



Published in final edited form as:

Cancer Epidemiol Biomarkers Prev. 2014 July ; 23(7): 1346–1355. doi:10.1158/1055-9965.EPI-13-1130.

Physicians, Clinics, and Neighborhoods: Multiple Levels of Influence on Colorectal Cancer Screening

Sandi L. Pruitt^{1,2,*}, Tammy Leonard³, Song Zhang¹, Mario Schootman⁴, Ethan A. Halm^{1,5}, and Samir Gupta^{6,7}

¹Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, TX U.S.A

²Harold C. Simmons Comprehensive Cancer Center, Dallas, TX U.S.A

³School of Economic, Political, and Policy Sciences, University of Texas-Dallas; Dallas, TX U.S.A

⁴Department of Epidemiology, College for Public Health and Social Justice, Saint Louis University; St. Louis, Missouri, U.S.A

⁵Departments of General Internal Medicine, University of Texas Southwestern Medical Center; Dallas, TX U.S.A

⁶Division of Gastroenterology, Department of Internal Medicine, Moores Cancer Center, University of California San Diego, San Diego, California, U.S.A

⁷Department of Veterans Affairs, San Diego Healthcare System, San Diego, California, U.S.A

Abstract

Background—We 1) Described variability in colorectal cancer (CRC) test use across multiple levels, including physician, clinic, and neighborhood; and 2) Compared the performance of novel cross-classified vs. traditional hierarchical models.

Methods—We examined multilevel variation in CRC test use among patients not up-to-date with screening in a large, urban safety net health system (2011–2012). Outcomes included: 1) fecal occult blood test (FOBT) or 2) colonoscopy and were ascertained using claims data during a 1-year follow-up. We compared Bayesian 1) Cross-classified 4-level logistic models nesting patients within separate, non-overlapping “levels” (physicians, clinics, and census tracts) vs. 2) Three hierarchical 2-level models using deviance information criterion. Models adjusted for covariates (patient sociodemographic factors, driving time to clinic, and census tract poverty rate).

Results—Of 3,195 patients, 157 (4.9%) completed FOBT and 292 (9.1%) completed colonoscopy during the study year. Patients attended 19 clinics, saw 177 physicians, and resided in 332 census tracts. Significant variability was observed across all levels in both hierarchical and

*Address for correspondence: University of Texas Southwestern Medical Center, Department of Clinical Sciences, Division of Outcomes and Health Services Research, 5323 Harry Hines Blvd E1.410D Dallas, TX 75390-9169, Phone: 214-648-3257, Fax: 214-648-3232, Sandi.Pruitt@utsouthwestern.edu.

CONFLICTS:

Dr. Gupta has received research support in the form of donated fecal immunochemical tests from the Polymedco Corporation. No potential conflicts of interest were disclosed by the other authors.

cross-classified models that was unexplained by measured covariates. For colonoscopy, variance was similar across all levels. For FOBT, physicians, followed by clinics, demonstrated the largest variability. Model fit using cross-classified models was superior or similar to 2-level hierarchical models.

Conclusions—Significant and substantial variability was observed across neighborhood, physician, and clinic levels in CRC test use, suggesting the importance of factors at each of these levels on CRC testing.

Impact—Future multilevel research and intervention should consider the simultaneous influences of multiple levels, including clinic, physician, and neighborhood.

INTRODUCTION

While U.S. guidelines recommended screening for healthy asymptomatic adults beginning at age 50, screening uptake is suboptimal. In 2010, about two-thirds (65.4%) of eligible adults in the U.S. met screening guidelines.(1) Colorectal cancer (CRC) screening behavior requires interaction with the health care system (physicians, clinics) and the larger environment in which that system exists (health systems, families, neighborhoods, state and national health policy).(2) Acknowledging these interactions, cancer prevention researchers are increasingly adopting multilevel frameworks to better understand and improve screening behavior and outcomes. Multilevel frameworks explicitly conceptualize health and health behaviors as a product of the dynamic inter-relation of multiple levels of influence, including the individual, social, structural, and spatial.(3) Multilevel models are a tool used to analyze hierarchically structured data(4)—that is, data organized across the *levels* in which humans are aggregated, (i.e. nested within) such as nations, neighborhoods, organizations, teams, families, and so forth.(3) Multilevel models contain variables measured at different levels of these hierarchies and statistically account for this hierarchical nesting.(4) *Multilevel* models should be distinguished from *multivariable* models which entail the inclusion of multiple independent or dependent variables without accounting for hierarchical nesting.

This growing body of literature has identified variation in CRC screening across multiple geographic and institutional levels of influence. For example, geographic variations in screening have been observed across different census tracts, zip codes, counties, and states. (5–8) Screening rates also differ widely by physicians.(9) Evidence also suggests organizational-level variations in screening, such as those occurring across primary care practices and clinics.(10, 11) The National Cancer Institute (NCI) has called for multilevel interventions(12) designed to improve cancer care and outcomes. However, it is not well understood how these different levels—both geographic and institutional—are related. For example, the presence of clinic-level variation may result in spurious neighborhood variation; or the two may arise from independent causal processes.

While multilevel conceptual frameworks acknowledge numerous levels of influence,(3) traditional multilevel analyses of CRC screening typically include two, or at most, three, strictly-hierarchical levels—an oversimplification of the true complexity present in the CRC screening continuum. For example, Figures 1A–1C depict hierarchical data structures

assumed in traditional multilevel models: patients are assumed to be nested in non-overlapping census tracts (Figure 1A), or assigned to single physicians (Figure 1B) or clinics (Figure 1C). Traditional multilevel models do not reflect the inherent complexity of the CRC screening continuum(2, 13) nor the complex health systems and environments experienced by patients, which are not necessarily hierarchical. A more realistic scenario is depicted in Figure 1D, wherein patients are simultaneously *cross-classified* across multiple overlapping, non-hierarchical levels. For example, patients from the same neighborhood may attend different clinics and physicians in some healthcare systems practice in more than one clinic.

Neglecting data cross-classification can result in mis-specified statistical models and potentially spurious conclusions.(14, 15) For example, model mis-specification can bias estimates of fixed effects, standard errors, and variance components. Thus, observed variations across different levels may be an artifact of other, unmeasured levels of influence or may be reduced because underlying similarities (e.g. common physicians) are not included in the analysis. For example, ignoring neighborhood variation may result in artificially inflated differences across physicians, clinics, or health systems. Importantly, identifying the key factor(s) responsible for outcome variation within a multilevel context may allow for the development of more effective screening interventions.

