Measuring Variability Between Clusters By Subgroup: An Extension of the Median Odds Ratio

Christopher Yarnell [a,b,c]

Ruxandra Pinto [d]

Rob Fowler [a,c,d,e,f]

2019-05-19

Affiliations

- [a] Interdepartmental Division of Critical Care Medicine, University of Toronto, Toronto, Canada
- [b] Mount Sinai Hospital, Toronto, Canada
- [c] Institute for Health Management, Policy and Evaluation, University of Toronto, Toronto, Canada
- [d] Sunnybrook Health Sciences Centre, Toronto, Canada
- [e] Department of Medicine, University of Toronto, Toronto, Canada
- [f] Institute for Clinical Evaluative Sciences, Toronto, Canada

Keywords

This is the accepted version of the article
which will be published in full form in Statistics in Medicine
DOI 10.1002/sim.8286

- median odds ratio
- clustered data
- multilevel analysis
- advanced care planning

Abstract

Investigating clustered data requires consideration of the variation across clusters, including consideration of the component of the total individual variance that is at the cluster level. The median odds ratio and analogues are useful intuitive measures available to communicate variability in outcomes across clusters using the variance of random intercepts from a multilevel regression model. However, the median odds ratio cannot describe variability across clusters for different patient subgroups because the random intercepts do not vary by subgroup. To empower investigators interested in equity and other applications of this scenario, we describe an extension of the median odds ratio to multilevel regression models employing both random intercepts and random coefficients. By example we conducted a retrospective cohort analysis of variation in care limitations (goals of care preferences) according to ethnicity in patients admitted to intensive care. Using mixed-effects logistic regression clustered by hospital we demonstrated that patients of non-Caucasian ethnicity were less likely to have care limitations but experienced similar variability across hospitals. Limitations of the extended median odds ratio include the large sample sizes and computational power needed for models with random coefficients. This extension of the median odds ratio to multilevel regression models with random coefficients will provide insight into cluster-level variability for researchers interested in equity and other phenomena where variability by patient subgroup is important.

Motivation and Background

Identifying the effects of clusters on individual outcomes is important for both statistical rigor and practical utility of regression models in epidemiology.[1] Investigating clusters requires assessment of the association between cluster and outcome as well as the assessment of variation in outcome across clusters. Useful and intuitive measures of overall variability exist for outcome variables that are continuous (intraclass correlation coefficient), binary (median odds ratio), time-to-event (median hazard ratio), and rates (median rate ratio).[2,3,4,5,6,7] Yet variability may not be the same for all subgroups of individuals in a population and measures of overall variability across clusters may conceal important information about variability in subgroups of individuals. Modelling variance as a function of individual-level variables has been demonstrated in other settings[8,9,10] but is infrequently used in clinical scenarios, perhaps due to difficulty in interpreting the variance components along

with the fixed and random effects. Here we describe a method of using the median odds ratio to communicate variability across clusters for individuals with different individual- or cluster-level covariates. Using the median odds ratio to communicate this variability may be beneficial due to clinicians' familiarity with odds ratios and the ease of comparison with odds ratios derived from fixed and random coefficients.

In practical terms, comparing the cluster-level variability for individuals with different characteristics is important for investigating questions of equity in healthcare and other settings. Consider the case of goals of care at the end of life for patients of different backgrounds; for example Caucasian as opposed to non-Caucasian ethnicities. Goals of care in hospital generally crystallize to the presence or absence of care limitations. Patients with a greater burden of frailty, age and comorbidity are less likely to recover from critical illness and sometimes opt for care limitations that guide treatment towards palliation and comfort if cardiac arrest or respiratory failure arise. Past research has documented that African-American patients in the United States are less likely to opt for care limitations on the use of life-supporting technology such as a "No cardiopulmonary resuscitation (CPR)" order.[11] Although individual-level differences may explain this finding, an additional potential explanation relates to the practices and culture of individual hospitals in which case patients of non-Caucasian ethnicities might have more variability in prevalence of care limitations than patients of Caucasian ethnicity. Other explanations focus on the impact of socioeconomic phenomena such as mistrust of healthcare institutions, in which case there may be an associated fixed effect of ethnicity but no differences in variability by ethnicity. We investigate this question as an illustrative example.

Full assessment of equity across a subgroup in clustered data requires attention to the average effect as well as the variation associated with that subgroup.[12] Here we propose and demonstrate an extension of the median odds ratio to facilitate intuitive measurement of this variation and empower investigators concerned with equity and other phenomena where variability across subgroups has relevance.

Using the Median Odds Ratio to Investigate Variability by Subgroup

The median odds ratio derived by Larsen et al is a convenient method of measuring variability at the cluster level because it is reported in odds ratio "units", similar to fixed effects coefficients.[2,13] The intuitive interpretation of the median odds ratio is the median increase in odds of an outcome when a patient is switched from a cluster with lower odds to a cluster with higher odds, holding all other variables constant. The median odds ratio is also useful because it describes inter-cluster variability as opposed to the intra-cluster variability described by the intraclass correlation coefficient.

Thus far the median odds ratio has been employed mainly for logistic regression models with a random intercept term which can be used to account for clustered data. Each cluster is assigned a different intercept. Each intercept is an instance of a random variable which by design follows a normal distribution with a single mean and variance. The median odds ratio converts the variance of this distribution into an odds ratio.

If patients from different subgroups are treated differently across clusters, then a random intercept alone may not capture the consequent variability in outcome, despite the outcome variance being intrinsically dependent on the mean. Instead, cluster-level slopes for the variable of interest can be added to the model as seen in a few empirical examples.[9,10,14] Analogous to the random intercepts, these slopes are instances of a random variable with a single mean and variance as well as a covariance with the random intercepts. This allows for the association between a covariate and an outcome to vary across clusters. In our example we use random coefficients to allow the association between care limitations and ethnicity to vary by hospital. However, the median odds ratio has not yet been described for models incorporating random coefficients.

