

Detecting and Quantifying Circularity Bias When Composite Quality Ratings Are Used as Covariates: A Methodological Demonstration Using CMS Hospital Excess Readmissions Data

Ali Akram

aakramm989@gmail.com

Independent Researcher

Research Article

Keywords: health services research methodology, over-adjustment bias, descendant of the outcome, composite quality ratings, variance decomposition, hospital readmissions, CMS data

Posted Date: June 30th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-9786241/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at Health Services and Outcomes Research Methodology on October 22nd, 2026. See the published version at <https://doi.org/10.1007/s10742-026-00398-5>.

Abstract

Publicly reported, risk-adjusted quality measures are increasingly used to compare providers and to allocate payment, and analysts frequently include composite quality ratings as covariates when modeling such measures. This paper identifies a methodological hazard in that practice: when a composite rating embeds the outcome of interest as one of its components, including the rating as a regressor means conditioning on a descendant of the outcome, which systematically attenuates the coefficients of other structural predictors. We formalize the problem, derive its expected direction, and propose a simple diagnostic: estimate the structural model with and without the embedding rating and compare the focal coefficient. We demonstrate the approach using the U.S. Centers for Medicare and Medicaid Services (CMS) excess readmission ratio for 2,446 acute care hospitals (fiscal year 2026), with hospital ownership as the focal predictor. Excluding the CMS overall star rating, for-profit ownership is associated with higher excess readmissions than nonprofit ownership (+ 0.0196, 95% CI 0.0126 to 0.0267, $p < 0.001$). Including the star rating, which embeds the readmission outcome, attenuates this coefficient by 48% to + 0.0113 on the same sample. A Monte Carlo simulation in which the unadjusted estimator recovers the known true effect confirms the attenuation (mean 50% across 1,000 replications) is systematic rather than a feature of these data. We also show, via a between- versus within-group variance decomposition, that only 1.1% of the variation in the measure lies between census regions. The diagnostic and the decomposition are transferable to other publicly reported composite quality measures.

1. Introduction

Publicly reported, risk-adjusted quality measures now play a central role in how hospitals and other providers are compared and paid. In the United States, the Centers for Medicare and Medicaid Services (CMS) publishes both individual measures, such as the excess readmission ratio used in the Hospital Readmissions Reduction Program (HRRP), and composite measures, such as the overall hospital quality star rating, which summarizes dozens of indicators into a single value (Zuckerman et al. 2016; Gettel et al. 2024). Researchers analyzing these measures routinely face a modeling choice: when studying how a structural characteristic such as ownership is associated with a specific outcome, should an available composite quality rating be included as a covariate to adjust for overall quality?

This paper argues that the answer is often no, and that doing so introduces a specific, quantifiable bias. The CMS overall star rating embeds readmission performance as one of its components. When such a rating is included as a regressor in a model whose outcome is itself a readmission measure, the rating is not a clean pretreatment control but a descendant of the outcome: it is constructed, in part, from the very quantity being modeled. Including a descendant of the dependent variable among the regressors places a proxy for the outcome on the right-hand side, so the rating coefficient absorbs part of the mechanical outcome-on-outcome relationship and the coefficients of genuine structural predictors are attenuated toward zero. This is a descendant-of-the-outcome (over-adjustment) problem, distinct from confounding and distinct from collider bias; it is well understood in the causal-inference literature but, in

our reading, frequently overlooked in applied analyses of publicly reported quality data, where composite ratings are convenient and readily available.

Unwarranted variation in quality has been a concern since the foundational work of Wennberg and Gittelsohn (1973), and thirty-day readmissions remain a policy-salient marker of it, with roughly one in five Medicare patients readmitted within thirty days (Jencks et al. 2009) at a cost in the tens of billions of dollars (MedPAC 2007). The validity of readmission measures as pure quality signals has been debated (Fischer et al. 2014; Krumholz et al. 2013), and HRRP penalties have been shown to fall unevenly across hospital types (Joynt and Jha 2013; Horwitz et al. 2017). Despite extensive study of patient-level readmission risk using clinical data (Kansagara et al. 2011; Goldstein et al. 2017) and institution-specific analyses (Park et al. 2014), the methodological consequences of using composite ratings as covariates in this setting have received little explicit attention.

