf{X} = \mathbf{x})], \tag{3}$$

where $E_{\mathbf{X}}$ is the expectation over the distribution of $\mathbf{X}$. Thus, we can rewrite Equations (1) and (2) as

$$\text{AME} = E_{\mathbf{X}}[\Pr(Y_1 = 1 \mid \mathbf{X} = \mathbf{x})] - E_{\mathbf{X}}[\Pr(Y_0 = 1 \mid \mathbf{X} = \mathbf{x})], \tag{4}$$

$$\text{MOR} = \frac{\text{odds}\{E_{\mathbf{X}}[\Pr(Y_1 = 1 \mid \mathbf{X} = \mathbf{x})]\}}{\text{odds}\{E_{\mathbf{X}}[\Pr(Y_0 = 1 \mid \mathbf{X} = \mathbf{x})]\}}. \tag{5}$$

We term the expression in Equation (5) the "adjusted marginal odds ratio," although it is the same estimand as the marginal odds ratio given in Equation (2). The expression in Equation (5) is useful when estimating marginal odds ratios in substantive research using observational data where confounding is ubiquitous. Given that we refer to Equation (5) as an adjusted MOR, we refer to Equation (2) as a gross or unadjusted MOR. We may think about them as effects controlling and not controlling for additional and potentially confounding covariates (Karlson, Popham, and Holm 2021). We later give an example showing the difference between the two.

Relationship to the Logit Model

Although we have defined the marginal odds ratio estimand in terms of potential outcomes, sociologists usually obtain odds ratios from a logistic response model. To show the relationship between the two, we first write the unconditional logistic model as

$$\Pr(Y_t = 1) = \text{logistic}(\alpha + \delta t)\,, \tag{6}$$

where $\text{logistic}(z)$ stands for $\exp(z)/[1-\exp(z)]$. In this model, the exponent to the treatment logit coefficient, $\exp(\delta)$, has a marginal odds ratio interpretation.

However, once we condition on other covariates, **X**, the interpretation of the odds ratio changes to a conditional one. To see this, assume that we include **X** in the regression equation,

$$\Pr(Y_t = 1 \mid \mathbf{X} = \mathbf{x}) = \text{logistic}\left(\widetilde{\alpha} + \widetilde{\delta}t + \mathbf{x}\beta\right). \tag{7}$$

In this model,

$$\text{COR}_{\mathbf{X}} = \frac{\text{odds}[\Pr(Y_1 = 1 \mid \mathbf{X} = \mathbf{x})]}{\text{odds}[\Pr(Y_0 = 1 \mid \mathbf{X} = \mathbf{x})]} = \exp\left(\widetilde{\delta}\right) \tag{8}$$

is a *conditional* odds ratio (COR), where conditional refers to the effects operating at the subgroup level defined by the covariates in **X**. The conditional odds ratio differs from the marginal counterpart adjusting for **X**, given by

$$\begin{aligned} \text{MOR}_{\mathbf{X}} &= \frac{\text{odds}\{E_{\mathbf{X}}[\Pr(Y_1 = 1 \mid \mathbf{X} = \mathbf{x})]\}}{\text{odds}\{E_{\mathbf{X}}[\Pr(Y_0 = 1 \mid \mathbf{X} = \mathbf{x})]\}} \\ &= \frac{\text{odds}\left\{E_{\mathbf{X}}\left[\text{logistic}\left(\widetilde{\alpha} + \widetilde{\delta} + \mathbf{x}\beta\right)\right]\right\}}{\text{odds}\{E_{\mathbf{X}}[\text{logistic}(\widetilde{\alpha} + \mathbf{x}\beta)]\}}, \end{aligned} \tag{9}$$

which has a population-averaged interpretation, that is, the average population response to changing treatment status (Zeger, Liang, and Albert 1988:1050). The COR in Equation (8) is the response of the subgroup defined by the covariates in **X** (and the COR is assumed to be constant across those groups as the covariates enter additively on the logit scale). In other words, the COR in Equation (8) and the MOR in Equation (9) refer to different estimands, have different interpretations, and cannot be directly compared (Pang et al. 2016; Breen et al. 2018; Daniel et al. 2021; Schuster et al. 2021).[4]

Marginal and conditional odds ratios are equally valid estimands, and their respective uses should depend on the research question. However, from a mathematical perspective, the difference between them arises from what statisticians call noncollapsibility: "Noncollapsibility of the OR [odds ratio] derives from the fact that when the expected probability of outcome is modeled as a nonlinear function of the exposure, the marginal effect cannot be expressed as a weighted average of the conditional effects" (Pang et al. 2016:1926). Indeed, the key difference between the two estimands is whether the averaging occurs on the log odds scale or on the probability scale.