Multilevel intervention research is in its infancy and there is a clear gap in the literature between multilevel descriptive studies and intervention research.(16) To date, while multilevel factors associated with CRC screening have been identified in descriptive research, these results haven't yet been leveraged during the development, implementation, or evaluation of multilevel interventions. Thus, to inform future multilevel interventions, our goals were to provide insight into the relative impact of multiple levels of influence on CRC test use and to compare two methods for assessing this variation. We examine descriptive data of a real-world scenario to identify the key leverage points contributing to screening variation at multiple relevant levels. Specifically, we examined the extent of variation in uptake of fecal occult blood tests (FOBT) and colonoscopy among patients in a usual care setting by cross-classifying patients within separate "levels" of physicians, clinics, and neighborhoods using two multilevel modeling techniques. To our knowledge, we are the first study to simultaneously examine CRC test use uptake across multiple, overlapping, non-hierarchical levels.

MATERIALS AND METHODS

Sample

We conducted a secondary data analyses of patients (n=3,898) randomized to the "usual care" control arm of a randomized, pragmatic, comparative effectiveness trial conducted in John Peter Smith Health System (JPS). There was no intervention in the usual care study arm. Patients received only opportunistic, visit-based offers to complete home- or office-based guaiac fecal occult blood test (FOBT), or facility-based tests including colonoscopy, barium enema, or sigmoidoscopy at the discretion of physicians. Physicians could prescribe any test modality and all patients had equal insurance coverage of either modality. Because the parent study was a pragmatic trial, usual care participants did not know they were

participating in a research study and thus were not influenced by selection/volunteer bias. (17) Further, usual care patients were not required to have any doctor visits during the follow up period. Thus, our study represents a *real-world* analysis of the role of physicians, clinics, and neighborhoods on CRC test use.

JPS is an urban publicly-funded safety-net healthcare system consisting of community- and hospital-based primary-care clinics and a tertiary-care hospital providing services to residents of Fort Worth and Tarrant County, Texas. The sample and parent study are described in detail elsewhere.(17) Briefly, the trial included patients ages 54–64 years, with a recent health system visit (any visit within 8 months before randomization), no CRC history, and who were uninsured but enrolled in a county-wide medical assistance program for the uninsured. Patients were excluded if they were up-to-date with CRC screening (defined as having a FOBT within 1 year, sigmoidoscopy or barium enema within 5 years, or colonoscopy within 8 years [8 not 10 years was used given availability of health system data.])

For the current study, we excluded 74 patients with addresses that could not be geocoded (e.g. P.O. Box) or who resided outside Tarrant County. We also excluded patients without an assigned primary care clinic (n=120) or physician (n=509). Institutional Review Boards at JPS and UT Southwestern Medical Center approved the study.

Clustering

Patients were clustered within 3 separate levels by: primary care physicians, clinics, and residential census tracts. Physicians, clinics, and patient addresses at baseline were ascertained using administrative and claims data. Addresses were geocoded to residential census tracts using ArcMap (ArcGIS, Version 9.3.1, ESRI Inc., Redlands, CA).

Measures

We measured two outcomes because we hypothesized that the influence of different “levels” (e.g. clinic, physician) varied by test modality. Outcomes included completion of either a 1) simple non-invasive stool blood test (FOBT) or 2) more complicated, facility-based tests (colonoscopy, barium enema, or sigmoidoscopy) within one year of randomization. Patients were considered to have FOBT if FOBT was their first test, regardless of whether FOBT was followed by colonoscopy. Because the predominant facility-based test (98%) was colonoscopy, we hereafter refer to facility-based tests as colonoscopy. Outcomes and covariates were ascertained using claims data. Because indication for testing could not be discerned from claims data, we refer to completion of tests as “CRC test use” following standard practice.(18, 19)

We measured covariates related to CRC test use in previous research:(20, 21) age, sex, race/ethnicity (non-Hispanic white, non-Hispanic black, Hispanic, and other), and primary language spoken at home (English, Spanish). We measured minutes of driving time from patient home to 1) primary-care clinic and 2) the central hospital, where the endoscopy facility is located because increased travel time may be associated with lower CRC test use. (22) Driving time was calculated with MapQuest’s Open Document API Web Service which

uses OpenStreetMap (OSM) data. OSM is a collaborative open-source participatory GIS map updated weekly with data provided by various, registered contributors, including international government agencies.(23, 24) Data requests for driving time calculations did not control for traffic or time of day because that functionality was not available at the time of our study. We included both measures in the models, as they represent access to different aspects of medical care, including CRC screening, primary and tertiary care; measures were moderately correlated ($r=.39$). Census tract poverty rate was measured using a 5-year estimate (2006–2010) drawn from the American Community Survey.

Analysis

We fitted 3 different Bayesian hierarchical 2-level random effects logistic models, with patients nested within 1) primary care physicians (Figure 1A); 2) primary care clinics (Figure 1B); or 3) census tracts (Figure 1C). Next, we fitted one Bayesian, cross-classified random effects logistic model(25) allowing patients to be cross-classified within multiple separate, non-hierarchical grouping “levels” (Figure 1D). All levels (physicians, clinics, and census tracts) were included in the cross-classified model.

For each outcome, we fit “empty” and multivariable models. Empty models include no predictor variables, but included a clustered structure, and are fit for the purpose of quantifying variation at various levels. We included a clustered structure and all covariates in multivariable models in order to identify remaining variation independent of potential confounders that may cluster spatially or across levels.

The full cross-classified model includes regression on the covariates as well as all random effects:

$$\log \left(\frac{p_i}{1-p_i} \right) = x_i' \beta + e_{j(i)} + \theta_{k(i)} + \delta_{l(i)} \text{ for } i=1, \dots, I. \quad (1)$$

Where Y_i generically denotes the binary outcome of the i th ($i = 1, \dots, I$) subject (1 for Yes, 0 for No). We use $x_i = x_{i1}, \dots, x_{ip}$, a vector of length p , to denote the vector of covariates from subject i . The vector of regression coefficients is defined as $\beta = (\beta_1, \dots, \beta_p)'$. Note that we set $x_{i1} = 1$ for $i = 1, \dots, I$, which implies that β_1 is the intercept. Suppose that subject i is nested in the j th ($j = 1, \dots, J$) physician; the k th ($k = 1, \dots, K$) clinic, and the l th ($l = 1, \dots, L$) census tract. We define e_j , θ_k , and δ_l to be the random effect for physician j , clinic k , and census tract l , respectively. All analyses are based on logistic regression models. We define $p_i = P(Y_i = 1)$. Here subscript $j(i)$, $k(i)$, and $l(i)$ index the physician, the clinic, and the census tract that subject i is nested in, respectively. For more detailed information for each model, including our specification of priors, see Supplementary Material and Methods.