We make three contributions. First, we describe the circularity problem and its expected direction. Second, we propose a simple, general diagnostic and demonstrate it on national CMS data, quantifying the bias and confirming it with a Monte Carlo simulation in which the true effect is known. Third, we apply a between- versus within-group variance decomposition to the same measure to show how a complementary descriptive tool clarifies where measured variation actually arises. Although we use hospital ownership and excess readmissions as the running example, the diagnostic and decomposition apply to any setting in which a composite rating embedding the outcome might otherwise be entered as a covariate.

2. Methods

2.1 The circularity problem

Let Y denote a risk-adjusted outcome measure (here, a hospital's excess readmission ratio) and let X denote a structural predictor of interest (here, ownership). Suppose a composite quality rating R is constructed, in part, as a decreasing function of Y , so that $R = g(Y, Q)$ where Q captures other quality domains. The relevant causal structure is simple: X influences Y , and Y influences R , so that R is a descendant of Y (X to Y to R , with Q also feeding R). R is therefore not a pretreatment confounder of the X to Y relationship; it is a partial image of the outcome itself. Regressing Y on X while also including R places a proxy for the dependent variable among the regressors. The coefficient on R then captures part of the mechanical relationship between Y and itself, and the variance it removes is partly outcome variance, so the coefficient on X is attenuated toward zero. The expected direction is attenuation, and its magnitude grows with the weight of Y inside R . This is over-adjustment by conditioning on a descendant of the outcome; it should not be confused with collider bias, which arises from conditioning on a common effect of two variables and opens a backdoor path. The practical implication is that a composite rating embedding the outcome should not be used as a covariate when the goal is to estimate structural associations with that outcome; it may, at most, be reported descriptively.

2.2 A diagnostic

We propose a minimal diagnostic that requires no additional data. On the subsample for which the rating is observed, estimate the structural model twice: once excluding the rating (Model A) and once including it (Model B). Compare the focal coefficient across the two. A material movement toward zero when the rating is added is the flag, but it is not by itself the diagnosis: attenuation is also exactly what legitimate confounding adjustment produces. The distinguishing question is structural, not numerical, namely whether the added variable is a pretreatment confounder or a descendant of the outcome. When it is a descendant, as a composite rating embedding the outcome is, the attenuation is over-adjustment and the adjusted estimate is the biased one; when it is a genuine pretreatment confounder, the attenuation is appropriate and the adjusted estimate is preferred. The movement raises the flag; the structural knowledge makes the diagnosis. Reporting both estimates, the percentage attenuation, and the structural rationale makes the issue transparent and supports the decision to treat the rating descriptively.

2.3 Data

We use two public datasets published by CMS for fiscal year 2026. The HRRP file reports, for each hospital and each of six conditions (acute myocardial infarction, coronary artery bypass graft, chronic obstructive pulmonary disease, heart failure, pneumonia, and elective hip or knee replacement), an excess readmission ratio, the ratio of predicted to expected risk-adjusted thirty-day readmissions, where a value above one indicates worse-than-expected performance. The Hospital General Information file reports ownership, the overall quality star rating, and location. The excess readmission ratio is computed by CMS within dual-eligibility peer groups under the 21st Century Cures Act, so it already partly accounts for patient social risk, although not at the hospital level.

2.4 Sample and variables

For each hospital we computed a discharge-weighted mean excess readmission ratio across the conditions it reported. We merged ownership, star rating, and state, retained acute care hospitals, mapped states to the four census regions, and grouped ownership as nonprofit, for-profit, or government. After restricting to complete cases on the weighted ratio, ownership, region, and discharge volume, the primary analytic sample comprised 2,446 hospitals (Fig. 1 reports the selection flow), of which 2,415 also had a star rating. The outcome is the discharge-weighted excess readmission ratio; predictors are ownership, census region, and the natural logarithm of total discharges; the star rating is used only in Model B and in the diagnostic.

