At Odds: Concerns Raised by Using Odds Ratios for Continuous or Common Dichotomous Outcomes in Research on Physical Activity and Obesity

Gina S. Lovasi[*], Lindsay J. Underhill, Darby Jack, Catherine Richards, Christopher Weiss and Andrew Rundle

Columbia University, Mailman School of Public Health, Department of Epidemiology, 722 W 168th St, Room 804, New York, NY 10032, USA

Abstract: *Purpose:* Research on obesity and the built environment has often featured logistic regression and the corresponding parameter, the odds ratio. Use of odds ratios for common outcomes such obesity may unnecessarily hinder the validity, interpretation, and communication of research findings. *Methods:* We identified three key issues raised by the use of odds ratios, illustrating them with data on walkability and body mass index from a study of 13,102 New York City residents. *Results:* First, dichotomization of continuous measures such as body mass index discards theoretically relevant information, reduces statistical power, and amplifies measurement error. Second, odds ratios are systematically higher (further from the null) than prevalence ratios; this inflation is trivial for rare outcomes, but substantial for common outcomes like obesity. Third, odds ratios can lead to incorrect conclusions during tests of interactions. The odds ratio in a particular subgroup might higher simply because the outcome is more common (and the odds ratio inflated) compared with other subgroups. *Conclusion:* Our recommendations are to take full advantage of continuous outcome data when feasible and to use prevalence ratios in place of odds ratios for common dichotomous outcomes. When odds ratios must be used, authors should document outcome prevalence across exposure groups.

Keywords: Physical activity, obesity, logistic regression, odds ratio, prevalence ratio, dichotomization.

INTRODUCTION

Research on environmental determinants of physical activity and obesity [1-4] has generated interest among urban planners and public health practitioners, and contributes to ongoing policy discussions [5, 6]. Yet one of the barriers to consistency and interpretability of results is the use of suboptimal data analysis strategies including logistic regression.

The Active Living Research online literature database which brings together much of the published work on environmental determinants of physical activity and obesity [7] reveals that 44% of papers with quantitative results reported odds ratios (Fig. **1**). Although common, reliance on odds ratios may hinder the validity and accurate communication of research. We identified key concerns with logistic regression in physical activity and obesity research (Table **1**), issues which also apply to other research fields [8, 9].

DICHOTOMIZATION OF CONTINUOUS MEASURES DISCARDS INFORMATION

The use of logistic regression frequently involves dichotomizing continuous measures such as physical activity or body mass index (BMI) (Fig. **1**). Although dichotomization may be the best choice for some questions and audiences [10], we have identified three key problems with dichotomizing continuous outcomes.
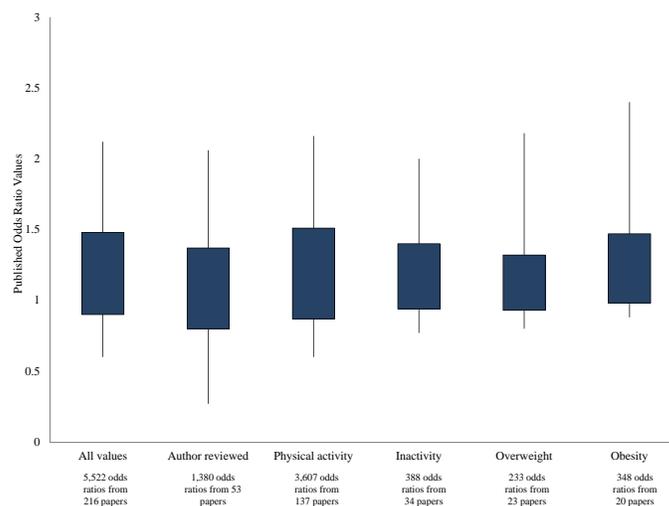
*Address correspondence to this author at the Columbia University, Mailman School of Public Health, Department of Epidemiology, 722 W 168th St, Room 804, New York, NY 10032, USA; Tel: 212-342-0446; Fax: 212-842-5169; E-mail: glovasi@columbia.edu

Fig. (1). Odds ratios commonly reported in active living research. Notes: Box plots indicate the interquartile range (the rectangle is bounded by the 75th and 25th percentiles) and the interdecile range (lines extend to the 90th and 10th percentiles); variation in the distribution is shown for the entire Active Living Research Literature Database [7], the subset for which authors had reviewed and confirmed the database entry, and for four commonly reported outcomes: meeting physical activity recommendations, inactivity, overweight, and obesity.