Mixed-Effects Model with Random Intercepts and Random Coefficients

Let patient i within cluster j have covariate vector $x(i,j) = [x_1(i,j), x_2(i,j), ...]$. The elements of $x(i,j)$ could be patient- or cluster-level covariates. Each cluster has an associated random intercept u_j and a random coefficient $\phi_{1,j}$ for variable x_1 where the variables u and ϕ are normally distributed with mean 0 and variance σ^2 and σ^2_ϕ respectively.

Consider a dichotomous response variable $y(i,j)$ that evaluates to 1 or 0. Let $\pi(i,j)$ be the probability that patient i in cluster j has $y(i,j) = 1$. The vector $\beta = [\beta_0, \beta_1, ...]$ contains the fixed coefficients. We can write the model as:

$$\begin{aligned} y(i,j) &\sim \text{Binomial}(1, \pi(i,j)) && (1) \\ \text{logit}[\pi(i,j)] &= (\beta_0 + u_j) + (\beta_1 + \phi_{1,j})x_1(i,j) + ... && (2) \end{aligned}$$

Equation 1 shows that the outcome variable follows a binomial distribution with one trial and success probability $\pi(i,j)$. Equation 2 shows that the logit function is the link function for the model and that each coefficient exists in "logit" space. The terms have been grouped for easier comprehension: the intercept is the sum of the global fixed intercept β_0 and the cluster random intercept u_j; the coefficient of x_1 for each patient is the sum of the global fixed slope β_1 and the cluster random coefficient $\phi_{1,j}$. A random coefficient is included only for x_1 although each covariate in x could have an associated random coefficient.

How to Find the Median Odds Ratio by Patient Subgroup

Using equation 2, if there are no random coefficients then $\phi_{1,j} = 0$ and the formula for the median odds ratio is:

$$\text{MOR} = \exp[\sqrt{2\sigma^2} \times \Phi^{-1}(3/4)] \quad (3)$$

where $\Phi^{-1}(3/4)$ is the 75th percentile of the cumulative standard normal distribution and "exp" is the exponential function.

Recall that the median odds ratio with random intercepts is the median increase in the odds of the outcome when a patient is moved from a lower-odds cluster to a higher-odds cluster. The analogous situation with random coefficients is the median increase in the odds of the outcome when a patient from a particular subgroup is moved from a lower-odds cluster to a higher-odds cluster. Therefore the median odds ratio in the setting of a random coefficient is conditional on the patient subgroup

x_1. This is meaningful only if there is some within-cluster variation in the variable x_1. The following two equations describe the median odds ratio for a model with a random coefficient, conditional on variable x_1 taking value a (see Appendix for derivation):

$$V(x_1 = a) = 2\sigma^2 + 2a^2\sigma_\phi^2 + 4a\text{Cov}(u, \phi) \quad (4)$$

$$\text{MOR}(x_1 = a) = \exp[\sqrt{V(x_1 = a)} \times \Phi^{-1}(3/4)]. \quad (5)$$

The term $V(x_1 = a)$ is a variance term (Appendix). As the variance σ_ϕ^2 of the random coefficient approaches zero, these equations approach the equations for median odds ratio without random coefficients. Note that the variable x_1 could be dichotomous, discrete or continuous and the derived equations will hold.

How to Find the Interval Odds Ratio by Patient Subgroup

The interval odds ratio is used to gauge the magnitude of variance between different values of a covariate. If the interval is wide then the variation between patients with different values of the covariate is large.[13,4] It is an interval covering a proportion of the odds ratios (usually 80%) comparing two patients' odds of the outcome where the two patients differ in only one covariate. Often it is a cluster-level covariate that is used in order to provide some insight into the magnitude of a cluster-level covariate with respect to the cluster-level variability.[13] If the interval odds ratio excludes 1, then the effect of the cluster-level covariate is large compared to the variability between clusters.[4] Consider patient i with covariate vector $x(i, j)$ in cluster j and covariate vector $x(i, k)$ in cluster k. Cluster-level variables differ, that is $x_2(j) \neq x_2(k)$. Recall that $\Phi^{-1}(b)$ gives the b-th quantile of the cumulative standard normal distribution. For a model without random coefficients the interval odds ratio lower and upper bounds have formulas:

$$\text{IOR}_{\text{lower}} = \exp\left[\beta_2(x_2(i) - x_2(j)) + \sqrt{2\sigma^2} \times \Phi^{-1}(0.1)\right] \quad (6)$$

$$\text{IOR}_{\text{upper}} = \exp\left[\beta_2(x_2(i) - x_2(j)) + \sqrt{2\sigma^2} \times \Phi^{-1}(0.9)\right] \quad (7)$$

so that the interval $[\text{IOR}_{\text{lower}}, \text{IOR}_{\text{upper}}]$ covers 80% of the odds ratios when a patient is "switched" from one cluster to another. The above formula could be easily adjusted to accommodate interval odds ratios covering other proportions, such as 75% or 95%.

Within a model using random coefficients the interval odds ratio communicates the spread of odds ratios across a cluster-level covariate conditional on the patient subgroup associated with the random coefficient. We use the same variance $V(x_1 = a)$ from equation 4 (Appendix):

$$\text{IOR}_{\text{lower}} = \exp\left[\beta_2(x_2(i) - x_2(j)) + \sqrt{V(a)} \times \Phi^{-1}(0.1)\right] \quad (8)$$

$$\text{IOR}_{\text{upper}} = \exp\left[\beta_2(x_2(i) - x_2(j)) + \sqrt{V(a)} \times \Phi^{-1}(0.9)\right]. \quad (9)$$

Again the interval odds ratio is given by $[\text{IOR}_{\text{lower}}, \text{IOR}_{\text{upper}}]$.