2.5 Statistical analysis

The primary model (Model A) regressed the weighted excess readmission ratio on ownership, region, and log discharge volume using ordinary least squares with heteroskedasticity-consistent (HC3) standard errors, excluding the star rating. Model B added the star rating to the same specification on the rating-complete subsample. The circularity diagnostic compared the for-profit coefficient between Model A and Model B on that common subsample. A Monte Carlo simulation generated data with a known ownership effect and a composite rating constructed to embed the outcome, and re-estimated the effect

with and without the rating to verify the direction and systematic nature of the bias. A between- versus within-region variance decomposition partitioned the total sum of squares in the outcome. Condition-stratified models re-estimated the for-profit association separately for each of the six conditions as a robustness check. Analyses used Python with pandas, numpy, statsmodels, and scikit-learn; the pipeline is reproducible from the public files.

3. Results

3.1 Sample

The primary sample of 2,446 acute care hospitals comprised 1,675 nonprofit, 463 for-profit, and 308 government hospitals, distributed across the South (998), Midwest (581), West (458), and Northeast (409). The mean weighted excess readmission ratio was 1.006 (standard deviation 0.059).

3.2 Primary model and the circularity diagnostic

In the primary model excluding the star rating (Model A, Table 1), for-profit ownership was associated with a higher excess readmission ratio than nonprofit ownership (+ 0.0196, 95% CI 0.0126 to 0.0267, $p < 0.001$). Government ownership was also modestly higher (+ 0.0089, $p = 0.005$). Western and Midwestern hospitals had lower ratios than Northeastern ones (West – 0.0217, $p < 0.001$; Midwest – 0.0157, $p < 0.001$), and larger hospitals had slightly lower ratios (log discharges – 0.0062, $p < 0.001$). The model explained 4.2% of the variance.

Applying the diagnostic on the rating-complete subsample ($n = 2,415$), the for-profit coefficient was + 0.0218 when the rating was excluded and + 0.0113 when it was included, an attenuation of 48% (Fig. 2). The star rating itself entered strongly (-0.0209 per star, $p < 0.001$) and Model B explained 19.8% of the variance, but this apparent improvement is largely mechanical: the rating embeds the outcome, so part of what it explains is the outcome itself. Treating the attenuated + 0.0113 as the structural ownership effect would understate it by nearly half.

Table 1
Primary model of the weighted excess readmission ratio (Model A; n = 2,446; HC3 standard errors).

Predictor	Coefficient	p-value	95% CI
For-profit (vs nonprofit)	+ 0.0196	< 0.001	0.0126 to 0.0267
Government (vs nonprofit)	+ 0.0089	0.005	0.0026 to 0.0152
Midwest (vs Northeast)	-0.0157	< 0.001	-0.0237 to -0.0076
South (vs Northeast)	-0.0102	0.010	-0.0180 to -0.0024
West (vs Northeast)	-0.0217	< 0.001	-0.0309 to -0.0125
Log discharges	-0.0062	< 0.001	-0.0084 to -0.0039

3.3 Simulation

To confirm the attenuation is systematic rather than an artifact of these data, we simulated data under a known data-generating process and repeated it across 1,000 replications of 3,000 observations each. In each replication a binary indicator had a true effect of + 0.020 on the outcome, the outcome depended only on the indicator and independent noise, and a composite rating was constructed to embed the outcome (rating decreasing in the outcome) plus an independent quality component uncorrelated with the indicator. Regressing the outcome on the indicator alone recovered the true effect almost exactly (mean + 0.0200 across replications), confirming that the unadjusted estimator is unbiased under this design. Adding the outcome-embedding rating biased the estimate toward zero (mean + 0.0099), a mean attenuation of 50%. Because the baseline recovers the truth and only the addition of the descendant rating moves it, the simulation isolates the over-adjustment mechanism cleanly, and its 50% attenuation closely matches the 48% observed in the CMS data.