First, theories about how environments influence physical activity and body weight generally suggest that the relationships are continuous. Conceptual models or frameworks guide the selection and organization of measures

Table 1.   Problems Arising from Over-Use of Odds Ratios, Along with Recommended Strategies and their anticcipated Benefits

Category	Potential Problems	Recommended Strategies	Anticipated Benefits
Dichotomization of continuous measures	Theory implies continuous, subtle influence of environment rather than referencing a threshold Information is discarded, reducing statistical power Unbiased measurement error will increase extreme values, altering the proportion exceeding a threshold	Use continuous measures to take full advantage of collected data Consider generalized linear models and elasticities Minimize measurement error and be cautious when discussing results derived from an imprecise, dichotomized measure	Flexibility to examine the form of the association is preserved Smaller sample sizes are needed to detect a statistically significant effect
Odds ratio interpretation	Odds ratios are difficult to interpret and communicate except when they approximate prevalence ratios Odds ratios diverge from prevalence ratios as the outcome prevalence in the reference group increases The magnitude of association may be misunderstood, leading to unrealistic estimates of intervention benefits and subsequent disappointment	Use prevalence ratios in place of, or to complement, odds ratios Provide outcome prevalence information to give context when odds ratios must be presented	Research results can be stated clearly and accurately in terms of how much more likely the outcome was in the exposed group Informed readers can use prevalence information to estimate prevalence ratios from published odds ratios
Interactions on an odds ratio scale	Interactions of scientific interest may be missed on an odds ratio scale Interactions on the odds ratio scale may result from a difference in outcome prevalence across subgroups, even if the prevalence ratios are the same within each subgroup	Consider using continuous outcome data to maximize statistical power to detect an interaction Conduct the interaction analysis that corresponds to the scientific question Provide outcome prevalence data for each subgroup if odds ratios are being compared across subgroups	Interaction analyses are more likely to yield scientifically interesting results Hypotheses regarding effect modification will receive support or be dismissed on the basis of valid tests

[11-14] which range from personal to societal, from distal to proximate, or along some other categorization scheme. Since numerous influences are proposed, each presumably has only a small effect on behavior or health. An exposure that is causally related to BMI will only occasionally cause an individual to cross a threshold value such as a BMI of 30. These exposures may nonetheless be important for their potential to generate incremental behavioral and health improvements for an entire population.

In addition, dichotomization is problematic because information is discarded [15-17]. A questionnaire, motion sensor, or anthropometry protocol captures a wide range of variation. Dichotomization ignores much of this variation and leads to a decrease in statistical power. Power loss depends on other study characteristics as well [18, 19], but a large increase in the sample size may be required to compensate for the dichotomization of an outcome variable [20]. As an example, consider a study of residential density and body mass index [21, 22]. This study, which had a large sample size (N = 13,102) reported a decrease of 0.4 BMI units for each 10,000 people/km^2 increase in residential density [23]. We re-analyzed subsets of the data using linear models of BMI and logistic regression for the dichotomous outcome of obesity (BMI≥30); methods and adjustments were otherwise identical. We were able to detect statistical significance for all continuous models with at least 783 randomly selected participants, but statistical significance was not consistently reached for our logistic regression models until the sample size was at least 1,248. Thus, we found that the sample size would have to be almost doubled

before a continuous association of interest was detected in logistic regression. Dichotomization of continuous measures may thus contribute to Type II error [17, 24].