The mathematical relationship between the two estimands (noncollapsibility) is well-described in the methods literature. From this literature, we highlight two key points. First, from Equations (8) and (9), we see that the MOR will differ from the COR whenever $\beta \neq 0$ (i.e., if there are other relevant predictors apart from the treatment variable). Moreover, the MOR will be attenuated relative to the COR. For example, if there is just a single covariate X, the relationship between the two can be approximated by

$$\ln \text{MOR}_X = \frac{\ln \text{COR}_X}{\sqrt{1 + 0.35\beta^2 \text{var}(X)}}. \tag{10}$$

(Zeger et al. 1988:1054).[5] Whenever $\beta = 0$, the COR collapses to the MOR. Whenever $\beta \neq 0$, that is, if the adjusting covariate has a nonzero effect on the outcome, the COR will be larger than the MOR *even if* X is not confounding the treatment effect.[6] Moreover, the attenuation of the MOR relative to the COR depends on the magnitude of β and the dispersion in X. In the example we later provide, we demonstrate how the attenuating effect of noncollapsibility operates.

Second, although there is only one MOR, there are in principle an infinite number of CORs. The interpretation of the conditional odds ratio will depend on the covariates included in the regression equation as it refers to effects specific to subgroups defined by those covariates. Each of these CORs is not directly comparable with the other CORs. In practical terms, whenever researchers successively add variables to a logit regression equation—a widespread practice in sociological research—the COR estimand changes and so coefficients of the treatment of interest are not directly comparable. Karlson, Holm, and Breen (2012) suggested solving this issue by holding one underlying COR estimand constant (what they refer to as the full model including all covariates) and then changing the set of conditioning control variables using residualized predictors (what they refer to as a reduced model). For this reason, the method by Karlson et al. (2012) recovers a COR estimand (Karlson et al. 2021).

Why Should Sociologists Use Marginal Odds Ratios?

Because MOR and COR both are valid estimands, there is no a priori argument for choosing one over the other. However, although we agree with the general point that the choice of estimand should depend on the research question, for most practical purposes we find that the MOR estimand is superior to COR estimands. We highlight four reasons. First, the MOR has an interpretation equivalent to an average marginal effect on the probability scale: it is a population-averaged effect focusing on the average "population response" to a treatment of interest. Given the increasing reporting of average marginal effects in sociological research, the MOR presents itself as a notable alternative or complement to the reporting of AMEs. Second, because MORs are unaffected by noncollapsibility, they can be used for comparing coefficients from same-sample models including different covariates (i.e., for mediation analyses or effect decompositions). Third, MORs are straightforward to compare across different studies or populations as their magnitude does not depend in arbitrary ways on the conditioning set (i.e., set of control variables). Fourth, because many COR estimands exist (depending on the conditioning set) but only one MOR estimand, researchers are free from presenting arguments for why a specific COR estimand is more interesting than another. Such arguments would require highly developed theoretical frameworks, which are rare in sociology.

Estimating Marginal Odds Ratios

Marginal odds ratios can be estimated in different ways. In a technical companion paper (Jann and Karlson 2023), we review these approaches and show which estimands each estimation technique recovers. Here, we present an estimation approach based on counterfactual predictions (known as G-computation) and focus on binary treatments for making the exposition as accessible as possible (Robins 1986). This approach compares counterfactual predictions from a (typically parametric) model involving the treatment and conditioning covariates (Zhang 2008). The approach proceeds in four steps:

1. Regress Y on T and $\mathbf{X}$ using a logit model (or, in principle, any other model).
2. Generate two sets of predictions of the success probability for each observation in the data, one setting everyone to be treated, $T = 1$, and one setting everyone to be untreated, $T = 0$ (i.e., $\hat{p}_{i,T=1}$ and $\hat{p}_{i,T=0}$, where i indexes observations).
3. Average each set of predictions to obtain the marginal or population-averaged success probabilities if treated ($\overline{p}_{T=1}$) and if untreated ($\overline{p}_{T=0}$), respectively.
4. To obtain the marginal odds ratio, plug in the average marginal predictions from step 3 into the formula for the marginal odds ratio,

$$\widehat{\text{MOR}} = \frac{\text{odds}(\overline{p}_{T=1})}{\text{odds}(\overline{p}_{T=0})}. \tag{11}$$

This approach based on counterfactual predictions is a straightforward way of obtaining marginal odds ratios. The approach is implemented in the user-written Stata command *lnmor*, which we present in our companion paper (Jann and Karlson 2023).[7] It is also worth noting that the average marginal effect can be obtained by the four steps outlined above, except that in the fourth step, one plugs in the average marginal predictions into the estimand formula for AMEs, that is, $\overline{p}_{T=1} - \overline{p}_{T=0}$. From this estimation perspective, the close relationship between AME and MOR also becomes apparent.