We quantified variation at each level in all models to provide insight into the relative impact of multiple levels of influence on CRC test use. We reported median odds ratios (MOR) to facilitate interpretation of the variance across levels on a scale similar to odds ratios associated with other model variables.(26) The MOR is based on the model median random effects variance component (V): $MOR = \exp(0.95 \sqrt{V})$. It is interpreted as the median value

of the ratio of predicted odds of the outcome for two patients randomly selected from different “levels” with equivalent covariates. The MOR ranges from 1 to infinity; if the MOR equals 1, it indicates no variation in outcome across levels. We obtained standard errors of variances to compute 95% credible interval (equivalent to “confidence intervals” per frequentist statistics) for MORs using Markov Chain Monte Carlo methods.

To evaluate methods for assessing variation, we compared traditional 2-level and novel cross-classified models using Deviance Information Criterion (DIC), a Bayesian measure of model fit (analogous to AIC in frequentist statistics). DIC assesses how well the model fits the data with a penalty on model complexity. Lower values indicate better fit. A rule of thumb states that a difference of 3 to 7 in DIC indicates material difference between models. (27) We used WinBUGS (version 1.4.3) to analyze the data. After 10,000 burn-in iterations, 10,000 additional iterations were kept for parameter estimates.

RESULTS

Our sample included 3,195 patients not up-to-date with screening at baseline. Patient characteristics by test use are provided in Table 1. Overall test use was low in the one-year follow-up period; only 8.1% (n= 299), received colonoscopy or other facility-based tests and only 4.3% (n=158) received FOBT.

The cross-classified data structure and the number of unique patients across the data structure are depicted in Figure 2. Patients attended 19 different clinics, saw 177 unique physicians, and resided in 332 different census tracts. The median number of patients per level was as follows: 8 per tract (range: 1–52), 6 per physician (range: 1–112), and 133 per clinic (range: 1–509). Overall, 3,195 patients were distributed across 2,288 unique combinations of physicians, clinics, and census tracts.

Multilevel Variation in FOBT

Empty 2-level hierarchical and cross-classified models demonstrated statistically significant variability across all levels in FOBT testing (Table 2). Variability remained in adjusted models, suggesting that none of the covariates explained variability in either test modality. Across all 2-level models, physicians demonstrated the largest variability followed by clinics, and lastly, neighborhoods. In 2-level hierarchical models, the adjusted model MORs were 2.31 (95% CI: 1.83–3.03) for physicians, 1.77 (1.46–2.46) for clinics, and 1.56 (1.37–1.85) for neighborhoods. In other words, if a patient switched from a physician with a low FOBT rate to a high FOBT rate, her odds of receiving FOBT would be 2.31 times higher (in median). MOR point estimates were similar in cross-classified models as in separate 2-level models (Table 2, Figure 3).

Multilevel Variation in Colonoscopy

Both empty and adjusted models demonstrated statistically significant variability across all levels in colonoscopy testing (Table 2). Variances were similar across all levels: adjusted model MORs from traditional hierarchical 2-level models were 1.60 (1.39–2.06) for clinics, 1.57 (1.38–1.83) for physicians, and 1.53 (1.36–1.76) for neighborhoods. Variation was similar in 2-level hierarchical and cross-classified models in both empty (data not shown)

and adjusted models (Figure 3). In other words, if a patient switched from a clinic, physician, or neighborhood with a low colonoscopy rate to a high colonoscopy rate, her odds of receiving colonoscopy would be 1.60, 1.57, and 1.53 times higher, respectively, in median.

Comparison of Cross-Classified Models to Traditional 2-level Models

We compared DIC of novel cross-classified models to traditional 2-level models separately for empty and adjusted models (Table 3). Of all FOBT models, cross-classified models and 2-level physician models demonstrated equivalent fit (DIC difference within 7) and were the best fitting models, compared to 2-level clinic and neighborhood models. This suggests that within our multilevel context, physicians had the most influence on variation in FOBT test use. For colonoscopy, cross-classified, 2-level physician and 2-level clinic models demonstrated equivalent fit (DIC difference within 7) and were the best fitting models, compared to the 2-level neighborhood models. This suggests that within our multilevel context, physicians and clinics introduced roughly equivalent variation in colonoscopy use, and their impact on variation was greater than that introduced by neighborhood factors.

DISCUSSION

We demonstrated significant variability across multiple “levels” in CRC testing in our large urban safety net health system. Variability across physicians, clinics, and neighborhoods was substantial and was not explained by the inclusion of measured correlates. Notably, even after accounting for physician and clinic variance in cross-classified models, neighborhood variance remained significant. These findings confirm the extensive literature documenting neighborhood disparities in cancer screening(5, 6, 8, 28) and add to the smaller but growing literature on physician- and practice-based differences.(9–11)

The first goal of our study was to provide insight into the relative impact of multiple levels of influence on CRC testing to inform future multilevel intervention research. Using two multilevel model specifications, we determined that of all levels, variability in FOBT was greatest across physicians. For colonoscopy, variability was significant and similar across all levels. Future multilevel research and intervention should consider leveraging all levels, with particular attention to physician variability in FOBT testing. For example, interventions conducted at the physician-level or at higher-levels that do not require physician action (e.g. system-level mailed FOBT kits) may reduce physician variability.

Our second goal was to compare two methods to assess multilevel variation. The question of when to consider a cross-classified model is complex and somewhat uncertain. In the current study, our conclusions were similar when comparing cross-classified vs. multiple traditional 2-level hierarchical models. Model fit using novel cross-classified models was superior or similar to that of several of the traditional 2-level hierarchical models. Further, MORs did not vary much between the different multilevel data structures. It is possible that the bias resulting from model mis-specification would be larger given a higher degree of data cross-classification. For example, in our data, similarly to many systems, cross-classification between physicians and clinics was relatively minor. Different health systems may have different degrees of cross-classification. It is likely that cross-classification across some

levels is more common; for example, clinics serving patients from multiple neighborhoods is a likely scenario that could lead to more significant mis-specification. To our knowledge, we are the first to apply cross-classified models in the cancer screening literature and to compare them to hierarchical models and there is currently little evidence regarding the empirical consequences of mis-specifying cross-classified random effects models.(14, 29) For these reasons, we follow earlier recommendations(14) that suggest if data are theoretically cross-classified, researchers should consider cross-classified models in order to avoid theoretical model mis-specification even in scenarios in which data have low levels of cross-classification. More empirical and simulation research is needed to better understand the appropriate role of cross-classified models in cancer prevention research.