3.4 Variance decomposition

The circularity diagnostic and the variance decomposition are two inexpensive guards against the same broad failure, reading published quality measures too literally: the first against treating an outcome-embedding rating as a clean control, the second against treating a statistically significant grouping coefficient as a substantively important one. Partitioning the total variance in the weighted excess readmission ratio, only 1.1% lay between census regions and 98.9% within them; state-level mean ratios ranged narrowly from 0.92 to 1.04 (Fig. 3). Although regional coefficients are statistically significant, regional membership accounts for almost none of the dispersion in the measure: variation is overwhelmingly a within-region, hospital-level phenomenon. The decomposition thus guards against over-interpreting statistically significant but substantively small geographic differences.

3.5 Robustness across conditions

Estimating the for-profit association separately for each condition (Fig. 1, lower panel), the coefficient was positive in all six and statistically significant in five: acute myocardial infarction (+ 0.0202, $p < 0.001$), chronic obstructive pulmonary disease (+ 0.0114, $p = 0.001$), heart failure (+ 0.0195, $p < 0.001$), pneumonia (+ 0.0238, $p < 0.001$), and hip or knee replacement (+ 0.0663, $p = 0.022$). The coronary artery bypass graft estimate was positive but not significant (+ 0.0129, $p = 0.355$), consistent with its much smaller sample. The hip or knee replacement coefficient (+ 0.0663) was roughly three times the others with a wide interval; whether this reflects genuine heterogeneity in surgical readmissions under for-profit ownership or a small-sample artifact (this condition has the fewest reporting hospitals) cannot be resolved here and is flagged rather than interpreted. The direction of the association is otherwise stable across conditions, while the diagnostic and the decomposition are unaffected by which condition is examined.

4. Discussion

4.1 Principal findings

The methodological message is that composite quality ratings embedding an outcome should not be entered as covariates when the goal is to estimate structural associations with that outcome. Doing so attenuated the focal coefficient by roughly half in national data and by a comparable fraction in a controlled simulation. The simple two-model diagnostic makes the problem visible and is cheap to apply. The complementary variance decomposition showed that statistically significant regional differences nonetheless explain almost none of the dispersion in the measure, a caution against equating significance with substantive importance. Substantively, for-profit ownership was associated with higher excess readmissions, and the direction was stable across conditions, but two cautions are essential. First, the rating-excluded estimate is less biased only in the sense that it is free of the over-adjustment described here; it is not a causal estimate. The excess readmission ratio is risk-adjusted by CMS within dual-eligibility peer groups but not at the hospital level, so the for-profit coefficient remains open to ordinary confounding by patient mix, the standard critique of this comparison. Removing circularity yields a cleaner descriptive association, not a causal effect of ownership. Second, structural characteristics explained only a small share of the total variation. The + 0.0196 coefficient functions in this paper as the focal estimate through which the diagnostic is demonstrated, not as a settled answer to the substantive ownership question.

4.2 Relation to existing methods

The bias described here is an instance of over-adjustment by conditioning on a descendant of the outcome, established in the causal-inference literature, but its specific manifestation with publicly reported composite ratings appears underappreciated in applied health-services work, where such ratings are convenient adjustments. The contribution is less a new estimator than a clearly stated hazard, a transparent diagnostic, and a demonstration of magnitude on data analysts actually use. The

same logic applies to other composite measures that bundle an outcome of interest, including overall star ratings in nursing homes, dialysis facilities, and clinician programs.

4.3 Limitations

The analysis is cross-sectional and the excess readmission ratio, while risk-adjusted by CMS, is an imperfect measure of quality. We also note an irony worth naming directly: the outcome is itself a discharge-weighted composite across six conditions, so a paper about the hazards of composite measures is built on a composite outcome. The weighting is transparent and, unlike the star rating, does not embed any of the structural predictors, so it does not create the circularity at issue; but it carries the usual caveats of composite outcomes and the per-condition models are reported precisely so the composite is not read too literally. The simulation illustrates the mechanism under a stylized data-generating process rather than estimating a structural parameter for hospitals.