Association Between Care Limitations and Patient Ethnicity

To illustrate the use of the median odds ratio for describing variation in outcome across clusters for patients from different subgroups, we investigate variation in care limitations across hospitals for patients with Caucasian and non-Caucasian ethnicities. Previous research in the United States has demonstrated that patients from ethnic minorities (Hispanic or non-Caucasian ethnicities) and in particular patients of African American ethnicity are less likely to have limitations on care such as a "Do Not Atttempt Resucitation" (DNAR) order.[15,16] Other work describes many potential contributing factors including communication approaches by healthcare providers, lack of trust in the medical system by patients and families, poor health literacy, different family decision-making processes, and access limitations based on geographic or socieoeconomic barriers.[11,17,18,19,20,21,22]

Some of the observed differences in end-of-life care to patients from ethnic minorities might be explained by hospital-level variation in care.[23,24,25] For example, some hospitals may have clinicians or staff with subconscious biases, while other hospitals may have mitigated those biases through training programs or increasing diversity through hiring practices.[26,27] This could introduce hospital-level variation in the extent to which an outcome, such as care limitations, vary by ethnicity.

Here we used a free restricted-access anonymized database of intensive care unit admissions in the United States[28] to demonstrate extension of the median odds ratio to models including random coefficients. We explored if there is more or less variability in care limitations across hospitals for patients of Hispanic or non-Caucasian ethnicities compared to patients of Caucasian ethnicity.

Study Population

The study population included anonymized patient-level data from the e-ICU Collaborative Research Database.[28,29] This database is maintained by the Phillips eICU Research Institute[30] and released under the Health Insurance Portability and Accountability Act (HIPAA) safe harbor provision. The re-identification risk was certified as meeting safe harbor standards by Privacert (Cambridge, MA) (HIPAA Certification no. 1031219-2). Each patient was admitted to an intensive care unit in the United States in 2014 and 2015 and for each patient only the first admission to an intensive care unit was analyzed.

Variables

The data is abstracted from the electronic medical record and includes patient demographics (age, gender, ethnicity), severity of illness (Acute Physiology And Chronic Health Evaluation (APACHE) IV severity scores), hospital of admission, and care limitations. The care limitation data are short phrases that can be broadly grouped into care with no limitations and care with limitations:

- No limitations on care: "Full Therapy": corresponding to a plan to provide all life-supporting technology including CPR.
- Limitations on care: "No augmentation of care", "Do not resuscitate", "No CPR", "Comfort measures only", "No intubation", "No cardioversion", "No vasopressors/inotropes": corresponding to a plan where some forms of life-supporting technology are not provided.

Reasons for care limitations might be related to patient or substitute decision maker preferences and recommendations by the physician and healthcare team. Data around motivations for limiting care is not available in the e-ICU database.

Analysis and Models

The analysis used logistic regression to model the relationship between ethnicity and care limitations for patients clustered within hospitals. In each model the response variable is the presence or absence of a care limitation. Age, gender, severity of illness (APACHE IV) and ethnicity are incorporated as covariates in all models.[31] Age was modeled as restricted cubic splines with four knots placed at the 5th, 35th, 65th and 95th percentiles.[32] Some models consider hospital ethnic diversity (hospital proportion of patients with non-Caucasian ethnicity above or below the median across hospitals). The response variable was the presence or absence of care limitations. Hospitals with less than 40 admissions (approximately the smallest 5%) and patients with missing baseline or outcome data were excluded. Analyses were performed using R v3.5 (R foundation) and several R packages including the `lme4` R package.[33,34] Full code is available in the Appendix.[35,36,37,38,39,40,41,42,43,44,45,46]

We proceeded sequentially from the base model to a model including random intercepts, random coefficients, and a cluster-level covariate:

- Model 1: Fixed effects (age, gender, severity of illness, ethnicity) - does not account for hospital-level clustering.
- Model 2: Model 1 with hospital-level random intercept - accounts for hospital-level clustering but the effect of ethnicity is fixed as in Model 1. As a consequence, the variation between hospitals is assumed to be the same for Caucasian and non-Caucasian patients.
- Model 3: Model 2 with hospital-level ethnicity random coefficient - accounts for hospital-level clustering and the effect of ethnicity has both fixed and random components. As a consequence, the variation between hospitals can be distinguished according to patient ethnicity.
- Model 4: Model 3 with cluster-level fixed effect for hospital ethnic diversity - introduces a cluster-level covariate to allow calculation of an interval odds ratio.

Models were compared with Akaike Information Criteria and modified likelihood ratio tests.[47,48,49]

Results

A total of 107,274 patients admitted to an intensive care unit in 174 hospitals were analyzed, of whom 23,635 (22%) were of non-Caucasian ethnicity and 49,098 (46%) were female (Table 1). The median age was 65 years (IQR 53-76), the median APACHE score was 50 (IQR 36-68) and the

ICU mortality rate was 6.4%. Median number of patients from analyzed hospitals was 1,306 (IQR 724-2,715). Care limitations were present in 16,485 (15%) patients including 13,931 (17%) patients with Caucasian ethnicity and 2,554 (11%) patients with non-Caucasian ethnicity.

Four models were constructed to describe factors associated with the presence of care limitations. Each successive model made modest improvement in describing the data based on the Akaike Information Criterion and likelihood-ratio test. Across all four models the odds of having a care limitation increased with APACHE IV score and female gender. Figure 1 shows that the odds of having a care limitation generally increase with age. The odds ratio estimates were similar between models with random effects, with the 95% confidence interval of the Model 4 odds ratio associated with ethnicity 0.78 (95% CI 0.72-0.83) encompassing the estimates for Models 2 and 3. The estimate for the fixed effect odds ratio of ethnicity in the unclustered model (Model 1) was more extreme than any of the subsequent estimates, indicating that some of the differences between care limitation prevalence by patient ethnicity are explained by clustering at the hospital level.