We observed larger variability in FOBT as compared to colonoscopy across all levels. Similarly, a Missouri study demonstrated larger area-level variation in FOBT testing as compared to endoscopy.(6) Confounding by indication may partially explain this finding. We would expect similar symptom prevalence across levels and that more colonoscopies in usual care were done for diagnostic reasons (i.e. symptoms). Thus, if more FOBTs were ordered strictly for screening (e.g. not ordered for symptoms), the more “discretionary” nature of screening could explain the greater multilevel variation in FOBT.

Mechanisms

The mechanisms underlying multilevel variation are incompletely understood. It is not known *why* CRC test use varies across physicians, clinics, and neighborhoods and exactly *how* physicians, clinics, and neighborhoods influence individual test use behaviors. Identifying modifiable mechanisms contributing to multilevel variation will be a crucial next step in the design and delivery of multilevel interventions.

Factors such as organizational structure and strategies may be important mechanisms.(2, 30) A number of studies have identified the implementation of organizational changes as among the most effective strategies to increase cancer screening uptake. These include, for example, practice-level reminder systems and the promotion of continuous patient care designed to make cancer screening services a part of routine patient care.(30–33) However, despite strong evidence supporting their effectiveness, such strategies are not adopted universally.(11) For example, 30% of physicians reported that their practices employed provider reminders and only 15% employed patient reminders in 2006–2007.(9) A more recent survey documented wide variation in the performance of many of the discrete CRC screening steps (e.g. reminders, rescheduling no-shows,) across 15 primary care practices. (10)

It is possible that physician “champions” or “detractors” that are strongly for or against FOBT are among the mechanisms driving the high observed physician-level variation. That is, some may be stronger proponents of FOBT than others. Provider recommendation has consistently been shown to be one of most influential factors predicting cancer screening. (34, 35) In a 2006–2007 national survey of primary care providers, 17% reported that they only recommend a single test modality, with colonoscopy the most recommended test. Just 1% recommended FOBT only.(9) This is worrisome, given growing evidence identifying large differences in screening participation depending on which test is offered.(17, 36) The

potential for physician preferences and/or practice styles to disproportionately influence FOBT (vs. colonoscopy) uptake should be examined in future research.

There may be multiple potential neighborhood-level mechanisms. For example, access to health care, local area socioeconomic deprivation, transportation availability, or neighborhood social norms about screening could lead to neighborhood-level variation. Furthermore, neighborhood and physician/clinic level factors may interact to drive variability; such variability can only be examined with appropriate cross-classified methods.

Strengths and Limitations

Our study faces several limitations. Our sample is drawn from a single, urban safety-net health system where all patients had equal access to very low cost health care. Further, other systems may experience different degrees of cross-classification (e.g. physicians may not practice in multiple clinics). Thus, results may not be generalizable to other health systems. We face a “missing cells” problem where not all possible combinations of levels are represented in the data, thus some statistical inferences are “off-support”(37) or not based on actual observations. We were unable to measure physician- or clinic- level covariates such as attitudes, knowledge, provider type, procedure volume, organizational culture, or institutional systems such as reminders or best practice alerts that may have accounted for observed variation. Lastly, following common practice, we defined neighborhoods using census tracts, a relatively arbitrary choice. It is possible more precise and/or granular definitions of neighborhoods could have better captured neighborhood-level variation.

Despite these limitations, our study has several strengths. To our knowledge, no prior studies have simultaneously examined variation in CRC test use across multiple, overlapping, non-hierarchical levels. In doing so, we provide early empirical evidence supporting the simultaneous consideration of variation across multiple, non-hierarchical levels in cancer prevention and control interventions. We also provide several conceptual and analytic implications for cancer prevention and control research. It is already widely acknowledged that failure to account for the hierarchical nature of data using multilevel analyses can bias standard errors and alter statistical inference.(38) Traditional 2-level models limit analyses to strictly hierarchical relationships. By neglecting to consider other more complex cross-classifications, they may provide biased or misleading conclusions.(14, 29) For example, neighborhood differences identified in 2-level models may be partly a result of clustering of patients across different physicians, clinics, or health systems. Ultimately, failure to examine cross-classification prevents identification of important sources of variation and relevant factors at multiple, overlapping, and not necessarily hierarchical levels.

Failure to identify the key leverage points contributing to screening variation at all relevant levels in future interventions will compromise the public health goal of increased and equitable cancer screening uptake. Multilevel analyses, including cross-classified analyses, will be a useful tool for identifying multilevel variation, monitoring intervention effectiveness, and identifying which factors at which levels to target in interventions. A robust multilevel research agenda designed to fill this gap will require continued investment in large-scale, high-quality multilevel data. Health system data, such as claims or electronic

medical record data, will facilitate a clearer understanding of multiple nesting structures, including physician, clinic, and residential neighborhood.

Summary

We demonstrated the importance of assessing cross-classification and multilevel variation in CRC screening. Using two different multilevel modeling techniques, we found that multiple levels of influence exert influence on CRC testing. Our conclusions were similar when using cross-classified vs. multiple traditional 2-level hierarchical models. We suggest consideration of cross-classified models in situations where multiple levels of influence are of primary interest to researchers and data are theoretically cross-classified, even if data have low levels of cross-classification. Our paper is timely because it adds to the evidence base needed to support the nascent field of multilevel interventions in cancer prevention and control and addresses the stated need for additional measurement research in this field.(12) In particular, our results add additional evidence suggesting the importance of multilevel factors across the cancer continuum, the continued need for multilevel intervention research, and the complexity of “real world” multilevel data structures. Future multilevel research and intervention should consider leveraging variations between, within, and *across* all levels, with particular attention to physician variation in FOBT.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

GRANT SUPPORT:

This work was supported by the Cancer Prevention Research Institute of Texas (CPRIT) PP100039 (PI: S. Gupta), CPRIT R1208 (PI: S.L. Pruitt), Agency for Healthcare Research and Quality R24 HS 22418-01 (PI: E.A. Halm), the National Cancer Institute R01 CA137750-02 (PI: M. Schootman) and U54CA163308 (S. Gupta). Contents of this paper are solely the responsibility of the authors and do not necessarily represent the official view of the NIH.

