

Ratio Variables in Regression: Implications for Cardiovascular Research

Zachary R. McCaw^{1*}, Thomas W. Soare¹, Colm O'Dushlaine^{1,2}

¹Instituto, South San Francisco, CA, USA

²Eli Lilly, Boston, MA, USA

*Correspondence should be addressed to Zachary R. McCaw, zmccaw@alumni.harvard.edu

Received date: November 30, 2025, **Accepted date:** March 11, 2026

Citation: McCaw ZR, Soare TW, O'Dushlaine C. Ratio Variables in Regression: Implications for Cardiovascular Research. J Clin Cardiol. 2026;7(1):58–61.

Copyright: © 2026 McCaw ZR, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Ratio variables, such as body mass index (BMI) and left ventricular ejection fraction (LVEF), are deeply ingrained in cardiovascular research and practice, yet the use of ratios in regression models can confuse associations and obscure underlying mechanisms. Drawing on examples from genome-wide association studies of adiposity traits and echocardiographic studies of cardiotoxicity, we illustrate how spurious correlation, mathematical coupling, and collider bias can arise when ratios are analyzed without attention to their component variables. Associations with ratio traits like BMI or LVEF are not unique to the ratio; instead, they intermix signals for the numerator and denominator. Regressing ratios on their components can generate tautological findings that are difficult to translate or interpret clinically. We propose a practical decision framework to guide analysis that begins with separate regressions of the numerator and denominator on the exposure, uses adjusted models when the exposure is independent of the denominator, and otherwise favors multivariate approaches, adopting a ratio outcome as a fallback. We encourage cardiovascular investigators to view ratios as one of several competing representations rather than default endpoints, and to choose modeling strategies that promote interpretability and clinical utility. In highlighting limitations of ratios for mechanistic and etiologic research, we do not challenge their use in evidence-based and guideline-supported clinical practice.

Keywords: Cardiovascular research, Body mass index (BMI), Left ventricular ejection fraction (LVEF), Heart failure

Ratio Variables in Cardiovascular Research

From lipid profiles to heart failure classification, ratios are ubiquitous in clinical cardiology. The ratios of total cholesterol to high-density lipoprotein (HDL), low-density lipoprotein to HDL, and triglycerides to HDL have all been studied for stratifying atherosclerotic disease risk [1,2]. The albumin-to-creatinine ratio is a critical indicator of kidney dysfunction and plays a key role in hypertension management [3]. Left ventricular ejection fraction, the ratio of stroke volume to end-diastolic volume, underpins the classification and treatment of heart failure [4]. Echocardiographic E/e' , the ratio of peak early mitral inflow velocity (E) to early diastolic mitral annular velocity (e'), contributes to the estimation of left ventricular filling pressure and the diagnosis of heart failure with preserved ejection fraction [5]. Fractional flow reserve, the ratio of maximal blood flow in a stenotic artery to normal maximal flow, complements angiography to guide decisions regarding percutaneous coronary intervention [6]. Body mass index (BMI), the ratio of weight in kilograms to

height² in meters², remains a central metric in cardiovascular risk assessment [7]. Although ratios play an important role in established guidelines and evidence-based clinical practice, they have important limitations as applied to understanding the relationships among clinical or physiological variables. Our recent work in *Human Genetics and Genomics Advances* [8] highlights pitfalls in performing genome-wide association studies (GWAS) with ratio outcomes, and these difficulties extend to any regression analysis involving ratios [9].

Ratios are appealing when the numerator and denominator covary biologically or clinically. Taller patients are expected to weigh more; larger ventricles are expected to eject more blood; and flow through wider vessels is expected to be greater. The goal of forming a ratio is to adjust the numerator for the denominator. The underlying hypothesis is that the value of the numerator *relative* to the denominator is more informative than the value of the numerator in isolation. Forming a ratio compresses two dependent variables into a single number that clinicians can interpret and act upon. While ratios have

clear applications when considering only two variables, they can obscure rather than clarify relationships when considering three or more. Two variables that are otherwise independent become dependent when normalized by a common denominator, a phenomenon described by Karl Pearson as *spurious correlation* [10]. In regression settings, Richard Kronmal [9] cataloged how using ratios as either outcomes or predictors can result in misattribution of effects.