A final concern about dichotomization is that it exaggerates misclassification. Physical activity and adiposity are difficult to measure, and common approaches have limited reliability and validity [25-27]. Even unbiased measurement error (mean error=0) in a continuous measure will affect the proportion exceeding a threshold [28]. One can visualize this by considering the distribution of BMI from the population of 13,102 adults discussed earlier [21, 22]. These data, based on heights and weights measured by trained staff, indicate that 28.8% of the study participants were obese. If we add random, unbiased error of up to 10 BMI units in either direction, we would find that 37.8% of participants met our criteria for obesity (Fig. **2**). The existence of nonrandom error such as social desirability bias may further complicate the picture by having differential effects across the BMI distribution or across other groups of interest.

Our recommendation is to use statistical approaches that take full advantage of continuous outcome data; useful strategies may include linear models, generalized linear models, zero inflated Poisson models, or proportional hazards models. In addition, studies should be designed to minimize measurement error, and should interpret with caution the proportion above a threshold in the presence of measurement error in the underlying continuous variable.
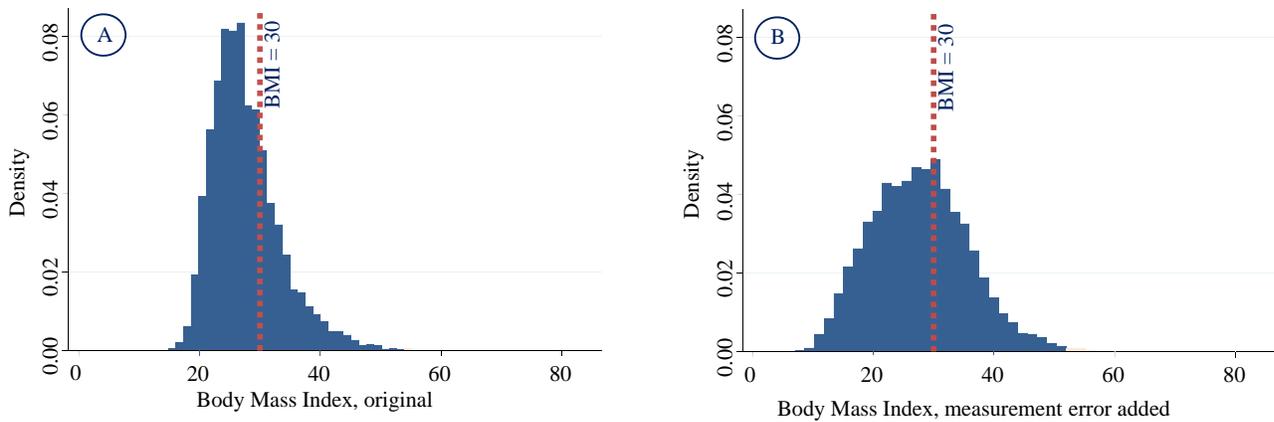
Fig. (2). Measurement error in a continuous variable affects the proportion exceeding a threshold. Notes: Histograms are shown for body mass index in (**A**) the 13,102 New York City residents in the years 2000-2002 [21], with 28.8% obesity based on the proportion of observations greater than or equal to 30 and (**B**) a hypothetical set of observations created by adding random error of up to 10 BMI units in either direction, with 37.8% obesity.

ODDS RATIOS MAY MISLEAD WHEN THE OUTCOME IS COMMON

For rare outcomes affecting <10% of the population [8, 29], the odds ratio approximates the prevalence ratio (also referred to as the probability ratio, risk ratio, or relative risk). However, for common outcomes odds ratios are systematically more extreme (further from the null) than the corresponding prevalence ratios [29-31]. For the common magnitudes of association, odds ratios are markedly different from the underlying prevalence ratios, being 50% to 400% further from the null value of 1 (Fig. **3**). In our study of obesity in New York City [21], participants living in low-density neighborhoods (defined as the lowest quartile of population density) had a 30% higher odds of obesity (OR=1.3). But the corresponding prevalence ratio of 1.2 indicates that the *probability* of obesity was only 20% higher in the lowest density quartile compared with the other three quartiles. For stronger associations or more common outcomes, the difference would be larger.
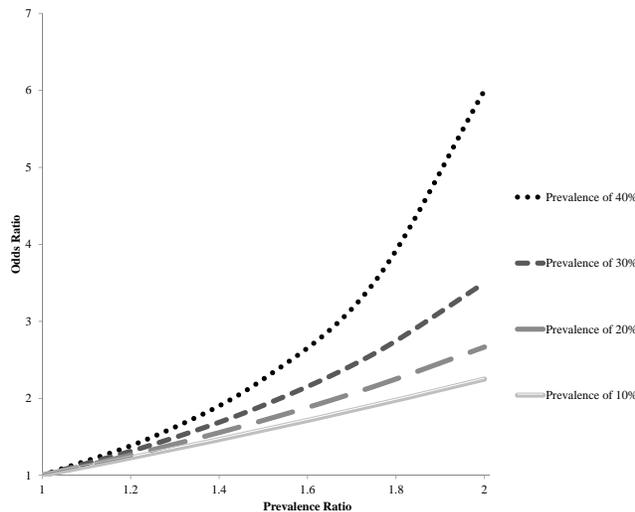