We thank Amy Speer for her assistance obtaining American Community Survey data and calculating driving times and Katherine McCallister and Bennett Grinsfelder for their assistance developing the figures.

References

1. Vital signs: colorectal cancer screening, incidence, and mortality --- United States, 2002–2010. *MMWR Morb Mortal Wkly Rep.* 2011;884–9. [PubMed: 21734636]
2. Taplin SH, Rodgers AB. Toward improving the quality of cancer care: addressing the interfaces of primary and oncology-related subspecialty care. *J Natl Cancer Inst Monogr.* 2010;3–10. [PubMed: 20386048]
3. Taplin SH, Anhang Price R, Edwards HM, Foster MK, Breslau ES, Chollette V, et al. Introduction: Understanding and influencing multilevel factors across the cancer care continuum. *J Natl Cancer Inst Monogr.* 2012;2–10. [PubMed: 22623590]
4. Kreft, I.; De Leeuw, J. *Introducing multilevel modeling.* Los Angeles, CA: Sage Publications; 1998.
5. Doubeni CA, Jambaulikar GD, Fouayzi H, Robinson SB, Gunter MJ, Field TS, et al. Neighborhood socioeconomic status and use of colonoscopy in an insured population--a retrospective cohort study. *PLoS One.* 2012:e36392. [PubMed: 22567154]
6. Lian M, Schootman M, Yun S. Geographic variation and effect of area-level poverty rate on colorectal cancer screening. *BMC Public Health.* 2008;358. [PubMed: 18925965]

7. Mobley LR, Kuo TM, Urato M, Subramanian S. Community contextual predictors of endoscopic colorectal cancer screening in the USA: spatial multilevel regression analysis. *Int J Health Geogr.* 2010;44. [PubMed: 20815882]
8. Shariff-Marco S, Breen N, Stinchcomb DG, Klabunde CN. Multilevel predictors of colorectal cancer screening use in California. *Am J Managed Care.* 2013;205–16.
9. Klabunde CN, Lanier D, Nadel MR, McLeod C, Yuan G, Vernon SW. Colorectal cancer screening by primary care physicians: recommendations and practices, 2006–2007. *Am J Prev Med.* 2009;8–16. [PubMed: 19442479]
10. Sarfaty M, Myers RE, Harris DM, Borsky AE, Sifri R, Cocroft J, et al. Variation in colorectal cancer screening steps in primary care: basis for practice improvement. *Am J Medical Quality: the official journal of the American College of Medical Quality.* 2012;458–66.
11. Yabroff KR, Zapka J, Klabunde CN, Yuan G, Buckman DW, Haggstrom D, et al. Systems strategies to support cancer screening in U.S. primary care practice. *Cancer Epidemiol Biomarkers Prev.* 2011;2471–9. [PubMed: 21976292]
12. Monograph. Understanding and influencing multilevel factors across the cancer care continuum. *J Natl Cancer Inst.* 2012;44.
13. Tiro, JA.; Kamineni, A.; Levin, TR.; Schottinger, JE.; Zhen, Y.; Chubak, J., et al. Conceptualizing the colorectal cancer (CRC) screening process in community settings: the population-based research optimizing screening through personalized regimens (PROSPR) network. Under review
14. Bell, BA.; Owens, CM.; Ferron, JM.; Kromrey, JD. Parsimony vs. Complexity: A Comparison of Two-Level, Three-Level, and Cross-Classified Models Using Add Health and AHAA Data. SESUG Proceedings (c) SESUG, Inc; (<http://www.sesug.org>) Paper PO091 from <http://analytics.ncsu.edu/sesug/2008/PO-091pdf>
15. Myers JL. The impact of inappropriate modeling of cross-classified data structures. *Multivariate Behavioral Research.* 2006;473–7.
16. Stange KC, Breslau ES, Dietrich AJ, Glasgow RE. State-of-the-art and future directions in multilevel interventions across the cancer control continuum. *J Natl Cancer Inst Monogr.* 2012;20–31. [PubMed: 22623592]
17. Gupta S, Halm EA, Rockey DC, Hammons M, Koch M, Carter E, et al. Comparative effectiveness of fecal immunochemical test outreach, colonoscopy outreach, and usual care for boosting colorectal cancer screening among the underserved: a randomized trial. *JAMA Internal Medicine.* 2013;1725–32. [PubMed: 23921906]
18. Pollack LA, Blackman DK, Wilson KM, Seeff LC, Nadel MR. Colorectal cancer test use among Hispanic and non-Hispanic U.S. populations. *Prev Chronic Dis.* 2006;A50. [PubMed: 16539791]
19. Seeff LC, Nadel MR, Klabunde CN, Thompson T, Shapiro JA, Vernon SW, et al. Patterns and predictors of colorectal cancer test use in the adult U.S. population. *Cancer.* 2004;2093–103. [PubMed: 15139050]
20. Cokkinides VE, Chao A, Smith RA, Vernon SW, Thun MJ. Correlates of underutilization of colorectal cancer screening among U.S. adults, age 50 years and older. *Prev Med.* 2003;85–91. [PubMed: 12473428]
21. Diaz JA, Roberts MB, Goldman RE, Weitzen S, Eaton CB. Effect of language on colorectal cancer screening among Latinos and non-Latinos. *Cancer Epidemiol Biomarkers Prev.* 2008;2169–73. [PubMed: 18708410]
22. Rossi PG, Federici A, Bartolozzi F, Sarchi S, Borgia P, Guasticchi G. Understanding non-compliance to colorectal cancer screening: a case control study, nested in a randomised trial [ISRCTN83029072]. *BMC Public Health.* 2005;135. [PubMed: 16356183]
23. OpenStreetMap Wiki. [Accessed October 2, 2013] http://wiki.openstreetmap.org/wiki/Main_Page
24. [Accessed October 2, 2013] MapQuest Developers - Open Data Map APIs and Web Services. <http://developer.mapquest.com/web/products/open>
25. Snijders, TAB.; Bosker, RJ. An introduction to basic and advanced multilevel modeling. London, England: Sage Publications; 1999. Multilevel analysis.
26. Merlo J, Chaix B, Ohlsson H, Beckman A, Johnell K, Hjerpe P, et al. A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *J Epi Community Health.* 2006;290–7.