Hospital-level variation in care limitations as measured by the median odds ratio was 1.51 for Model 2, slightly greater in magnitude than the odds ratios associated with female relative to male gender or Caucasian relative to non-Caucasian ethnicity and similar to an increase in APACHE IV score by 15 points holding all other covariates constant. This suggests hospital-level variation is a component of variation in care limitation prevalence with magnitude comparable to individual patient-level factors such as gender. However, the median odds ratio does not communicate the ratio of latent cluster-level to individual-level variation, which in this case is only about 5% based on the variance partitioning coefficient using Model 2.[4]

Models 3 and 4 allow calculation of the median odds ratio for Caucasian patients (1.48) and non-Caucasian patients (1.44). Clinically, these median odds ratios are indistinguishable, implying that the variation in care limitation prevalence between hospitals is similar for patients of different ethnicities (Figure 2).

The interval odds ratio for care limitation comparing patients from hospitals with lower versus higher ethnic diversity was broad. For Caucasian patients the interval odds ratio is 0.66 to 2.93 and for non-Caucasian patients the interval odds ratio is 0.7 to 2.77. The proportional change in variance after adding the neighbourhood-level covariate is 14.3%. The proportion of opposed odds ratios (POOR)[1] is 34%, meaning that approximately 34% of pairs of hospitals with different levels of

ethnic diversity (high versus low) have odds ratios in the opposite direction indicating heterogeneity with respect to the association between hospital ethnic diversity and goals of care limitation. Taken together, this shows that the hospital ethnic diversity measured in this manner may not explain hospital-level variation in prevalence of care limitations.

Discussion

Cluster-level variability is an important aspect of modeling multilevel data in healthcare and other fields. The median odds ratio is one helpful tool for communicating and interpreting the extent of inter-cluster variability. In this paper the median odds ratio and interval odds ratio were extended from the case of random intercepts to the case of random intercepts and random coefficients to empower statistical investigations of variability by subgroup.

Investigation of hospital-level variation in care limitations according to patient ethnicity was used to illustrate the extended median odds ratio. The median odds ratio by ethnicity revealed that hospital-level variation is similar for patients of Caucasian and non-Caucasian ethnicities. The addition of a cluster-level ethnicity variable improved model fit but the interval odds ratio showed residual hospital-level variability. These findings imply that differences in care limitation preferences are driven by factors consistent across hospitals, as opposed to factors that vary by hospital. For example, these findings do not support the theory that differences in care limitation prevalence by ethnicity relate to racist behavior on the part of few outlier hospitals. Instead they support the theory that differences relate to factors common to all hospitals such as mistrust of medical institutions or ongoing unintentional systemic racism.[50,51] The findings are limited by the proportion of excluded patients due to missing severity of illness information, the specific setting of intensive care unit admission, and the possibility of residual confounding from relevant variables not included in the model (such as diagnosis) or the database (such as functional status).

The extension of the median odds ratio to the setting of random coefficient models has limitations. The most important limitation is that interpretation of random coefficient and intercept models is subtle and can be misleading. The median odds ratio itself is not enough to fully understand the proportion of individual-level variation existing at the cluster level, and other quantities such as the variance partitioning coefficient or the area under the receiver operating characteristics curve

may also be helpful.[52] However, our proposed median odds ratio extension may aid in appropriate interpretation of random coefficients in multilevel models. A second limitation is that models with random coefficients may not be possible or necessary for many datasets due to sample size or the balance of individual-level as opposed to cluster-level variability. A final limitation is that models with random coefficients are computationally intensive and require more time and processing power than models with only fixed effects.

Conclusion

The median odds ratio can be extended to models with random coefficients to assess variability across clusters by patient subgroup. In our example we demonstrated that non-Caucasian patients have lower odds of care limitation after adjusting for age, gender and severity; however there was no clinically significant difference in variability across hospitals for patients of Caucasian as opposed to non-Caucasian ethnicities. This argues in favour of broad societal factors affecting care limitation prevalence by ethnicity as opposed to variation in individual hospital culture and practices.

Other outcomes regarding ethnicity and care limitation prevalence were possible. For example, if we had found that the median odds ratio for non-Caucasian patients was much larger than the median odds ratio for Caucasian patients we would have described inequity that is not captured by the fixed effect coefficients. This could prompt more detailed consideration of individual hospital characteristics and how those might be related to variation in treatment of non-Caucasian patients.

Since the introduction of the median odds ratio, several different median measures such as the median hazard ratio have emerged which allow description of variability across clusters. Each of these measures relies on the assumption that variability is uniform across subgroups. In certain scenarios, such as questions of health equity, this assumption may require investigation. Our extension of the median odds ratio provides researchers a way to verify this assumption and investigate variability across clusters by patient subgroup.

Funding

Physician Services Incorporated Resident Research Grant, Heart and Stroke Foundation of Ontario.

Acknowledgements

We thank the following for helpful comments: Donald A Redelmeier, Peter Austin.

References

1. Merlo J, Wagner P, Ghith N, Leckie G. An original stepwise multilevel logistic regression analysis of discriminatory accuracy: The case of neighbourhoods and health Moerbeek M, ed. *PLoS ONE*. 2016;11(4):e0153778. Available at: http://dx.plos.org/10.1371/journal.pone.0153778.

2. Larsen K, Petersen JH, Budtz-Jørgensen E, Endahl L. Interpreting Parameters in the Logistic Regression Model with Random Effects. *Biometrics*. 2000;56(3):909–914. Available at: http://doi.wiley.com/10.1111/j.0006-341X.2000.00909.x.

3. Goldstein H, Browne W, Rasbash J. Partitioning Variation in Multilevel Models. *Understanding Statistics*. 2002;1(4):223–231. Available at: http://www.tandfonline.com/doi/abs/10.1207/S15328031US0104{_}02

4. Austin PC, Merlo J. Intermediate and advanced topics in multilevel logistic regression analysis. *Statistics in Medicine*. 2017;36(20):3257–3277. Available at: http://doi.wiley.com/10.1002/sim.7336.