4.4 Conclusion

When a composite quality rating embeds the outcome under study, including it as a covariate is not a neutral adjustment but a source of attenuation bias. A simple comparison of the focal coefficient with and without the rating detects the problem, and a variance decomposition clarifies where measured variation truly lies. Demonstrated on national CMS readmissions data, these tools change the estimated for-profit association by nearly half and reframe statistically significant regional differences as substantively minor. Both are transferable to the many settings in which publicly reported composite quality measures are analyzed.

Declarations

Competing Interests:

The author declares no competing interests.

Ethics Approval:

This study used only publicly available, de-identified, hospital-level aggregate data and did not involve human subjects; ethics approval was not required.

Funding:

This research received no external funding.

Author Contribution

A.A. conceived and designed the study, conducted the analysis and simulation, prepared all figures and tables, and wrote and reviewed the manuscript.

Data Availability

All data analyzed in this study are publicly available from the CMS Provider Data Catalog. Two datasets for fiscal year 2026 were used: the Hospital Readmissions Reduction Program file (<https://data.cms.gov/provider-data/dataset/9n3s-kdb3>) and the Hospital General Information file (<https://data.cms.gov/provider-data/dataset/xubh-q36u>).

References

1. Fischer, C., Lingsma, H.F., Marang-van de Mheen, P.J., et al.: Is the readmission rate a valid quality indicator? A review of the evidence. *PLoS One*. **9**, e112282 (2014). <https://doi.org/10.1371/journal.pone.0112282>
2. Gettel, C.J., Bagshaw, K., Qin, L., et al.: Calculation of overall hospital quality star ratings with and without inclusion of the readmission measures. *Health Serv. Res.* **59**, e14256 (2024). <https://doi.org/10.1111/1475-6773.14256>
3. Goldstein, B.A., Navar, A.M., Pencina, M.J., Ioannidis, J.P.A.: Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review. *J. Am. Med. Inf. Assoc.* **24**, 198–208 (2017). <https://doi.org/10.1093/jamia/ocw042>
4. Horwitz, L.I., Bernheim, S.M., Ross, J.S., et al.: Hospital characteristics associated with risk-standardized readmission rates. *Med. Care.* **55**, 528–534 (2017). <https://doi.org/10.1097/MLR.0000000000000713>
5. Jencks, S.F., Williams, M.V., Coleman, E.A.: Rehospitalizations among patients in the Medicare fee-for-service program. *N Engl. J. Med.* **360**, 1418–1428 (2009). <https://doi.org/10.1056/NEJMsa0803563>
6. Joynt, K.E., Jha, A.K.: Characteristics of hospitals receiving penalties under the Hospital Readmissions Reduction Program. *JAMA.* **309**, 342–343 (2013). <https://doi.org/10.1001/jama.2012.94856>
7. Kansagara, D., Englander, H., Salanitro, A., et al.: Risk prediction models for hospital readmission: a systematic review. *JAMA.* **306**, 1688–1698 (2011). <https://doi.org/10.1001/jama.2011.1515>
8. Krumholz, H.M., Lin, Z., Keenan, P.S., et al.: Relationship between hospital readmission and mortality rates for patients hospitalized with acute myocardial infarction, heart failure, or pneumonia. *JAMA.* **309**, 587–593 (2013). <https://doi.org/10.1001/jama.2013.333>
9. Medicare Payment Advisory Commission (MedPAC): Report to the Congress: Promoting Greater Efficiency in Medicare. MedPAC, Washington, DC (2007)

10. Park, L., Andrade, D., Mastey, A., Sun, J., Hicks, L.: Institution specific risk factors for 30 day readmission at a community hospital. BMC Health Serv. Res. **14**, 40 (2014). <https://doi.org/10.1186/1472-6963-14-40>