Examples

Academic Ability and Intergenerational College Mobility

Stratification scholars are interested in quantifying the extent to which academic abilities explain or mediate family background inequalities in educational attainment (Boudon 1974; Erikson et al. 2005; Jackson 2013). Gaps by family background in educational attainment that operate independently of demonstrated academic abilities are theorized to represent the "secondary effects" of family background, that is, how class-based aspirations, preferences, and outlooks feed into educational decisions over and above those differences that come about through unequal skill levels. In our example, we examine the extent to which academic ability (as measured by a cognitive test) accounts for the gap in college attainment between children born to parents with and without a college degree. To fully illustrate the difference between MOR and COR, we conduct this analysis on representative samples from the United States and Denmark. Comparing the United States and Denmark is substantively interesting because, for the birth cohorts we analyze here (born in the mid-1950s through the mid-1960s), Denmark was a more educationally mobile country than the United States (Karlson and Landersø 2021).

For the United States, we analyze data from the National Longitudinal Survey of Youth, 1979 cohort, which followed a national probability sample of children aged 14 through 21 in 1979 (Bureau of Labor Statistics 2019). For Denmark, we examine data from the national Danish Longitudinal Survey of Youth, which followed a national-probability sample of seventh graders in 1968/1969 (Hansen 1995). Both data sets provide information on parental college attainment, respondent college attainment (as adults), and a standardized cognitive ability test.[8] Our analytical strategy is straightforward: we compare the unadjusted or gross gap by parental college attainment with the one adjusted for academic ability.

Table 1 shows the results. We find that the unadjusted or gross marginal odds ratio is about twice as large in the United States (7.6) as in Denmark (3.8), meaning that Denmark is significantly more educationally mobile (row 1). This estimate has a marginal or population-averaged interpretation as the "population response in child college attainment" to changing from non–college-educated to college-educated parents. Adjusting for academic ability (row 2), we find that the marginal odds ratio reduces to 2.5 in both countries. We interpret this adjusted marginal odds ratio as the impact, *on average in each population*, of parental college attainment on child college attainment, accounting for the unequal distribution of academic abilities across family background. Because the adjusted MOR reduces to the same number (2.5), it means that the "secondary effects" of family background are of similar magnitude in the two countries. Moreover, because the unadjusted MOR is much larger in the United States than in Denmark, it means that academic ability "mediates" a significantly larger portion of the gap by parental college attainment in college completion in the United States than in Denmark.[9]

In contrast to the adjusted marginal odds ratio, the conditional counterpart in row 3 is 3.4 for the United States and 2.9 for Denmark. Thus, had we been using this adjusted COR for comparing the two countries, we would have concluded that, net of academic ability, Denmark is a (albeit only slightly) more educationally mobile country. The adjusted COR has an interpretation that is different from the marginal counterpart: it is the odds ratio for groups with similar levels of demonstrated academic ability, and it does not refer to population-level effects. Moreover, the difference between the adjusted MOR and COR points to the attenuating impact of noncollapsibility. This impact is significantly larger in the United States than in Denmark, meaning that academic ability is a much stronger predictor of college attainment in the United States than in Denmark (net of parental college attainment).[10]

Using the Karlson–Holm–Breen (KHB) approach, we present the unadjusted COR in row 4 (Karlson et al. 2021).[11] The KHB approach holds constant the COR estimand (i.e., the COR within groups of children with different ability levels). We make three observations regarding this estimand. First, had we used this COR for comparing the two countries' gross mobility levels, Denmark would be almost three times as mobile by this measure (compared with twice as mobile with the unadjusted MOR). Had we adjusted for additional covariates (e.g., aspirations) that also have more predictive power in the United States than in Denmark, this

Table 1: Odds ratios and average marginal effects of parental college attainment gap in college attainment unadjusted and adjusted for academic ability: the United States and Denmark

	United States ($N = 10,068$)	Denmark ($N = 2,185$)	United States/ Denmark
1: MOR, unadjusted	7.7 (0.46)	3.8 (0.55)	2.03*
2: MOR, adjusted	2.5 (0.13)	2.5 (0.33)	1.00
3: COR, adjusted	3.4 (0.23)	2.9 (0.45)	1.17
4: COR, unadjusted (khb)	14.2 (1.05)	4.9 (0.78)	2.90*
5: AME, unadjusted	0.43 (0.01)	0.31 (0.03)	1.38*
6: AME, adjusted	0.17 (0.01)	0.20 (0.20)	0.86

Notes: Standard errors in parentheses. MOR is marginal odds ratio; COR is conditional odds ratio; AME is average marginal effect; khb is the Karlson–Holm–Breen decomposition method (using orthogonalized predictors). U.S. data are from the National Longitudinal Survey of Youth, 1979 cohort; Danish data are from the Danish Longitudinal Survey of Youth. $*$ indicates that the country difference in log odds ratios is statistically significant at a five percent level.

ratio would only grow (and the estimand would change). As we stated earlier, CORs are valid estimands, but as there are an infinite amount of them (depending on the conditioning set) and sociological theory rarely is sufficiently detailed to make informed choices about which COR is the better one, it is difficult to argue for choosing one over the other. The MOR does not have this property (there is only one estimand) and so appears to be the best choice for any initial comparison of mobility levels in this example.