Fig. (3). Odds ratios diverge from prevalence ratios as outcome prevalence in the reference group increases.

Odds ratios and prevalence ratios contain essentially similar information, but are numerically different. If described and interpreted correctly, the difference between these approaches and the parameters of interest need not present a problem. A problem often arises, however, when investigators try to explain the magnitude of odds ratios [8, 9, 32]. The magnitude of association may become particularly important when research is used to assess attributable risk, drive cost benefit analyses, or shape policy goals. It is very tempting to interpret an odds ratio of 3 in an obesity study as meaning that obesity is three-times as likely in the exposed group. However, an odds ratio of 3 may correspond to a prevalence ratio of only 2 (Fig. **3**).

As above, we caution against dichotomizing continuous measures, preferring methods that use all of the theoretically relevant data. However, we recognize that there are circumstances that may encourage or compel an investigator to use a dichotomous version of a continuous measure [10]. For example, clinical and policy audiences may prefer a message framed in terms of reducing obesity risk, rather than decreasing body mass index. Relative risk regression [29, 33, 34] can be used in place of logistic regression (Box **1**). A simple formula is available for estimating prevalence ratios from published odds ratios [30], but requires that the outcome prevalence in the reference group be known. This formula offers only an approximation for adjusted models, in comparison to regression methods that directly estimate prevalence ratios with adjustment on the same scale.

THE ODD MEANING OF INTERACTIONS ON AN ODDS RATIO SCALE

The loss of statistical power due to dichotomization of a continuous outcome may undermine one's ability to detect effect modification. More importantly, apparent interactions may appear in analyses using odds ratios that would not be evident in analyses based on prevalence ratios; conversely an interaction on the prevalence ratio scale may be obscured by using odds ratios [35-37].

A higher odds ratio in a particular subgroup might be observed simply because the outcome is more common in that group. As an example, consider fast food restaurant

Box 1.   Methods for calculating prevalence ratios

① **DIRECT CALCULATION OF A PREVALENCE RATIO**

$$\frac{\text{Probability of outcome in exposed group}}{\text{Probability of outcome in reference group}}$$

② **APPROXIMATING A PREVALENCE RATIO FROM A PUBLISHED OR (FROM ZHANG ET AL. JAMA 1998):**

$$\frac{\text{OR}}{(1\text{-probability in reference group}) + (\text{probability in reference group x OR})}$$

This post hoc calculation can be an aid to interpretation, but is not entirely satisfactory in a multivariable analysis because the scale of adjustment does not correspond to the scale of the parameter of interest.

③ **ESTIMATING PREVALENCE RATIOS WITH ROBUST STANDARD ERRORS USING STATISTICAL SOFTWARE:**

Stata,
    glm y x, link(log) family(binomial) eform *(options to try: difficult, search)*
    glm y x, link(log) family(poisson) eform robust *(if above doesn't converge)*

R or S+:
    glm y ~ x, family=binomial(log) *(R only, S-plus reads log as logit)*
    glm y ~ x, family=poisson(log)

SAS,
    proc genmod; class id; model y = x /dist=bin link=log;
    proc genmod data = poissonreg; class id; model y = x /dist=poisson; repeated subject=id/type=ind;