11. Wennberg, J.E., Gittelsohn, A.: Small area variations in health care delivery. Science. **182**, 1102–1108 (1973). <https://doi.org/10.1126/science.182.4117.1102>

12. Zuckerman, R.B., Sheingold, S.H., Orav, E.J., Ruhter, J., Epstein, A.M.: Readmissions, observation, and the Hospital Readmissions Reduction Program. N Engl. J. Med. **374**, 1543–1551 (2016). <https://doi.org/10.1056/NEJMsa1513024>

Figures

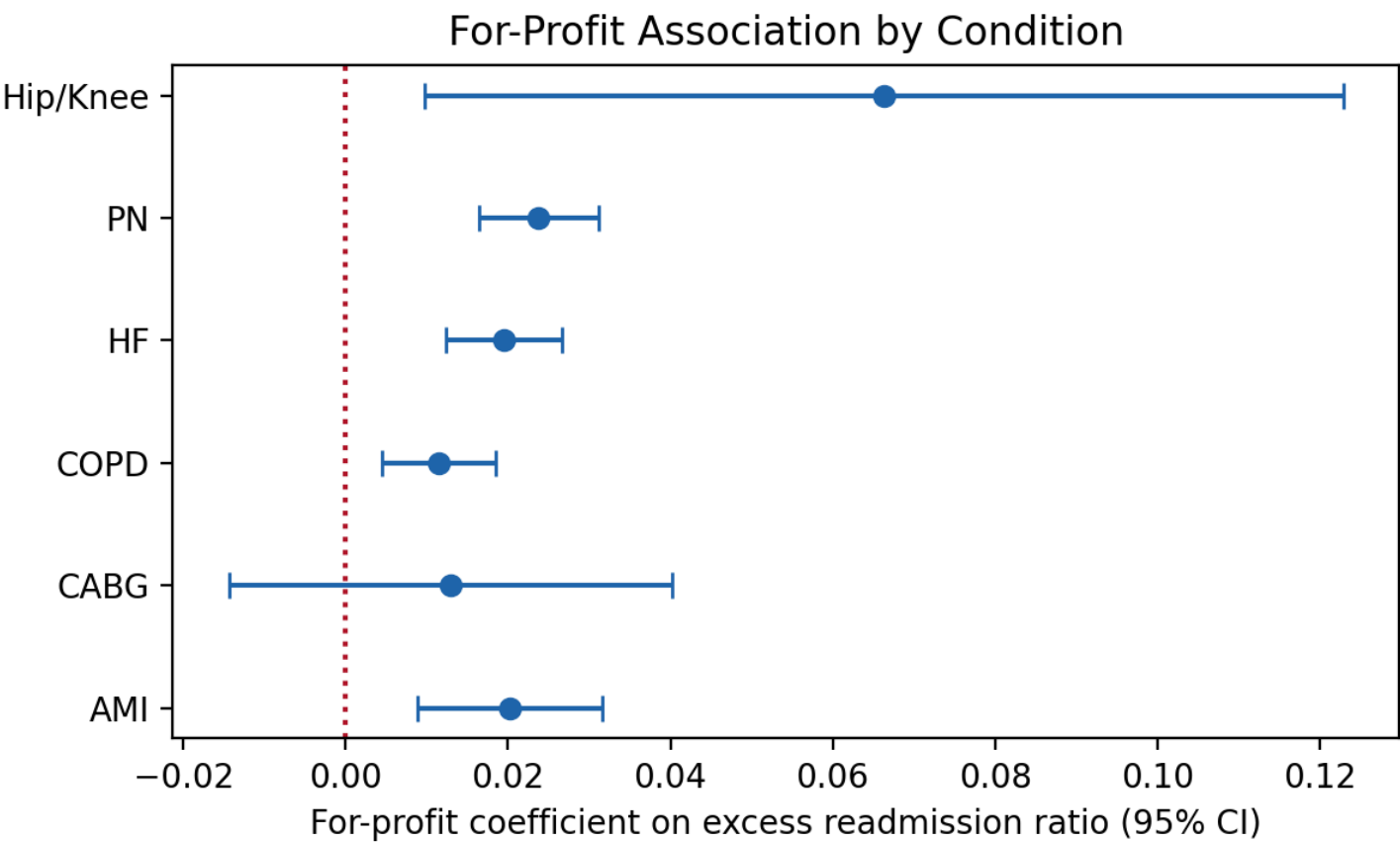


Figure 1

Selection of the analytic sample and, in the lower panel, the for-profit association estimated separately by condition (95% confidence intervals). [Selection flow: 3,055 hospitals in the HRRP file; 2,477 with valid excess readmission ratio and discharges; 2,472 acute care; 2,446 complete cases.]

Attenuation of the For-Profit Coefficient When an Outcome-Embedding Quality Rating Is Included