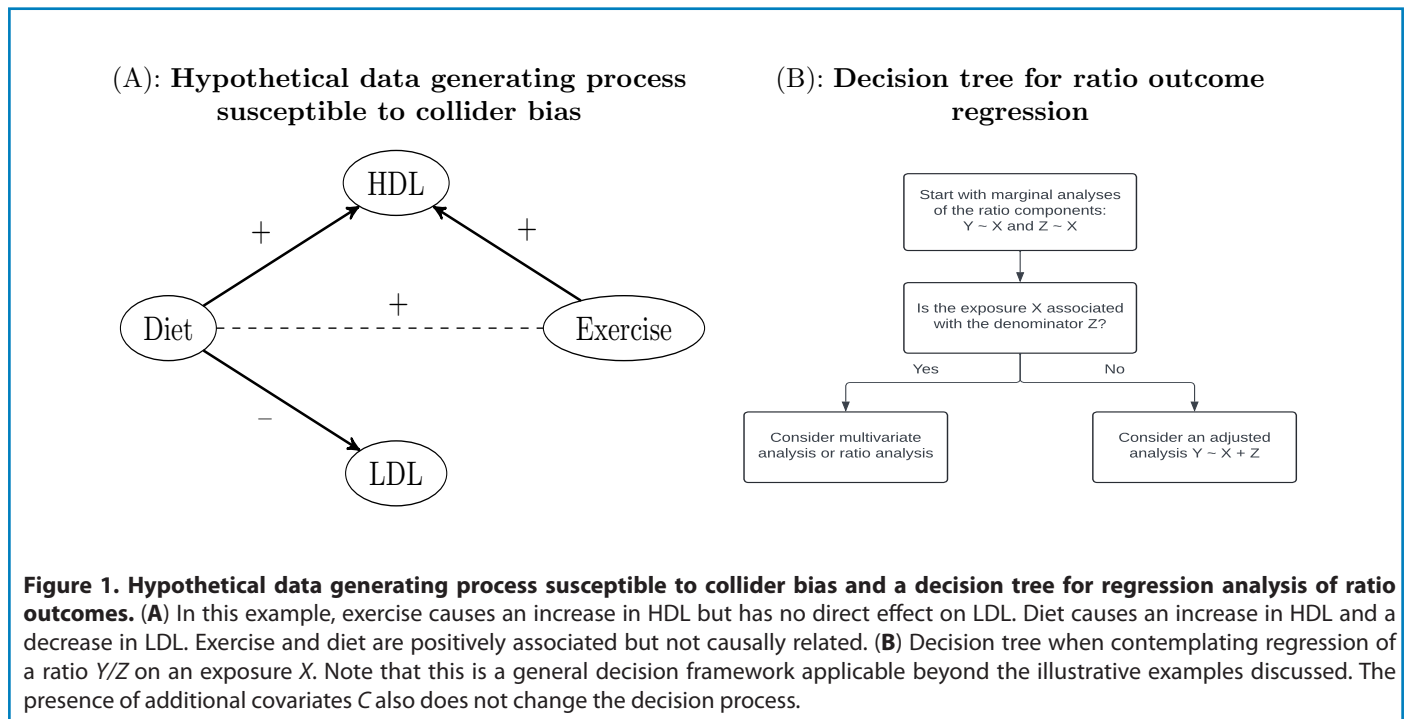
BMI has become an established index of adiposity due to its ease of measurement [11]. The genetics of BMI have been studied since the earliest applications of GWAS [12], and remain under active investigation [13]. While BMI provides a pragmatic height-normalized measure of excess adiposity across body sizes, its clinical utility does not guarantee suitability for etiologic research. Uncovering the genetic determinants of adiposity through BMI is challenging in part because BMI is a composite phenotype formed as the ratio of two heritable traits. Genetic associations with ratios are not unique to the ratio [8]; they can arise from association with the numerator, the denominator, or both. In samples of sufficient size, any variant associated with either the numerator or the denominator will be associated with the ratio. Consequently, the set of genetic variants truly associated with BMI is a mixture of variants affecting height and variants affecting weight – some, but not all, of which are relevant to adiposity. Although forming the ratio may improve power for detecting BMI-associated variants, it complicates the interpretation of all variants by intermixing adiposity variants with signals for the component traits. Importantly, a stronger statistical association with a ratio does not imply that the ratio constitutes an independent biological construct; signal strength can be amplified simply because the associations with numerator and denominator have opposite directions [8]. Waist-to-hip ratio, a key measure of central obesity and major cardiovascular risk factor [14], suffers the same drawbacks.