27. Spiegelhalter DJ. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society Series B: Statistical Methodology*. 2002;583–616.
28. Pruitt SL, Shim MJ, Mullen PD, Vernon SW, Amick BC. The association of area socioeconomic status and breast, cervical, and colorectal cancer screening: A systematic review. *Cancer Epidemiol Biomarkers Prev*. 2009;2579–99. [PubMed: 19815634]
29. WL, Kwok O. The impacts of ignoring a crossed factor in analyzing cross-classified data. *Multivariate Behavioral Research*. 2009;182–212.
30. Anhang Price R, Zapka J, Edwards H, Taplin SH. Organizational factors and the cancer screening process. *J Natl Cancer Inst Monogr*. 2010;38–57. [PubMed: 20386053]
31. Stone EG, Morton SC, Hulscher ME, Maglione MA, Roth EA, Grimshaw JM, et al. Interventions that increase use of adult immunization and cancer screening services: a meta-analysis. *Ann Intern Med*. 2002;641–51. [PubMed: 11992299]
32. Sarfaty M, Wender R. How to increase colorectal cancer screening rates in practice. *CA Cancer J Clin*. 2007; 57(6):354–66. [PubMed: 17989130]
33. Hudson SV, Ohman-Strickland P, Cunningham R, Ferrante JM, Hahn K, Crabtree BF. The effects of teamwork and system support on colorectal cancer screening in primary care practices. *Cancer Detect Prev*. 2007;417–23. [PubMed: 18031947]
34. Subramanian S, Klosterman M, Amonkar MM, Hunt TL. Adherence with colorectal cancer screening guidelines: a review. *Prev Med*. 2004;536–50. [PubMed: 15066356]
35. Brawarsky P, Brooks DR, Mucci LA, Wood PA. Effect of physician recommendation and patient adherence on rates of colorectal cancer testing. *Cancer Detect Prev*. 2004;260–8. [PubMed: 15350629]
36. Inadomi JM, Vijan S, Janz NK, Fagerlin A, Thomas JP, Lin YV, et al. Adherence to colorectal cancer screening: a randomized clinical trial of competing strategies. *Arch Intern Med*. 2012;575–82. [PubMed: 22493463]
37. Manski, CF. Identification problems in the social sciences. In: PVM, editor. *Sociological methodology*. San Francisco, CA: Jossey-Banks; 1993.
38. Snijders, TAB.; Bosker, R. *An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage; 1999. Multilevel analysis.

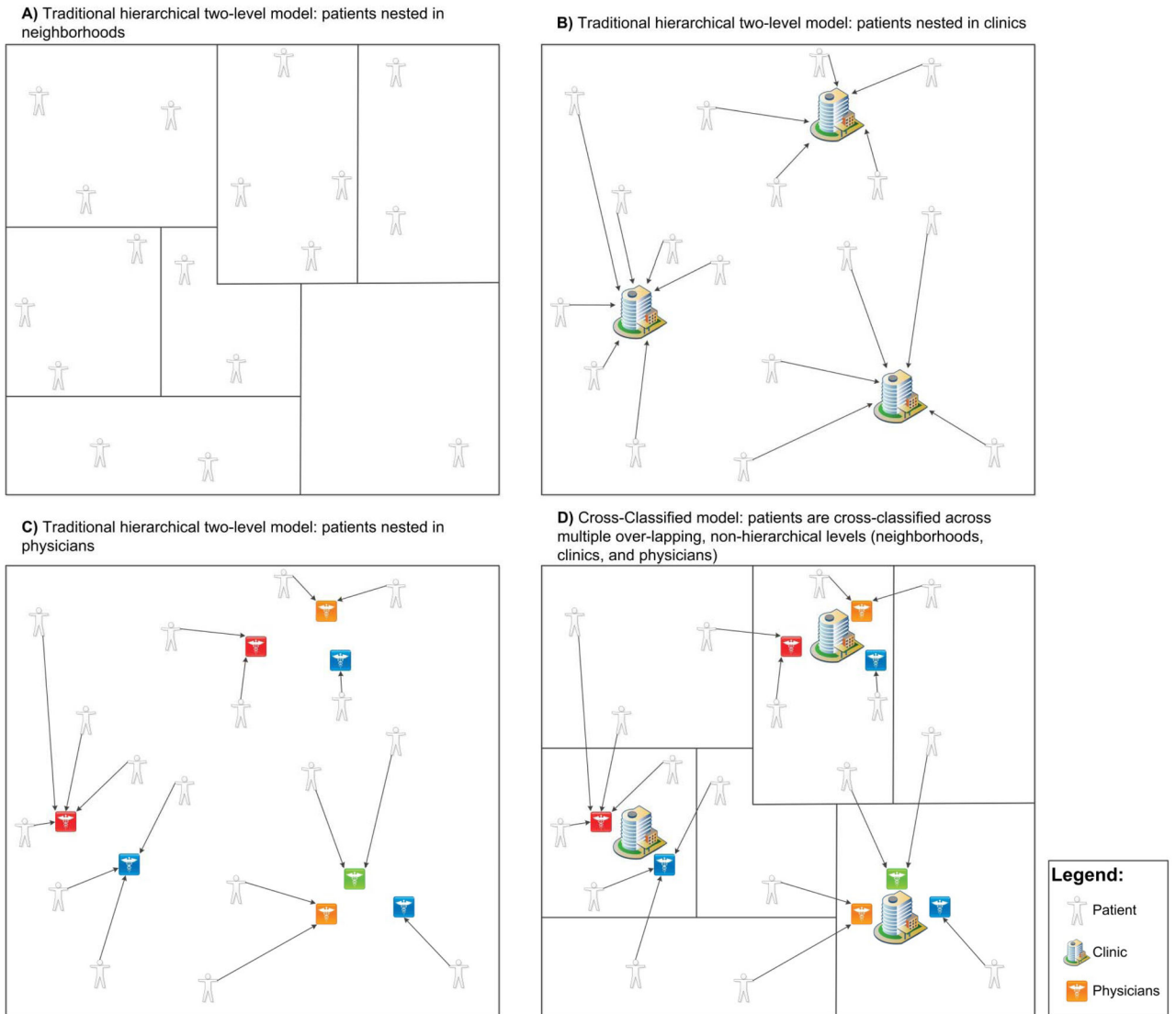


Figure 1.
Figure 1A–D. Hypothetical multilevel data structures.