Second, as is the case with MORs, the unadjusted COR can be compared with the adjusted COR to gauge mediation. Here we find that the percentage mediated is virtually identical to that based on MORs, indicating that conclusions about mediation are similar using MORs or the KHB approach (Karlson et al. 2021).[12] Third, comparing the unadjusted COR with the unadjusted MOR is informative about the impact of noncollapsibility (as was comparing the adjusted odds ratio counterparts). Here again we find that the bias stemming from noncollapsibility is larger in the United States than in Denmark (resulting from academic ability being a stronger predictor in the United States).

In rows 5 and 6, we also present average marginal effects, that is, the probability difference estimand (cf. Eqs. [1] and [4]). Similar to the odds ratios, we find that the unadjusted AME is larger in the United States (43 percentage points) than in Denmark (31 percentage points). Moreover, adjusting for academic ability significantly reduces the AMEs and, in relative terms, to an extent similar to the reductions seen in log odds ratios.[13] However, the adjusted AME is now *smaller* in the United States than in Denmark (about 14 percent), pointing to the well-known "flipped-signs phenomenon" of interaction terms depending on the scale of measurement (Bloome and Ang 2022). Still, had we relied exclusively on AMEs, we would have concluded that the "secondary effects" of family background are (slightly) larger in Denmark than in the United States, a conclusion that would run counter to the conclusion based on the MOR (similarity) and, in particular, the COR (opposite country difference).

Trends in the College Gap in Attitudes toward Racial Segregation

Political sociologists are interested in how schooling shapes attitudes. We examine the gap in attitudes toward racial segregation between respondents with and without a four-year college degree. In particular, we study whether this gap has changed over two decades, focusing on the results based on average marginal effects (absolute gaps) and marginal odds ratios (relative gaps) when we also control for a range of other covariates. We examine data from the General Social Survey cumulative file (Smith et al. 2019), focusing here on the years 1976 through 1996 when information on attitudes toward racial segregation was collected. Our outcome variable is the response to a question about whether white people have a right to keep black people out of their neighborhoods if they feel like it (and that black people should respect that right). We collapse the outcome variable into a binary variable indicating agreement (1) or disagreement (0) with the stated opinion. We measure college attainment as having completed at least 16 years of schooling. Moreover, we include additional covariates, including survey year (for studying trends from 1976 through 1996), age, gender, race, marital status, and a generic seven-point political views variable indicating whether the respondent thinks of him- or herself as liberal (1) or as conservative (7). The final sample with valid information on all variables comprises 12,239 respondents.[14]

In this example, we are not interested in quantifying the degree of confounding but merely in summarizing the trends in the college gap net of other factors. We specify a logit model in which calendar year is fully interacted with the college dummy and all other covariates (allowing for the effects of the covariates to change over time). We estimate the model specified both as a linear probability model and as a logit model. For all models, we derive average marginal effects and marginal odds ratios evaluated at calendar years 1976, 1981, 1986, 1991, and 1996 and report these implied quantities in Table 2 (estimates of the coefficients in the underlying regression models are available in Table S1 in the online supplement).

The main finding in Table 2 is that the absolute college gap (as measured by average marginal effects) in the attitude toward racial segregation has declined significantly over the 20-year period, whereas the relative gap (as measured by marginal odds ratios) has remained unchanged. In 1976, the absolute college gap net of the other covariates is around 21 percentage points on average (for both AMEs implied by the linear probability model and the logit model, respectively), suggesting that college-educated individuals were much less likely than individuals without a college degree to support racial segregation. In 1996, the absolute gap

Table 2: Average marginal effects and marginal odds ratios of the college gap in the attitude toward racial segregation in 1976, 1981, 1986, 1991, and 1996

	AME_{LPM}	AME_{LOGIT}	MOR	lnMOR
1976	−0.215	−0.213	0.366	−1.006
	(0.019)	(0.021)	(0.043)	(0.118)
1981	−0.180	−0.181	0.374	−0.983
	(0.013)	(0.012)	(0.030)	(0.079)
1986	−0.146	−0.147	0.383	−0.959
	(0.009)	(0.008)	(0.025)	(0.065)
1991	−0.111	−0.115	0.393	−0.934
	(0.011)	(0.009)	(0.035)	(0.089)
1996	−0.076	−0.088	0.403	−0.909
	(0.017)	(0.010)	(0.053)	(0.130)
1976–1996 difference	0.138	0.125	—	0.097
	(0.031)	(0.028)		(0.211)
1976–1996 prop. reduction	64.4%	58.5%	—	9.7%
	(9.6)	(7.7)		(20.0)