The confusion caused by ratios is not confined to GWAS. In a prospective longitudinal analysis among breast cancer patients treated with doxorubicin and/or trastuzumab, Narayan and colleagues [15] related within patient changes in left-ventricular ejection fraction (LVEF) to changes in echocardiographic parameters. On average, LVEF was modestly decreased at 1-3 years post treatment initiation. Increases in LV end-systolic volume (ESV) and end-diastolic volume (EDV) were each associated with lower LVEF. Because EDV appears in the denominator of LVEF ($= SV/EDV$), an increase in EDV automatically decreases LVEF even if stroke volume (SV) remains unchanged. In this way, LVEF and EDV are said to be *mathematically coupled* [16]. End-systolic volume is likewise coupled with LVEF through the relation $ESV = EDV - SV$. An increase in ESV can result from an increase in EDV, a decrease in SV, or both. Regression of a ratio outcome (LVEF) on mathematically coupled components (ESV and EDV) is

tautological: an association exists by construction. Moreover, analysis of the ratio obscures the physiological mechanism. Without analyzing the components separately, it is unclear whether the reduction in LVEF is due to a decrease in SV, an increase in EDV, or both. This example illustrates how default reliance on ratios can muddle rather than clarify mechanisms, yielding conclusions that offer limited insight. More broadly, this motivates alternative modeling strategies that disentangle component effects rather than collapsing them.

To obtain more interpretable effects, the association between the exposure and the numerator can be estimated while adjusting for the denominator as a covariate. For example, an alternative to regressing BMI on genotype is to regress weight on genotype while adjusting for height. The coefficient on genotype from this *adjusted model* estimates the expected change in weight per unit change in genotype (e.g., per additional minor allele) while holding height constant. The motivation here is the same as that which underpins forming the ratio to begin with: that greater weight at fixed height is suggestive of adiposity. While the adjusted model clarifies the interpretation of identified variants, conditioning on the denominator introduces the risk of *collider bias*.

Collider bias can affect the adjusted model when 1. the exposure affects the denominator and 2. there exists a background common cause of both the denominator and the outcome. For an example from cardiology, consider estimating the effect of an exercise program on blood lipid levels at 6-month follow-up. Participants are randomized to receive an exercise program or dietary counseling. Let X denote an indicator of randomization to exercise (the exposure), Y the 6-month LDL, and Z the 6-month HDL. Suppose for exposition that exercise increases HDL while having no direct effect on LDL, and that participants who adhere to the exercise program are more likely to adopt dietary changes that both reduce LDL and increase HDL. A schematic of this scenario is shown in **Figure 1A**. Consider two possible analyses: the first regresses the 6-month LDL/HDL ratio on an indicator of randomization to exercise ($Y/Z \sim X$), and the second regresses LDL on the exercise indicator, adjusting for HDL ($Y \sim X + Z$). The ratio outcome analysis is expected to show that the exercise program reduces LDL/HDL. Although exercise has no direct effect on LDL, it reduces the ratio by increasing HDL. Conditioning HDL in the adjusted analysis induces an association between exercise and LDL where none exists marginally. This is because participants who achieve the greatest increase in HDL will tend to be those who both adhere to the exercise program and make dietary changes, with the latter decreasing LDL. Collider bias likewise impacts GWAS [8,17]. Genetic variants that modulate HDL but have no direct effect on LDL can associate with LDL in an adjusted analysis if there exist background common causes (e.g., diet) of both LDL and HDL.



If ratio outcome analysis lacks interpretability while the adjusted analysis incurs collider bias, what can be done in practice? We make the following recommendations (**Figure 1B**), intended as a guide for investigators designing analyses, reviewers assessing ratio-based claims, and readers interpreting published effect estimates. First, examine marginal associations between the exposure and each component of the ratio to determine if the exposure associates with the numerator, denominator, or both. Second, assess whether the exposure is associated with the denominator, which determines whether adjusted analyses are susceptible to collider bias. Third, when the exposure and denominator are associated, consider multivariate modeling approaches, adopting a ratio outcome only as a fallback. Multivariate analysis can be performed via mixed effects modeling or generalized estimating equations. Even if a ratio outcome is ultimately selected, the preliminary marginal analyses provide critical context, revealing whether the exposure is associated with the ratio through the numerator, denominator, or both. As a final caution, when using a ratio outcome, investigators should check for mathematical coupling with candidate covariates, since such coupling induces mechanical associations.