5. Austin PC, Wagner P, Merlo J. The median hazard ratio: a useful measure of variance and general contextual effects in multilevel survival analysis. *Statistics in Medicine*. 2017;36(6):928–938. Available at: http://www.ncbi.nlm.nih.gov/pubmed/27885709 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PI

6. Austin PC, Stryhn H, Leckie G, Merlo J. Measures of clustering and heterogeneity in multilevel Poisson regression analyses of rates/count data. *Statistics in Medicine*. 2018;37(4):572–589. Available at: http://www.ncbi.nlm.nih.gov/pubmed/29114926 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5 http://doi.wiley.com/10.1002/sim.7532.

7. Merlo J, Wagner P, Austin PC, Subramanian SV, Leckie G. General and specific contextual effects in multilevel regression analyses and their paradoxical relationship: A conceptual tutorial. *SSM - Population Health*. 2018;5:33–37. Available at: https://www.sciencedirect.com/science/article/pii/S235282731830048X.

8. Goldstein H. Heteroscedasticity and Complex Variation. *Encyclopedia of Statistics in Behavioral*

Science. 2005;2:790–795. Available at: http://www.bristol.ac.uk/media-library/sites/cmm/migrated/documents/complex-variation.pdf.

9. Ohlsson H, Merlo J. Understanding the effects of a decentralized budget on physicians' compliance with guidelines for statin prescription – a multilevel methodological approach. *BMC Health Services Research*. 2007;7(1):68. Available at: http://www.ncbi.nlm.nih.gov/pubmed/17488496 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC1885428 https://bmchealthservres.biomedcentral.c 6963-7-68.

10. Merlo J, Asplund K, Lynch J, Rastam L, Dobson A. Population Effects on Individual Systolic Blood Pressure: A Multilevel Analysis of the World Health Organization MONICA Project. *American Journal of Epidemiology*. 2004;159(12):1168–1179. Available at: https://academic.oup.com/aje/article-lookup/doi/10.1093/aje/kwh160.

11. Barnato AE, Chang C-CH, Saynina O, Garber AM. Influence of Race on Inpatient Treatment Intensity at the End of Life. *Journal of General Internal Medicine*. 2007;22(3):338–345. Available at: http://link.springer.com/10.1007/s11606-006-0088-x.

12. Merlo J, Ohlsson H, Lynch KF, Chaix B, Subramanian SV. Individual and collective bodies: Using measures of variance and association in contextual epidemiology. *Journal of Epidemiology and Community Health*. 2009;63(12):1043–1048. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19666637 http://jech.bmj.com/cgi/doi/10.1136/jech.2009.088310.

13. Larsen K, Merlo J. Appropriate assessment of neighborhood effects on individual health: Integrating random and fixed effects in multilevel logistic regression. *American Journal of Epidemiology*. 2005;161(1):81–88. Available at: https://academic.oup.com/aje/article-lookup/doi/10.1093/aje/kwi017.

14. Merlo J, Yang M, Chaix B, Lynch J, Råstam L. A brief conceptual tutorial on multilevel analysis in social epidemiology: Investigating contextual phenomena in different groups of people. *Journal of Epidemiology and Community Health*. 2005;59(9):729–736. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16100308 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC

15. Kwak J, Haley WE. Current research findings on end-of-life decision making among racially or ethnically diverse groups. *Gerontologist*. 2005;45(5):634–641. Available at: https://academic.oup.

com/gerontologist/article-lookup/doi/10.1093/geront/45.5.634.

16. Erickson SE, Vasilevskis EE, Kuzniewicz MW, Cason BA, Lane RK, Dean ML, Rennie DJ, Dudley RA. The effect of race and ethnicity on outcomes among patients in the intensive care unit: A comprehensive study involving socioeconomic status and resuscitation preferences. *Critical Care Medicine*. 2011;39(3):429–435. Available at: http://search.ebscohost.com/login.aspx?direct=true{\&}db=c8h{\&}AN=104852792{\&}site=ehost-live.

17. Mack JW, Paulk ME, Viswanath K, Prigerson HG. Racial disparities in the outcomes of communication on medical care received near death. *Archives of Internal Medicine*. 2010;170(17):1533–1540. Available at: http://www.ncbi.nlm.nih.gov/pubmed/20876403 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3739279 http://archinte.jamanetwork.com/article.

18. Burt J, Plant H, Omar R, Raine R. Equity of use of specialist palliative care by age: Cross-sectional study of lung cancer patients. *Palliative Medicine*. 2010;24(6):641–650. Available at: http://search.ebscohost.com/login.aspx?direct=true{\&}db=c8h{\&}AN=105008926{\&}site=ehost-live.

19. Carrion IV. Communicating terminal diagnoses to Hispanic patients. *Palliative and Supportive Care*. 2010;8(2):117–123. Available at: http://ovidsp.ovid.com/ovidweb.cgi?T=JS{\&}CSC=Y{\&}NEWS=N{\&}

20. Mazanec PM, Daly BJ, Townsend A. Hospice utilization and end-of-life care decision making of African Americans. *The American journal of hospice & palliative care*. 2010;27(8):560–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21071435.

21. Song MK, Donovan HS, Piraino BM, Choi J, Bernardini J, Verosky D, Ward SE. Effects of an intervention to improve communication about end-of-life care among African Americans with chronic kidney disease. *Applied Nursing Research*. 2010;23(2):65–72. Available at: http://ovidsp.ovid.com/ovidweb.cgi?T=JS{\&}CSC=Y{\&}NEWS=N{\&}PAGE=fulltext{\&}D=med5{\&}AN=

22. Barnato AE, Mohan D, Downs J, Bryce CL, Angus DC, Arnold RM. A randomized trial of the effect of patient race on physicians' intensive care unit and life-sustaining treatment decisions for an acutely unstable elder with end-stage cancer. *Critical Care Medicine*. 2011;39(7):1663–1669. Avail-

able at: http://ovidsp.ovid.com/ovidweb.cgi?T=JS{\&}CSC=Y{\&}NEWS=N{\&}PAGE=fulltext{\&}D=medc{\

23. Barnato AE, Berhane Z, Weissfeld LA, Chang CCH, Linde-Zwirble WT, Angus DC. Racial variation in end-of-life intensive care use: A race or hospital effect? *Health Services Research*. 2006;41(6):2219–2237. Available at: http://doi.wiley.com/10.1111/j.1475-6773.2006.00598.x.