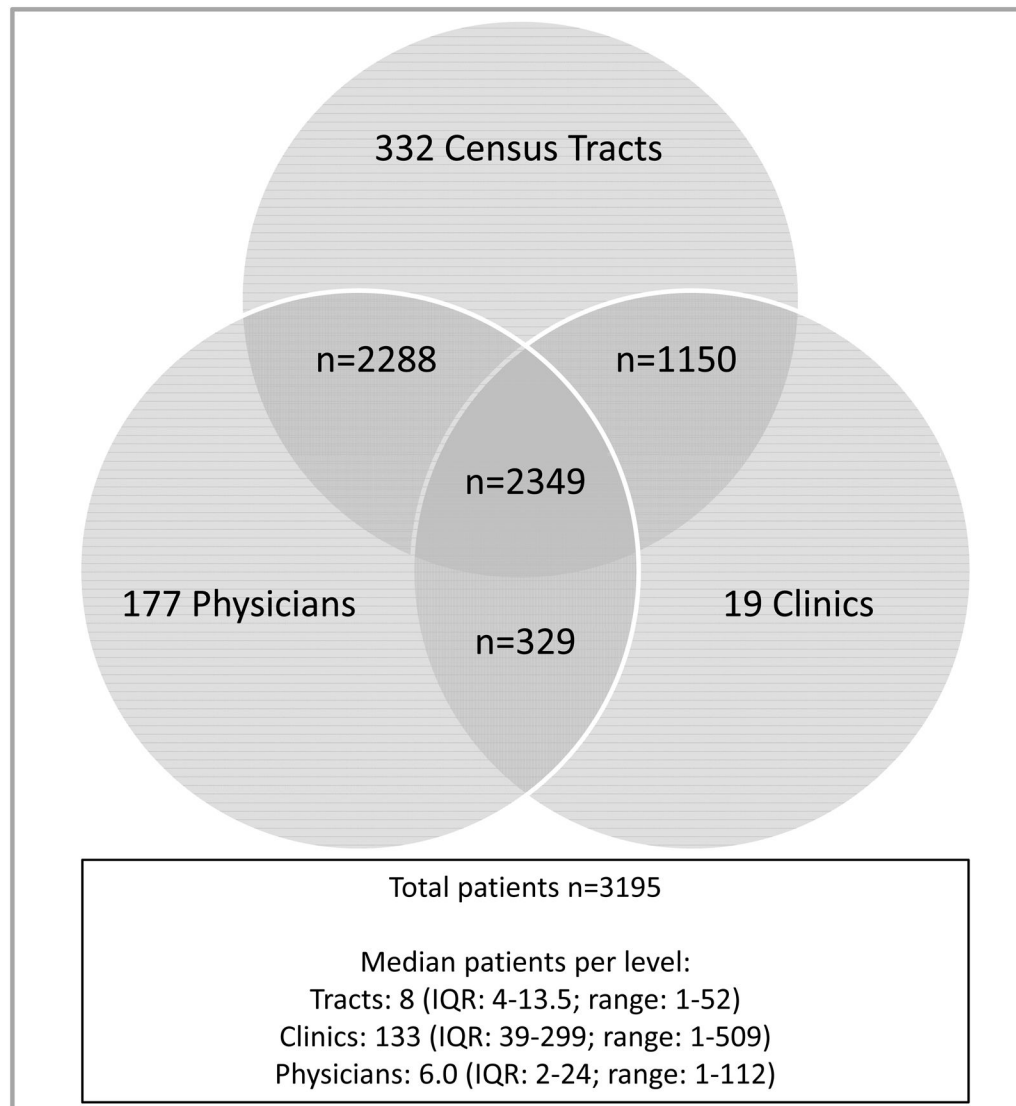


Figure 2. Cross-classified structure of the data by physicians, clinics, and residential census tracts. Numbers in overlapping circles indicate the number of patients in the unique combinations of each data structure.
 IQR=interquartile range

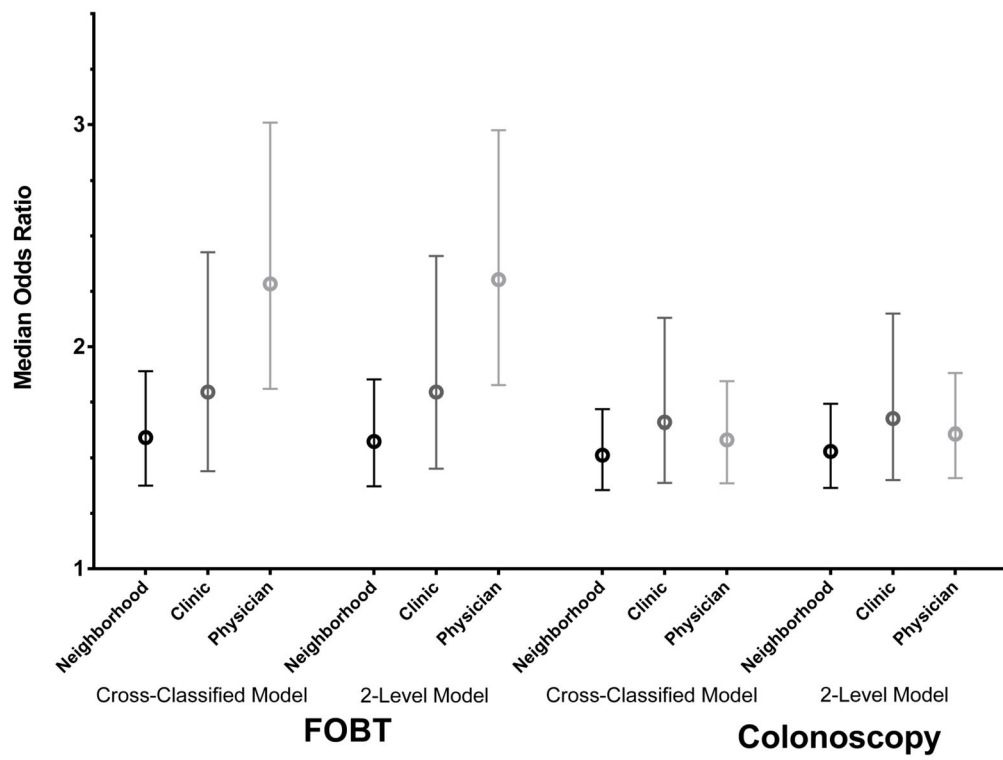


Figure 3. Variation across neighborhoods, clinics, and physicians in hierarchical 2-level and cross-classified adjusted models for FOBT and colonoscopy test use. Note: Circles represent point estimates and whiskers represent 95% credible intervals of MOR (median odds ratios), data shown in Table 2.