SPSS,
    genlin y with x /model x distribution = binomial link = log /criteria covb = robust
    genlin y with x /model x distribution = poisson link = log /criteria covb = robust

proximity and obesity prevalence among each of 4 age groups. Suppose obesity prevalence varies from 10 to 40 percent among age groups. If the prevalence ratio in each age group was 2.0, odds ratios would be 2.25 in the low-prevalence group and 6.0 in the high-prevalence group (see right side of Fig. **3**). This odds ratio "interaction" is difficult to explain, potentially misleading, and not well aligned with a scientific interest the pattern of association between fast food restaurants and the probability of obesity. This interaction fallacy [35] not only affects interaction analyses within a single study, but also has potential to bias meta-analyses that integrate effect estimates from multiple studies, particularly if thresholds used are not consistent across studies [19, 24, 31, 37].

The proposed alternative of using prevalence ratios rather than odds ratios (Table **1**, Box **1**) should be given strong consideration when assessing interactions. When odds ratios are used to define an interaction, outcome prevalence should be shown by subgroup.

CONCLUSIONS

For research on physical activity, obesity, or other common outcomes, odds ratios should be viewed critically because of the information lost through dichotomization of continuous measures and the mismatch between the odds ratio scale and the scientific questions of interest. Continuous outcomes should be used to take full advantage of the collected data, particularly in the context of small sample sizes or substantial measurement error. When a dichotomous outcome must be used, prevalence ratios are easier to understand and communicate. When odds ratios must be used, presentation of outcome prevalence can facilitate interpretation.

CONFLICT OF INTEREST

The authors have no conflicts of interest to report.

REFERENCES

[1]   Papas MA, Alberg AJ, Ewing R, *et al*. The built environment and obesity. Epidemiol Rev 2007; 29: 129-43.

[2]   Davison KK, Lawson CT. Do attributes in the physical environment influence children's physical activity? A review of the literature. Int J Behav Nutr Phys Act 2006; 3: 19.

[3]   TRB. TRB Special Report 282: Does the built environment influence physical activity? Examining the evidence. Washington, DC: The National Academies, Transportation Research Board 2005 January. Report No.: 282.

[4]   Lee C, Moudon AV. Physical activity and environment research in the health field: implications for urban and transportation planning practice and research. J plann Lit 2004; 19(2): 147-81.

[5]   Ottoson JM, Green LW, Beery WL, *et al*. Policy-contribution assessment and field-building analysis of the Robert Wood Johnson Foundation's Active Living Research Program. Am J Prev Med 2009; 36(2 Suppl): S34-43.

[6]   York CoN. Active Living Guidelines: Promoting Physical Activity and Health in Design. New York City 2010.

[7]   ALR. Active Living Research Database. 2007 [December, 2009]; Available from: http: //www.activelivingresearch.org/resourcesearch/literaturedatabase.

[8]   Sinclair JC, Bracken MB. Clinically useful measures of effect in binary analyses of randomized trials. J Clin Epidemiol 1994; 47(8): 881-9.

[9]   Liberman AM. How much more likely? The implications of odds ratios for probabilities. Am J Evaluation 2005; 26(2): 253.

[10]   DeCoster J, Iselin AM, Gallucci M. A conceptual and empirical examination of justifications for dichotomization. Psychol Methods 2009 ; 14(4): 349-66.

[11]   Stokols D. Establishing and maintaining healthy environments: Toward a social ecology of health promotion. Am Psychol 1992; 47: 6 - 22.

[12]   Sallis JF, Cervero RB, Ascher W, Henderson KA, Kraft MK, Kerr J. An ecological approach to creating active living communities. Annu Rev Public Health 2006; 27: 297-322.

[13]   Schulz AJ, Kannan S, Dvonch JT, *et al*. Social and physical environments and disparities in risk for cardiovascular disease: the healthy environments partnership conceptual model. Environ Health Perspect 2005; 113(12): 1817-25.

[14]   Moudon AV, Lee C. Walking and bicycling: an evaluation of environmental audit instruments. Am J Health Promot 2003 ; 18(1): 21-37.

[15]   Cohen J. The cost of dichotomization. Appl Psychol Meas 1983; 7(3): 249.