24. Barnato AE, Chang C-CH, Lave JR, Angus DC. The Paradox of End-of-Life Hospital Treatment Intensity among Black Patients: A Retrospective Cohort Study. *Journal of Palliative Medicine*. 2017;21(1):jpm.2016.0557. Available at: http://online.liebertpub.com/doi/10.1089/jpm.2016.0557.

25. Lackan NA, Eschbach K, Stimpson JP, Freeman JL, Goodwin JS. Ethnic differences in in-hospital place of death among older adults in California: Effects of individual and contextual characteristics and medical resource supply. 2009;47:138–145. Available at: http://ovidsp.ovid.com/ovidweb.cgi?T=JS{\&}CSC=Y{\&}NEWS=N{\&}PAGE=fulltext{\&}D=med5{\&}AN=

26. Paradies Y, Truong M, Priest N. A systematic review of the extent and measurement of healthcare provider racism. *Journal of General Internal Medicine*. 2014;29(2):364–387. Available at: http://www.ncbi.nlm.nih.gov/pubmed/24002624 http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3 http://link.springer.com/10.1007/s11606-013-2583-1.

27. Williams DR, Wyatt R. Racial bias in health care and health: Challenges and opportunities. *JAMA - Journal of the American Medical Association*. 2015;314(6):555–556. Available at: http://jama.jamanetwork.com/article.aspx?doi=10.1001/jama.2015.9260.

28. Goldberger AL, Amaral LA, Glass L, Hausdorff JM, Ivanov PC, Mark RG, Mietus JE, Moody GB, Peng CK, Stanley HE. A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, PhysioNet: Com- ponents of a new research resource for complex physiologic signals," *Circulation*. 2000;101(23):E215–20. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10851218.

29. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. *Scientific Data*. 2018;5:180178. Available at: http://www.nature.com/articles/sdata2018178.

30. Phillips eICU Research Initiative. eICU Collaborative Research Database v 2.0 [database]. 2017.

Available at: https://eicu-crd.mit.edu/about/eicu/.

31. Zimmerman JE, Kramer AA, McNair DS, Malila FM. Acute Physiology and Chronic Health Evaluation (APACHE) IV: Hospital mortality assessment for today's critically ill patients. *Critical Care Medicine*. 2006;34(5):1297–1310. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16540951 https://insights.ovid.com/crossref?an=00003246-200605000-00001.

32. Harrell FEJ. *Regression Modeling Strategies - With Applications to Linear Models, Logistic Regressions, and Survival Analysis*. Springer New York; 2001:251–264.

33. R Development Core Team. *R: a Language and Environment for Statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2011. Available at: https://www.r-project.org/.

34. Bates D, Mächler M, Bolker B, Walker S. Fitting Linear Mixed-Effects Models using lme4. *Journal of Statistical Software*. 2014;67(1):1–48. Available at: http://arxiv.org/abs/1406.5823.

35. Wickham H, François R, Henry L, Müller K. *dplyr: A grammar of data manipulation.*; 2018. Available at: https://cran.r-project.org/package=dplyr.

36. Wickham H. *tidyr: Easily tidy data with 'spread()' and 'gather()' functions.*; 2016. Available at: http://cran.r-project.org/package=tidyr.

37. Wickham H, Hester J, Francois R. *readr: Read Rectangular Text Data.*; 2017. Available at: https://cran.r-project.org/package=readr.

38. Knowles JE, Frederick C, Maintainer]. *Package 'merTools' Title Tools for Analyzing Mixed Effect Regression Models.*; 2016. Available at: https://www.github.com/jknowles/merTools.

39. Gómez-Rubio V. *ggplot2 - Elegant Graphics for Data Analysis (2nd Edition)*. Springer-Verlag New York; 2017:3–5. Available at: http://www.jstatsoft.org/v77/b02/.

40. Dowle M, Srinivasan A. *data.table: Extension of 'data.frame'.*; 2018. Available at: https://cran.r-project.org/package=data.table.

41. Xie Y. *Dynamic Documents with R and knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC; 2013:188. Available at: http://www.crcpress.com/product/isbn/9781482203530.

42. Xie Y. *knitr: A General-Purpose Tool for Dynamic Report Generation in R.*; 2012:1–12. Available

at: http://cloud.github.com/downloads/yihui/knitr/knitr-manual.pdf{\%}5Cnpapers3://publication/uuid/5088673
8AFE-4961-8CDD-186D699DFB3A.

43. Xie Y. knitr: A Comprehensive Tool for Reproducible Research in R. in: Stodden V, Leisch F, Peng RD, eds. *Implementing reproducible research*. Chapman; Hall/CRC; 2014:3–32. Available at: http://www.crcpress.com/product/isbn/9781466561595.

44. Silge J, Robinson D. tidytext: Text Mining and Analysis Using Tidy Data Principles in R. *The Journal of Open Source Software*. 2016;1(3):37. Available at: http://joss.theoj.org/papers/10.21105/joss.00037.

45. Yoshida K, Bohn J. *Package 'tableone'. Create 'Table 1' to Describe Baseline Characteristics.*; 2018:1–33. Available at: https://cran.r-project.org/package=tableone.

46. Zhu H. Construct complex table with 'kable' and Pipe syntax. 2018:24.

47. Akaike H. A New Look at the Statistical Model Identification. *IEEE Transactions on Automatic Control*. 1974;19(6):716–723. Available at: http://ieeexplore.ieee.org/document/1100705/.

48. Rothman KJ. *Epidemiology : an introduction*. Oxford University Press; 2012:268. Available at: https://global.oup.com/academic/product/epidemiology-9780199754557?cc=ca{\&}lang=en{\&}.