Table 1
Patient characteristics by uptake of FOBT and colonoscopy test use among usual care patients (n=3195)

	FOBT				Colonoscopy					
	No	Yes	No	Yes	No	Yes	No	Yes		
	N(%)	Mean (SD)	N(%)	Mean (SD)	p (Chi ² /t-test)	N(%)	Mean (SD)	N(%)	Mean (SD)	p (Chi ² /t-test)
Age	59.0 (2.9)	58.9 (3.0)	59.0 (2.9)	59.2 (2.9)	.652	59.0 (2.9)	59.2 (2.9)	59.0 (2.9)	59.2 (2.9)	.386
Sex										
Male	1002 (33.0)	46 (29.3)	950 (32.7)	98 (33.6)	.338	950 (32.7)	98 (33.6)	950 (32.7)	98 (33.6)	.772
Female	2036 (67.0)	111 (67.7)	1953 (67.28)	194 (66.4)		1953 (67.28)	194 (66.4)	1953 (67.28)	194 (66.4)	
Race										
non-Hispanic White	1203 (39.6)	58 (36.9)	1177 (40.5)	84 (28.8)	.754	1177 (40.5)	84 (28.8)	1177 (40.5)	84 (28.8)	<.001
non-Hispanic Black	724 (23.8)	40 (45.5)	670 (23.1)	94 (32.2)		670 (23.1)	94 (32.2)	670 (23.1)	94 (32.2)	
Hispanic	931 (30.7)	47 (29.9)	881 (30.4)	97 (33.2)		881 (30.4)	97 (33.2)	881 (30.4)	97 (33.2)	
Other	180 (5.9)	12 (7.6)	175 (6.0)	17 (5.8)		175 (6.0)	17 (5.8)	175 (6.0)	17 (5.8)	
Primary Language										
English	2513 (82.7)	129 (82.2)	2399 (82.6)	243 (83.2)	.858	2399 (82.6)	243 (83.2)	2399 (82.6)	243 (83.2)	.803
Spanish	525 (17.3)	28 (17.8)	504 (17.4)	49 (16.8)		504 (17.4)	49 (16.8)	504 (17.4)	49 (16.8)	
Driving time (minutes) to Colonoscopy Clinic	14.8 (6.5)	14.5 (5.8)	14.7 (6.5)	15.3 (5.9)	.520	14.7 (6.5)	15.3 (5.9)	14.7 (6.5)	15.3 (5.9)	.133
Driving time (minutes) to Primary Care Clinic	10.9 (6.3)	11.1 (6.1)	10.9 (6.3)	11.1 (5.9)	.638	10.9 (6.3)	11.1 (5.9)	10.9 (6.3)	11.1 (5.9)	.554
Neighborhood Poverty Rate	20.2 (13.3)	19.4 (13.3)	20.3 (13.5)	19.2 (19.2)	.434	20.3 (13.5)	19.2 (19.2)	20.3 (13.5)	19.2 (19.2)	.200

Table 2

Primary care physician, clinic, and neighborhood (census tract) variability in hierarchical 2-level and cross-classified empty and adjusted models in FOBT and colonoscopy test use among usual care patients (n=3,195).

		Empty Models				Adjusted Models ^a						
	Parameter	95% CI	Parameter	95% CI	Parameter	95% CI	Parameter	95% CI	Parameter	95% CI		
FOBT Test Use												
Hierarchical 2-Level Models												
	Model 1 PCP variability		Model 2 Clinic variability		Model 3 Neighborhood variability		Model 1 PCP variability		Model 2 Clinic variability	Model 3 Neighborhood variability		
Variance	0.735	0.399–1.306	0.334	0.153–0.850	0.213	0.110–0.418	0.773	0.404–1.350	0.356	0.159–0.889	0.216	0.111–0.417
MOR	2.27	1.83–2.97	1.74	1.45–2.41	1.55	1.37–1.85	2.31	1.83–3.03	1.77	1.46–2.46	1.56	1.37–1.85
Cross-Classified Model 4												
	PCP variability		Clinic variability		Neighborhood variability		PCP variability		Clinic variability	Neighborhood variability		
Variance	0.716	0.387–1.33	0.332	0.146–0.863	0.224	0.112–0.445	0.765	0.413–1.40	0.359	0.149–0.989	0.227	0.114–0.452
MOR	2.24	1.81–3.01	1.73	1.44–2.43	1.57	1.38–1.89	2.30	1.85–3.09	1.77	1.45–2.58	1.57	1.38–1.89
Colonoscopy Test Use												
Hierarchical 2-Level Models												
	Model 1 PCP variability		Model 2 Clinic variability		Model 3 Neighborhood variability		Model 1 PCP variability		Model 2 Clinic variability	Model 3 Neighborhood variability		
Variance	0.235	0.129–0.439	0.263	0.125–0.643	0.190	0.106–0.339	0.221	0.114–0.402	0.242	0.118–0.576	0.197	0.103–0.352
MOR	1.59	1.41–1.88	1.63	1.40–2.15	1.52	1.36–1.74	1.57	1.38–1.83	1.60	1.39–2.06	1.53	1.36–1.76
Cross-Classified Model 4												
	PCP variability		Clinic variability		Neighborhood variability		PCP variability		Clinic variability	Neighborhood variability		
Variance	0.219	0.117–0.412	0.253	0.118–0.628	0.180	0.102–0.322	0.215	0.113–0.399	0.240	0.112–0.601	0.194	0.104–0.349

	Empty Models				Adjusted Models ^a							
	Parameter	95% CI	Parameter	95% CI	Parameter	95% CI	Parameter	95% CI				
MOR	1.56	1.39-1.84	1.62	1.39-2.13	1.50	1.36-1.72	1.56	1.38-1.83	1.60	1.38-2.09	1.52	1.36-1.76

MOR: Median odds ratio; CI: credible interval; PCP: primary care provider

^a Adjusted model covariates include: sex, race/ethnicity, age, primary language, neighborhood poverty, and driving time to endoscopy clinic and primary care clinic

Table 3

Model fit indicated with deviance information criterion (DIC) comparing traditional hierarchical 2-level and novel cross-classified empty and adjusted models in FOBT and colonoscopy test use among usual care patients (n=3,195).

	Hierarchical 2-Level Models			
	Cross-classified Model	Physician Model	2 Clinic Model	Neighborhood Model
FOBT Models				
Empty	1186.43*	1187.52*	1239.77	1262.70
Adjusted	1195.38*	1196.87*	1250.64	1274.06
Colonoscopy Models				
Empty	1953.42*	1946.88*	1947.37*	1961.79
Adjusted	1944.90*	1938.27*	1939.24*	1948.81

Note:

* Indicates best fitting mode(s) of all models in row. Lower values indicate better fit; a difference of 3 to 7 in DIC indicates material difference between models.(27)