Notes: Standard errors in parentheses. MOR is marginal odds ratio; LPM is linear probability model; AME is average marginal effect. Estimates are adjusted for gender, race, age, marital status, and overall political view. Data are from the General Social Survey cumulative file, $N = 12,239$.

is reduced to about seven or eight percentage points on average, pointing to a decline of around 60 percent in just 20 years. This decrease is also highly statistically significant. By way of contrast, the relative gap implied by the marginal odds ratios is virtually constant (we can detect a minor change in the odds ratio toward one, but this trend is not statistically significant).

Thus, although both average marginal effects and marginal odds ratios point to a substantial college divide in attitudes toward racial segregation (with non–college-educated individuals being more supportive of this opinion)—even when we control for potentially "confounding" variables—they disagree on the trend in this gap. To see why this is the case, we report in Figure 1 the average marginal predictions from the logit model by college attainment and survey year. From this figure, we can easily see that the absolute gap reduces over time because there is a general decline in support of racial segregation; this decline is steeper among the non–college educated in absolute terms because they start at a higher level than the college educated. However, the relative difference does not change much, resulting in constant odds ratios.[15] In conclusion, the support for racial segregation declined steadily over the period in question, resulting in a decline in the absolute college gap, but the relative difference between the non–college educated and college educated did not change.

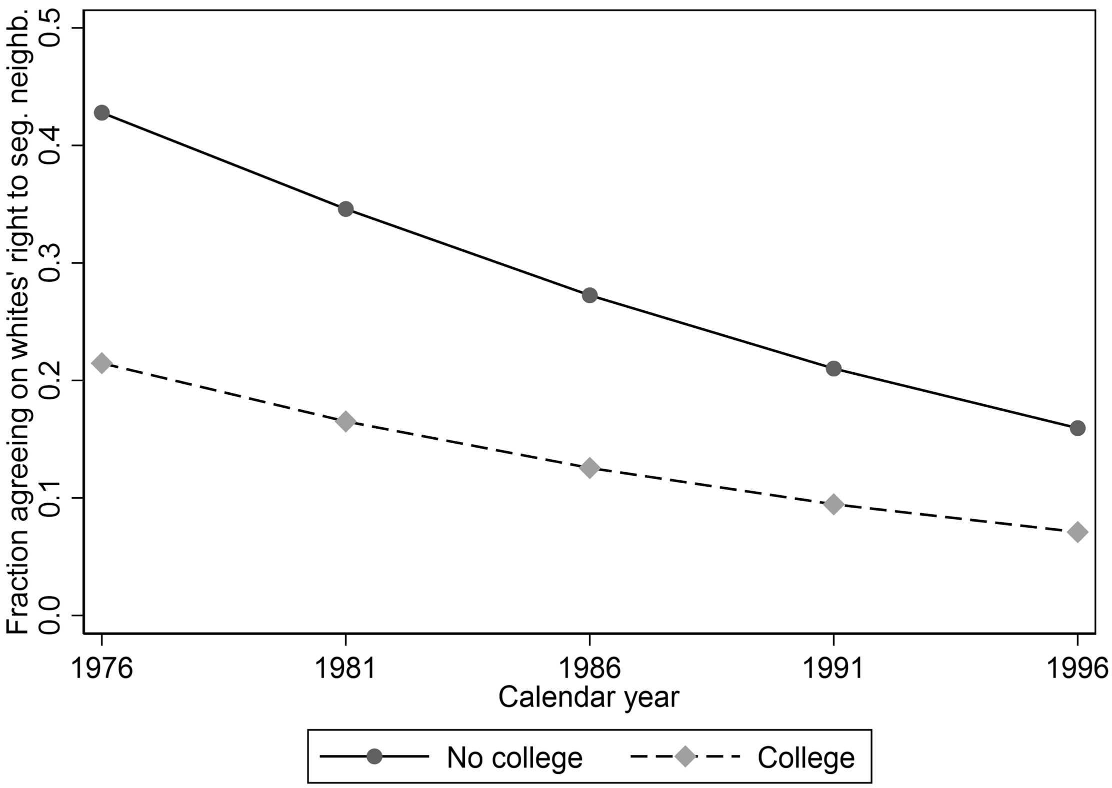

Figure 1: Trends among college-educated and non–college-educated individuals in the attitude toward whites' right to live segregated from blacks, 1976 to 1996: average marginal predictions. *Notes*: Estimates are adjusted for gender, race, age, marital status, and overall political view. Data are from the General Social Survey cumulative file, $N = 12,239$.