49. Austin PC, Leckie G. The effect of number of clusters and cluster size on statistical power and Type I error rates when testing random effects variance components in multilevel linear and logistic regression models. *Journal of Statistical Computation and Simulation*. 2018;88(16):3151–3163. Available at: https://www.tandfonline.com/doi/full/10.1080/00949655.2018.1504945.

50. Oluo I. *So you want to talk about race*. Da Capo Press; 2018:1–238.

51. Caplan A. *Bad blood: The Tuskegee syphilis experiment*. Free Press; 2007:275–276. Available at: http://link.springer.com/10.1017/S1745855207225529.

52. Wagner P, Merlo J. Discriminatory Accuracy of a Random effect in Multilevel Logistic Regression. *International Journal of Epidemiology*. 2015;44(suppl_1):i49–i50. Available at: https://academic.oup.com/ije/article/2573789/Discriminatory.

Table 1: Baseline Patient Information

	Summary [1]
Total Patients	
Included	107,274 (77%)
Missing information [2]	32,093 (23%)
Gender	
Female	64,040 (46%)
Male	75,177 (54%)
Unknown / Missing	42 (0%)
Age (years)	65 [53-76]
Ethnicity	
Caucasian	107,669 (77%)
Non-Caucasian	29,916 (21%)
Missing	1,782 (1%)
APACHE IV	
Score	50 [36-67]
Missing	27,338 (20%)
Hospital characteristics	
Patients per hospital	1,306 [724-2,715]
Included hospitals	174 (84%)
Proportion non-Caucasian (percent)	18 [9-30]
ICU mortality	
Death before ICU discharge	8,708 (6%)
Missing	15 (0%)
Goals of care information	
Present	130,374 (94%)
Missing	8,993 (6%)

[1] Each row summarized as mean (%) or median [IQR].
[2] Comprising all patterns of missingness among covariates and outcome.

Table 2: Summary of Models Predicting Care Limitations

	Model 1	Model 2	Model 3	Model 4
Odds ratio (95% CI)				
Non-Caucasian ethnicity	0.66 (0.63-0.7)	0.78 (0.74-0.82)	0.76 (0.71-0.81)	0.78 (0.72-0.83)
APACHE IV score [1]	1.36 (1.35-1.37)	1.37 (1.36-1.38)	1.37 (1.36-1.38)	1.37 (1.36-1.38)
Male gender	0.82 (0.79-0.86)	0.83 (0.8-0.86)	0.83 (0.8-0.86)	0.83 (0.8-0.86)
Lower hospital diversity [2]	–	–	–	1.39 (1.21-1.59)
Median odds ratio				
Overall	–	1.51	–	–
Non-Caucasian ethnicity	–	–	1.53	1.48
Caucasian ethnicity	–	–	1.48	1.44
Model fit				
AIC [3]	73,041	72,007	72,002	71,983
Log of likelihood ratio	–	517.9	4.32	10.61
Variance parameters				
Intercept variance	–	0.1867	0.1986	0.1702
Coefficient variance	–	–	0.0333	0.0327
Int.-coeff. covariance	–	–	-0.0318	-0.0288

[1] Odds ratios correspond to a 10-point increase in APACHE IV score

[2] Denotes if a given hospital proportion of non-Caucasian patients is less than the median

[3] Akaike Information Criterion

[4] Model 4: Caucasian decedent variance is 0.1702 and non-Caucasian decedent variance is 0.2029