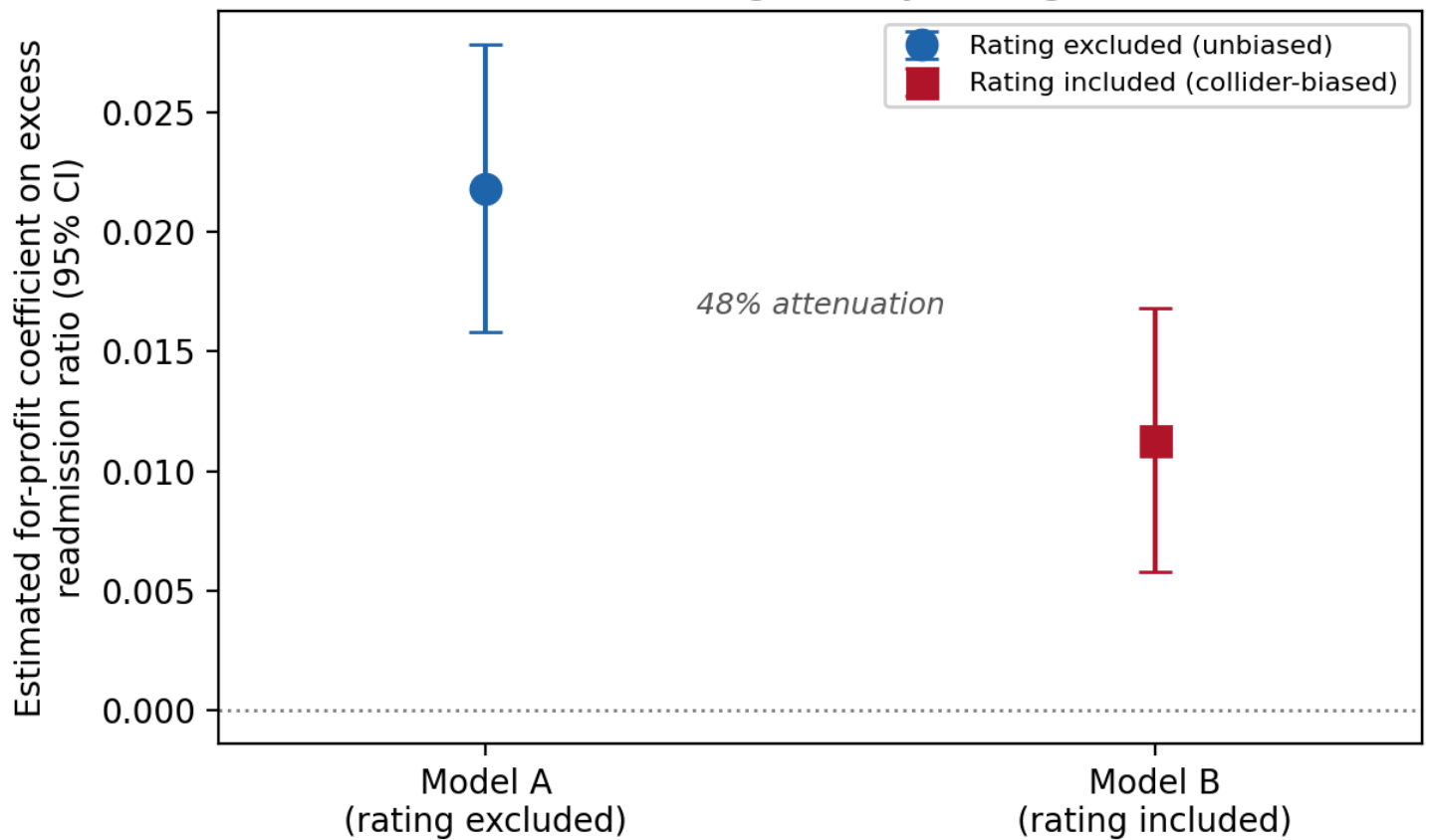


Figure 2

For-profit coefficient on the excess readmission ratio estimated with the outcome-embedding star rating excluded (Model A) versus included (Model B), on the common rating-complete subsample ($n = 2,415$). Including