Discussion

We have introduced to sociologists the marginal odds ratio, an odds ratio that "behaves like" the increasingly popular average marginal effect (on the probability scale): the marginal odds ratio is unaffected by noncollapsibility, has a population-averaged interpretation, and is "derived from" a given model. We have demonstrated the close relationship between the marginal odds ratio and average marginal effects, and we have outlined why we believe that marginal odds ratios should be preferred over conditional odds ratios in many areas of sociology.

In addition to introducing to sociologists the marginal odds ratio as a complement to the reporting of average marginal effects, our defining the marginal odds ratio in terms of potential outcomes also highlights the crucial distinction between estimands and estimation (Lundberg, Johnson, and Stewart 2021). Many sociologists think of odds ratios as the exponentiated coefficients from logistic response models and have been trained in interpreting these coefficients as if they behave like coefficients from linear regression models. By separating the estimands from their estimation, as we do in this article, we hope to contribute to sociologists being more precise about the quantities they are interested in estimating.

We have also presented empirical examples to illustrate the uses and interpretation of marginal odds ratios relative to the conditional counterparts or the average marginal effect. Although these examples are stylized, they represent types of analyses that are widespread in mainstream sociology. We show how overall conclusions can depend on the chosen estimand. As all of the estimands are equally valid from a statistical perspective, the choice should depend on the research question. For the examples we provided, the marginal odds ratio appears as an obvious candidate. However, in most applied research, it will be useful to report and interpret estimates of several estimands. In particular, reporting both average marginal effects and marginal odds ratios could be very informative about absolute and relative differences, even if it results in the "flipped-signs phenomenon" (Bloome and Ang 2022).

For readers interested in an in-depth description of estimation techniques and the estimands they each recover, including user-written Stata software that implements the discussed methods, we refer to our technical companion paper (Jann and Karlson 2023). In the replication package for this article, we share code and sample data that reproduce the two examples reported earlier. We hope that these tools will urge sociologists to consider using the marginal odds ratio and reporting it as a complement to average marginal effects in substantive research.

Notes

1 Being a ratio, the marginal odds ratio can take on values between zero and infinity, and a value of one means that there is no effect. To obtain a symmetric measure (with zero corresponding to a null effect), one could take the log. Although sociologists employ both the odds ratio and the log odds ratio, we mainly focus on the former in this article as it is more straightforward to interpret in empirical work.

2 Sociologists have also recently begun discussing marginal odds ratios (see Erikson et al. 2005; Breen et al. 2018:46; Kuha and Mills 2020:521–22; Karlson et al. 2021). Moreover, there is a well-established literature in statistics on this topic for clustered or "multilevel" data (see, e.g., Zeger et al. 1988; Agresti 2002). We also draw on these literatures.

3 This definition only holds under the stable unit treatment value assumption (SUTVA).

4 An alternative to using "marginal" in marginal odds ratios would be "unconditional," which is used in the methodological literature on quantile regression where one distinguishes conditional from unconditional quantile regression (Firpo, Fortin, and Lemieux 2009; Killewald and Bearak 2014). However, we adopt the former term because this terminology is already established in the literature (Stampf et al. 2010).

5 This approximation assumes that X is normally distributed, and the number 0.35 is the approximation of $[16\sqrt{3}/(15\pi)]^2$ from the expression derived in Zeger et al. (1988:1054). We obtain a similar approximation if we formulate the logit model in terms of an underlying latent variable model (see Breen et al. 2018).

6 This situation is sometimes referred to as rescaling bias (Karlson et al. 2012; Breen, Karlson, and Holm 2013).

7 Stata command *lnmor* also supports continuous treatments and provides consistent standard errors.

8 In the replication package for this article, we provide the Stata code used for generating the results in this analysis. The National Longitudinal Survey of Youth, 1979 cohort, is available from the Bureau of Labor Statistics; the Danish Longitudinal Survey of Youth is available from the Danish National Archives. The final National Longitudinal Survey of Youth sample size is 10,068; the final Danish Longitudinal Survey of Youth sample size is 2,185.

9 In log odds ratios, academic ability explains 56 and 33 percent of the gap in United States and Denmark, respectively.

10 By the "strength of the predictor," we refer to $\beta^2 \text{var}(X)$ in Equation (10); that is, both the impact of ability and the dispersion in ability affect the degree of attenuation. Because academic ability is a latent variable, we cannot meaningfully disentangle the two here. Had we controlled for a variable with a natural metric (e.g., parental income or number of books in the home), we could have decomposed the attenuating impact into the contribution of each of these two components.