[16]   MacCallum RC, Zhang S, Preacher KJ, Rucker DD. On the practice of dichotomization of quantitative variables. Psychol Methods 2002; 7(1): 19-40.

[17]   Maxwell SE, Delaney HD. Bivariate median splits and spurious statistical significance. Psychological Bulletin 1993; 113(1): 181.

[18]   Delucchi KL. Sample size estimation in research with dependent measures and dichotomous outcomes. Am J Public Health 2004; 94(3): 372-7.

[19]   Selvin S. Two issues concerning the analysis of grouped data. Eur J Epidemiol 1987; 3(3): 284-7.

[20]   Bhandari M, Lochner H, Tornetta P, 3rd. Effect of continuous versus dichotomous outcome variables on study power when sample sizes of orthopaedic randomized trials are small. Arch Orthop Trauma Surg 2002; 122(2): 96-8.

[21]   Rundle A, Roux AV, Freeman L, Miller D, Neckerman KM, Weiss CC. The urban built environment and obesity in New York City: a multilevel analysis. Am J Health Promot 2007; 21(4 Suppl): 326-34.

[22]   Lovasi GS, Neckerman KM, Quinn JW, Weiss CC, Rundle A. Effect of individual or neighborhood disadvantage on the association between neighborhood walkability and body mass index. Am J Public Health 2009 ; 99(2): 279-84.

[23]   Lovasi GS, Quinn JW, Neckerman KM, Weiss CC, Thomas JS, Rundle A. Walkability and obesity in New York City: an examination of differences by scale and subpopulation. Walk21 Conference; New York, NY2009.

[24]   Hunter JE, Schmidt FL. Dichotomization of continuous variables: The implications for meta-analysis. J Appl Psychol 1990; 75(3): 334.

[25]   Sallis JF, Saelens BE. Assessment of physical activity by self-report: status, limitations, and future directions. Res Q Exerc Sport 2000; 71(2 Suppl): S1-14.

[26]   Ferrari P, Friedenreich C, Matthews CE. The role of measurement error in estimating levels of physical activity. Am J Epidemiol 2007; 166(7): 832-40.

[27]   Garcia-Marcos L, Valverde-Molina J, Sanchez-Solis M, *et al*. Validity of parent-reported height and weight for defining obesity among asthmatic and nonasthmatic schoolchildren. Int Arch Allergy Immunol 2006; 139(2): 139-45.

[28]   Armstrong B, White E, Saracci R. Exposure measurement error and its effects. Principles of Exposure Measurement in Epidemiology. New York: Oxford University Press Inc. 2000; p. 49-77.

[29]   McNutt LA, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. Am J Epidemiol 2003; 157(10): 940-3.

[30]   Zhang J, Yu KF. What's the relative risk? A method of correcting the odds ratio in cohort studies of common outcomes. JAMA. 1998; 280(19): 1690-1.

[31]   Ragland DR. Dichotomizing continuous outcome variables: dependence of the magnitude of association and statistical power on the cutpoint. Epidemiology 1992; 3(5): 434-40.

[32]   Pepe MS, Janes H, Longton G, Leisenring W, Newcomb P. Limitations of the odds ratio in gauging the performance of a diagnostic, prognostic, or screening marker. Am J Epidemiol 2004; 159(9): 882-90.

[33]   Lumley T, Kronmal RA, Ma Y. Relative risk regression in medical research: models, contrasts, estimators, and algorithms. Biostatistics Working Paper Series2006.

[34]   Spiegelman D, Hertzmark E. Easy SAS calculations for risk or prevalence ratios and differences. Am J Epidemiol 2005; 162(3): 199-200.

[35]   Morabia A, Ten Have T, Landis JR. Interaction fallacy. J Clin Epidemiol 1997; 50(7): 809-12.

[36]   Campbell UB, Gatto NM, Schwartz S. Distributional interaction: Interpretational problems when using incidence odds ratios to assess interaction. Epidemiol Perspect Innov 2005; 2(1): 1.

[37]   Veiel HO. Base-rates, cut-points and interaction effects: the problem with dichotomized continuous variables. Psychol Med 1988; 18(3): 703-10.