the rating attenuates the coefficient by 48%.

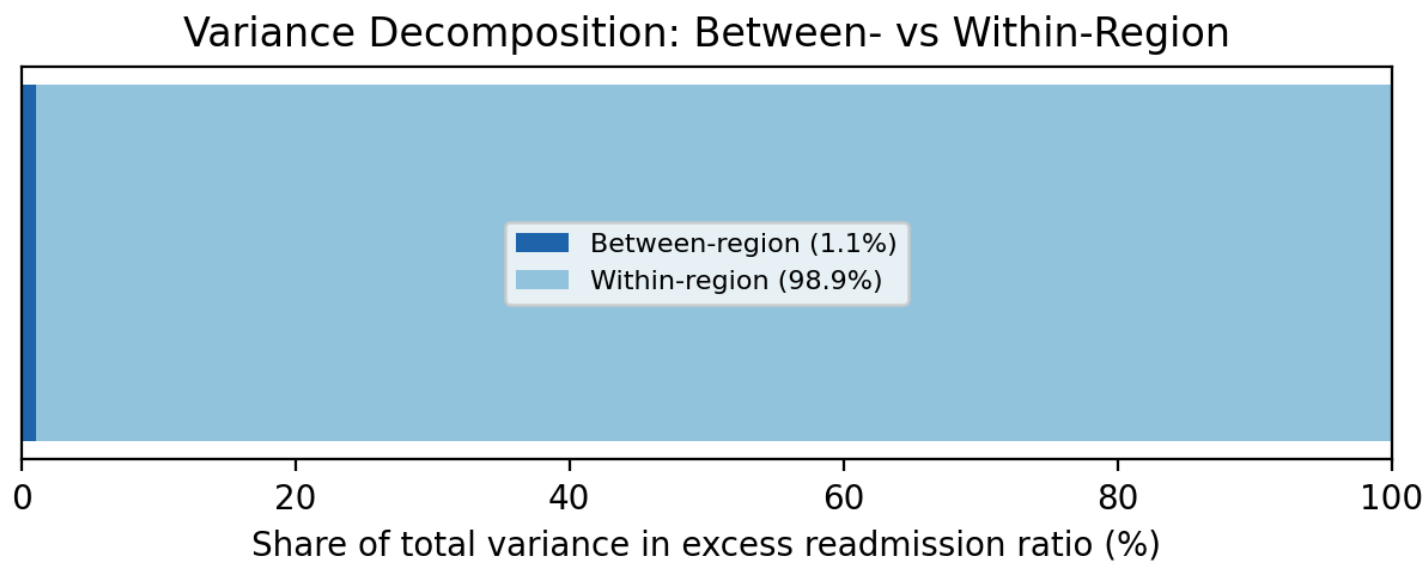


Figure 3

Between- versus within-region variance decomposition of the weighted excess readmission ratio (n = 2,446). Regional membership accounts for 1.1% of total variance.