11 The KHB approach uses residualized control variables to make constant the scales of the coefficients across logit models with different covariates (see Karlson et al. 2012).

12 For the CORs reported in Table 1, in log odds ratios academic ability explains 54 and 33 percent of the gap in the United States and Denmark, respectively.

13 For the United States, the percentage explained is 59 percent; for Denmark, 35 percent.

14 In the replication package for this article, we provide the final sample of the General Social Survey used in this example and Stata code for generating the results in this analysis.

15 We also calculated risk ratios with this outcome definition (agreeing to the opinion), and the results are similar to those based on odds ratios, indicating that the relative gap has remained constant over time.

References

Agresti, Alan. 2002. *Categorical Data Analysis*. New York: John Wiley & Sons. https://doi.org/10.1002/0471249688.

Allison, Paul D. 1999. "Comparing Logit and Probit Coefficients across Groups." *Sociological Methods & Research* 28(2):186–208. https://doi.org/10.1177/0049124199028002003.

Breen, Richard, Kristian Bernt Karlson, and Anders Holm. 2013. "Total, Direct, and Indirect Effects in Logit and Probit Models." *Sociological Methods & Research* 42(2):164–91. https://doi.org/10.1177/0049124113494572.

Breen, Richard, Kristian Bernt Karlson, and Anders Holm. 2018. "Interpreting and Understanding Logits, Probits, and Other Nonlinear Probability Models." *Annual Review of Sociology* 44:39–54. https://doi.org/10.1146/annurev-soc-073117-041429.

Bloome, Deidre, and Shannon Ang. 2022. "Is the Effect Larger in Group A or B? It Depends: Understanding Results from Nonlinear Probability Models." *Demography* 59(4):1459–88. https://doi.org/10.1215/00703370-10109444.

Boudon, Raymond. 1974. *Education, Opportunity, and Social Inequality: Changing Prospects in Western Society.* New York: John Wiley and Sons.

Bureau of Labor Statistics, U.S. Department of Labor. 2019. *National Longitudinal Survey of Youth 1979 Cohort, 1979–2016 (Rounds 1–27).* The Ohio State University. Columbus, Ohio: Center for Human Resource Research (CHRR).

Cramer, J. S. 2007. "Robustness of Logit Analysis: Unobserved Heterogeneity and Misspecified Disturbances." *Oxford Bulletin of Economics and Statistics* 69(4):545–55. https://doi.org/10.1111/j.1468-0084.2007.00445.x.

Daniel, Rhian, Jingjing Zhang, and Daniel Farewell. 2021. "Making Apples from Oranges: Comparing Noncollapsible Effect Estimators and Their Standard Errors after Adjustment for Different Covariate Sets." *Biometrical Journal* 63(3):528–57. https://doi.org/10.1002/bimj.201900297.

Erikson, Robert, and John H. Goldthorpe. 1992. *The Constant Flux*. Oxford: Oxford University Press.

Erikson, Robert, John H. Goldthorpe, Michelle Jackson, Meir Yaish, and D. R. Cox. 2005. "On Class Differentials in Educational Attainment." *Proceedings of the National Academy of Sciences* 102(27):9730–33. https://doi.org/10.1073/pnas.0502433102.

Firpo, Sergio, Nicole M. Fortin, and Thomas Lemieux. 2009. "Unconditional Quantile Regressions." *Econometrica* 77(3):953–73. https://doi.org/10.3982/ECTA6822.

Hansen, E. J. 1995. *En Generation Blev Voksen*. Copenhagen: SFI.

Holm, Anders, Mette Ejrnæs, and Kristian Karlson. 2015. "Comparing Linear Probability Model Coefficients across Groups." *Quality & Quantity* 49(5):1823–34. https://doi.org/10.1007/s11135-014-0057-0.

Jackson, Michelle. 2013. *Determined to Succeed? Performance versus Choice in Educational Attainment*. Stanford, CA: Stanford University Press. https://doi.org/10.11126/stanford/9780804783026.001.0001.

Jann, Ben, and Kristian Bernt Karlson. 2023. "Estimation of Marginal Odds Ratios." Working paper. https://ideas.repec.org/p/bss/wpaper/44.html.

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(1):286–313. https://doi.org/10.1177/0081175012444861.

Karlson, Kristian B., and Rasmus Landersø. 2021. "The Making and Unmaking of Opportunity: Educational Mobility in 20th Century Denmark." IZA Discussion Paper Series, IZA DP No. 14135. https://www.iza.org/publications/dp/14135/.

Karlson, Kristian Bernt, Frank Popham, and Anders Holm. 2021. "Marginal and Conditional Confounding Using Logits." *Sociological Methods & Research*, published in advance online. https://doi.org/10.1177/004912412199554.