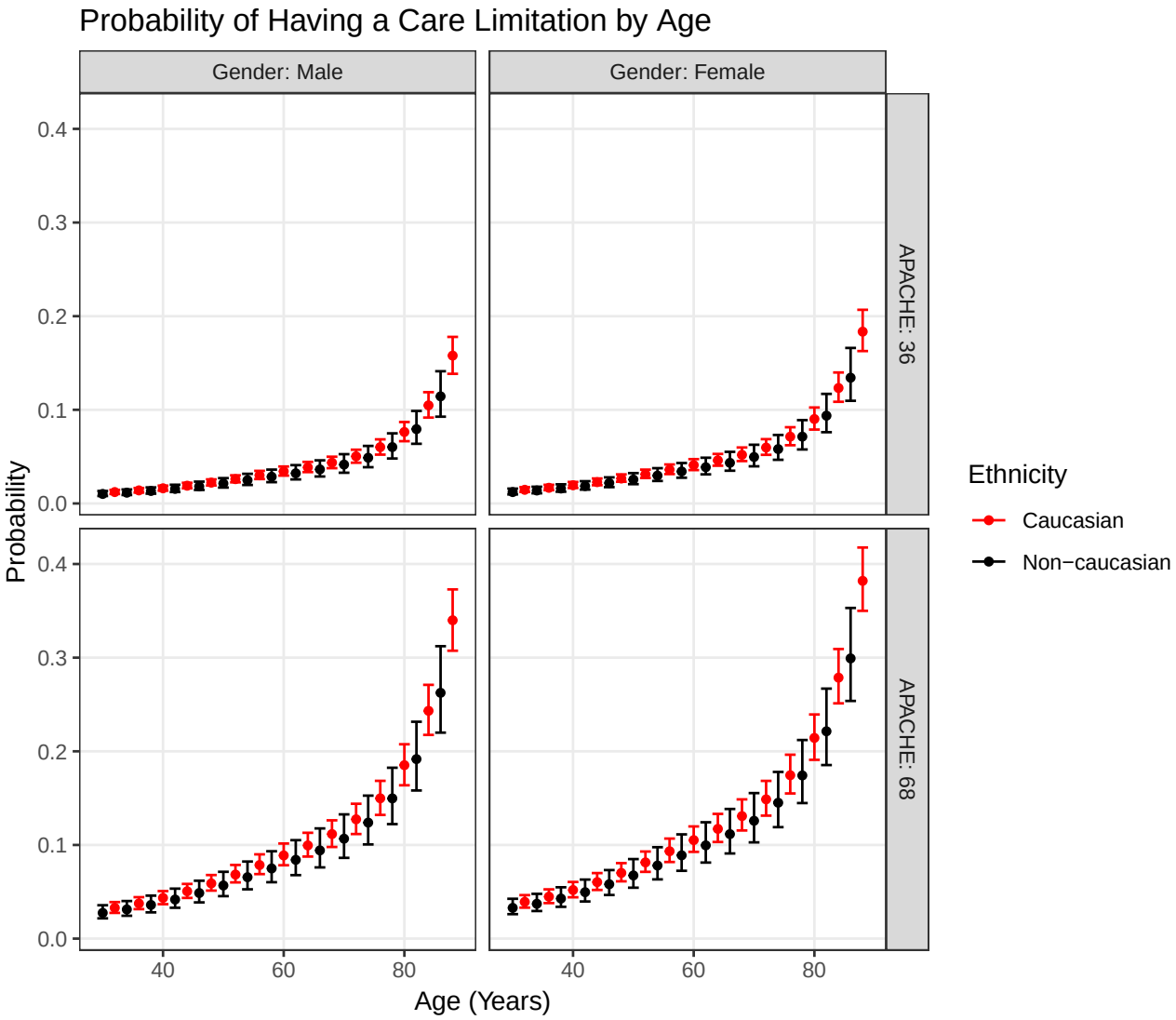

Figure 1: This figure shows the predicted probability of having a care limitation on the y axis and patient age on the x axis calculated using Model 3. Ethnicities are separated by colour. Hospital was set to the modal hospital. Curves with APACHE scores at the 25th (score 36) and 75th (score 68) percentile are shown. The dot denotes the estimated probability and the errorbars surrounding denote an estimate of the 95% prediction interval. The intervals were generated using the merTools package in R (see Appendix for more details).

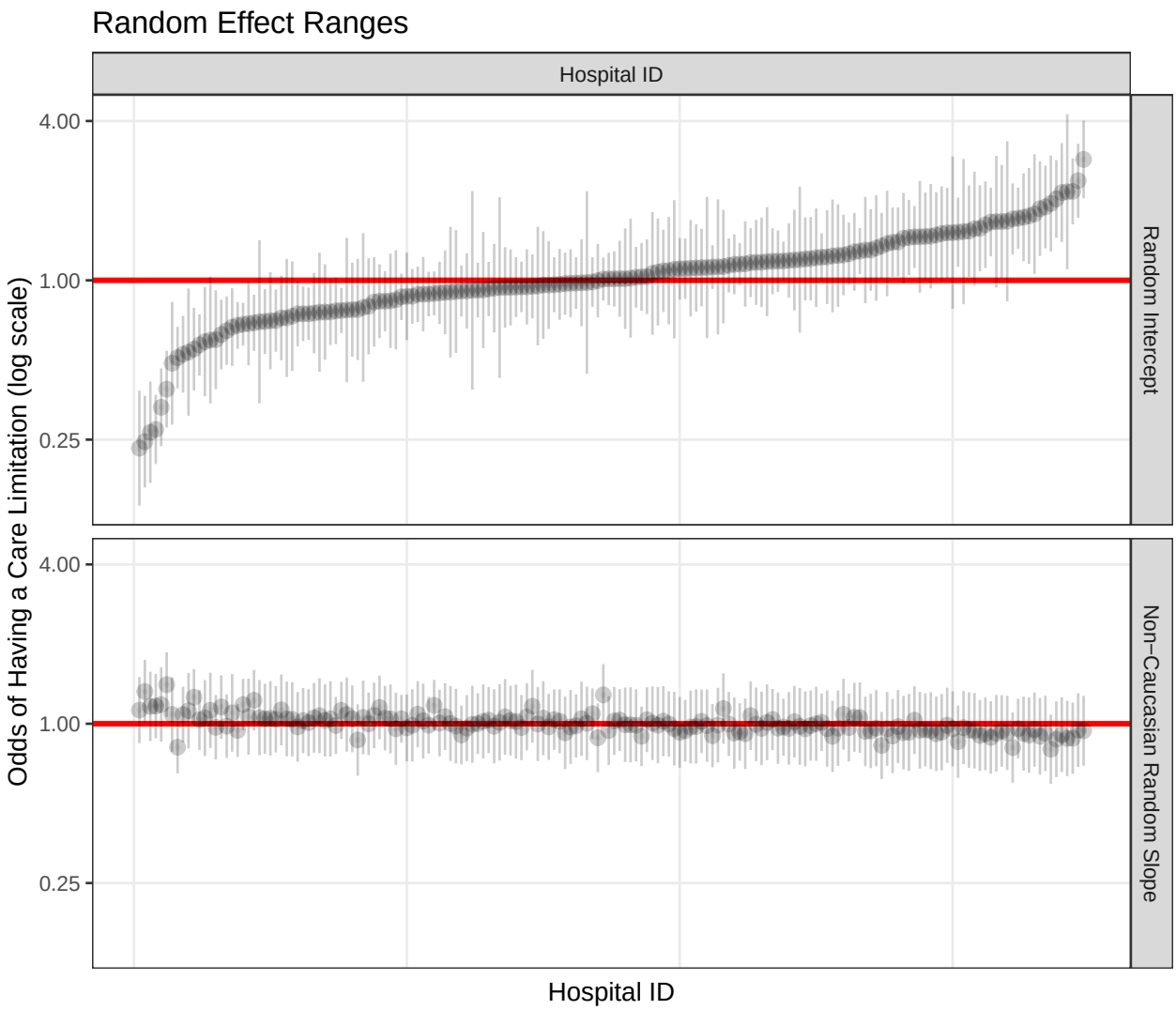

Figure 2: This figure shows the random intercepts and slopes estimated for each hospital based on Model 3. The y axis shows the odds ratio plotted in logarithm scale. Each point has an associated vertical line representing the 95% confidence interval of the estimate for that point. The upper panel shows the distribution of hospital intercepts. The lower panel shows the distribution of random slopes associated with non-Caucasian ethnicity. This figure demonstrates that the hospital random intercepts vary more than the hospital random slopes suggesting that variability in care does not depend on ethnicity.