In conclusion, we encourage researchers and clinicians to be mindful of the ambiguities that ratios can cause, and to evaluate the alternatives before including ratios in mechanistic or etiologic research. Notwithstanding their limitations, this commentary does not seek to challenge the use of ratios in evidence-based and guideline-supported clinical care.

References

1. Millán J, Pintó X, Muñoz A, Zúñiga M, Rubiés-Prat J, Pallardo LF, et al. Lipoprotein ratios: Physiological significance and clinical usefulness in cardiovascular prevention. *Vasc Health Risk Manag.* 2009;5:757–65.
2. Chen Y, Chang Z, Liu Y, Zhao Y, Fu J, Zhang Y, et al. Triglyceride to high-density lipoprotein cholesterol ratio and cardiovascular events in the general population: A systematic review and meta-analysis of cohort studies. *Nutr Metab Cardiovasc Dis.* 2022 Feb;32(2):318–29.
3. McEvoy JW, McCarthy CP, Bruno RM, Brouwers S, Canavan MD, Ceconi C, et al. 2024 ESC Guidelines for the management of elevated blood pressure and hypertension. *Eur Heart J.* 2024 Oct 7;45(38):3912–4018.
4. Heidenreich PA, Bozkurt B, Aguilar D, Allen LA, Byun JJ, Colvin MM, et al. 2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines. *J Am Coll Cardiol.* 2022 May 3;79(17):e263–421.
5. Nagueh SF, Smiseth OA, Appleton CP, Byrd BF 3rd, Dokainish H, Edvardsen T, et al. Recommendations for the Evaluation of Left Ventricular Diastolic Function by Echocardiography: An Update from the American Society of Echocardiography and the European Association of Cardiovascular Imaging. *J Am Soc Echocardiogr.* 2016 Apr;29(4):277–314.
6. Tonino PA, De Bruyne B, Pijls NH, Siebert U, Ikeno F, van' t Veer M, et al. Fractional flow reserve versus angiography for guiding percutaneous coronary intervention. *N Engl J Med.* 2009 Jan 15;360(3):213–24.

7. Koskinas KC, Van Craenenbroeck EM, Antoniadou C, Blüher M, Gorter TM, Hanssen H, et al. Obesity and cardiovascular disease: an ESC clinical consensus statement. *Eur Heart J.* 2024 Oct 7;45(38):4063–98.
8. McCaw ZR, Dey R, Sominen H, Amar D, Mukherjee S, Sandor K, et al. Pitfalls in performing genome-wide association studies on ratio traits. *HGG Adv.* 2025 Apr 10;6(2):100406.
9. Kronmal R. Spurious Correlation and the Fallacy of the Ratio Standard Revisited. *JRSSA* 1993; 156:379–92.
10. Pearson K. Mathematical contributions to the theory of evolution.—On a form of spurious correlation which may arise when indices are used in the measurement of organs. *Proc R Soc Lond* 1897; 60:489–97.
11. Khosla T, Lowe CR. Indices of obesity derived from body weight and height. *Br J Prev Soc Med.* 1967 Jul;21(3):122–8.
12. Loos RJ, Lindgren CM, Li S, Wheeler E, Zhao JH, Prokopenko I, et al. Common variants near MC4R are associated with fat mass, weight and risk of obesity. *Nat Genet.* 2008 Jun;40(6):768–75.
13. Baya NA, Sur Erdem I, Venkatesh SS, Reibe S, Charles PD, Navarro-Guerrero E, et al. Combining evidence from human genetic and functional screens to identify pathways altering obesity and fat distribution. *Am J Hum Genet.* 2025 Oct 2;112(10):2316–37.
14. Yusuf S, Hawken S, Ounpuu S, Dans T, Avezum A, Lanas F, et al. Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study. *Lancet.* 2004 Sep 11-17;364(9438):937–52.
15. Narayan HK, Finkelman B, French B, Plappert T, Hyman D, Smith AM, et al. Detailed Echocardiographic Phenotyping in Breast Cancer Patients: Associations With Ejection Fraction Decline, Recovery, and Heart Failure Symptoms Over 3 Years of Follow-Up. *Circulation.* 2017 Apr 11;135(15):1397–412.
16. Archie JP Jr. Mathematic coupling of data: a common source of error. *Ann Surg.* 1981 Mar;193(3):296–303.
17. Aschard H, Vilhjálmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. *Am J Hum Genet.* 2015 Feb 5;96(2):329–39.