Killewald, A., and J. Bearak. 2014. "Is the Motherhood Penalty Larger for Low-Wage Women? A Comment on Quantile Regression." *American Sociological Review* 79(2):350–57. https://doi.org/10.1177/0003122414524574.

Kuha, Jouni, and Colin Mills. 2020. "On Group Comparisons with Logistic Regression Models." *Sociological Methods & Research* 49(2):498–525. https://doi.org/10.1177/0049124117747306.

Long, J. Scott, and Sarah A. Mustillo. 2021. "Using Predictions and Marginal Effects to Compare Groups in Regression Models for Binary Outcomes." *Sociological Methods & Research* 50(3):1284–320. https://doi.org/10.1177/0049124118799374.

Lundberg, Ian, Rebecca Johnson, and Brandon M. Stewart. 2021. "What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory." *American Sociological Review* 86(3):532–65. https://doi.org/10.1177/00031224211004187.

Mare, Robert D. 1981. "Change and Stability in Educational Stratification." *American Sociological Review* 46(1):72–87. https://doi.org/10.2307/2095027.

Mize, Trenton D. 2019. "Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects." *Sociological Science* 6:81–117. https://doi.org/10.15195/v6.a4.

Mize, Trenton D., Long Doan, and J. Scott Long. 2019. "A General Framework for Comparing Predictions and Marginal Effects across Models." *Sociological Methodology* 49(1):152–89. https://doi.org/10.1177/0081175019852763.

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(1):67–82. https://doi.org/10.1093/esr/jcp006.

Norton, Edward C., and Bryan E. Dowd. 2018. "Log Odds and the Interpretation of Logit Models." *Health Services Research* 52(2):859–78. https://doi.org/10.1111/1475-6773.12712.

Pang, Menglan, Jay S. Kaufman, and Robert W. Platt. 2016. "Studying Noncollapsibility of the Odds Ratio with Marginal Structural and Logistic Regression Models." *Statistical Methods in Medical Research* 25(5):1925–37. https://doi.org/10.1177/0962280213505804.

Robins, James. 1986. "A New Approach to Causal Inference in Mortality Studies with a Sustained Exposure Period—Application to Control of the Healthy Worker Survivor Effect." *Mathematical Modelling* 7(9–12):1393–1512. https://doi.org/10.1016/0270-0255(86)90088-6.

Rubin, D. B. 1974. "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies." *Journal of Educational Psychology* 66(5):688–701. https://doi.org/10.1037/h0037350.

Smith, Tom W., Michael Davern, Jeremy Freese, and Stephen L. Morgan. 2019. *General Social Surveys, 1972–2018* [machine-readable data file]. Principal Investigator, Tom W. Smith; Co-Principal Investigators, Michael Davern, Jeremy Freese, and Stephen L. Morgan; Sponsored by National Science Foundation. Chicago: NORC.

Schuster, Noah A., Jos W. R. Twisk, Gerben ter Riet, Martijn W. Heymans, and Judith J. M. Rijnhart. 2021. "Noncollapsibility and Its Role in Quantifying Confounding Bias in Logistic Regression." *BMC Medical Research Methodology* 21(1):136. https://doi.org/10.1186/s12874-021-01316-8.

Stampf, Susanne, Erika Graf, Claudia Schmoor, and Martin Schumacher. 2010. "Estimators and Confidence Intervals for the Marginal Odds Ratio Using Logistic Regression and Propensity Score Stratification." *Statistics in Medicine* 29(7–8):760–9. https://doi.org/10.1002/sim.3811.

Zeger, Scott L., Kung-Yee Liang, and Paul S. Albert. 1988. "Models for Longitudinal Data: A Generalized Estimating Equation Approach." *Biometrics* 44(4):1049–60. https://doi.org/10.2307/2531734.

Zhang, Zhiwei. 2008. "Estimating a Marginal Causal Odds Ratio Subject to Confounding." *Communications in Statistics - Theory and Methods* 38(3):309–21. https://doi.org/10.1080/03610920802200076.

Acknowledgments: We thank the following for invaluable comments and feedback: Tim Liao, Mike Hout, Rudolf Farys, and Jesper Fels Birkelund, as well as participants at the Hans Schadee Research Methods Center Seminar on November 3, 2022, at Trento University; the Seminar on Analytical Sociology on November 14–17, 2022, at Venice International University; and the 2022 Swiss Stata Meeting on November 18, 2022, at University of Bern. For Kristian Bernt Karlson, the research leading to the results presented in this article has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement no. 851293). Replication materials for the examples reported in this article are available here: https://osf.io/xkre6/.

Kristian Bernt Karlson: Department of Sociology, University of Copenhagen. E-mail: kbk@soc.ku.dk.

Ben Jann: Institute of Sociology, University of Bern. E-mail: ben.jann@unibe.ch.