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

Sandi L. Pruitt1,3, Tammy Leonard4, Song Zhang1, Mario Schootman5, Ethan A. Halm1,2, and Samir Gupta6,7

Abstract

Background: We (i) described variability in colorectal cancer (CRC) test use across multiple levels, including physician, clinic, and neighborhood; and (ii) compared the performance of novel cross-classified models versus 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: (i) fecal occult blood test (FOBT) or (ii) colonoscopy and were ascertained using claims data during a 1-year follow-up. We compared Bayesian (i) cross-classified four-level logistic models nesting patients within separate, nonoverlapping “levels” (physicians, clinics, and census tracts) versus (ii) three hierarchical two-level models using deviance information criterion. Models were 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 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. Cancer Epidemiol Biomarkers Prev; 23(7); 1346–55. ©2014 AACR.

Introduction

Although 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 United States 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 interrelation 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

Authors’ Affiliations: 1Department of Clinical Sciences; 2Department of General Internal Medicine, University of Texas Southwestern Medical Center; 3Harold C. Simmons Comprehensive Cancer Center; 4School of Economic, Political, and Policy Sciences, University of Texas-Dallas, Dallas, Texas; 5Department of Epidemiology, College for Public Health and Social Justice, Saint Louis University, St. Louis, Missouri; 6Department of Gastroenterology, Department of Internal Medicine, Moores Cancer Center, University of California San Diego; and 7Department of Veterans Affairs, San Diego Healthcare System, San Diego, California

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (http://cebp.aacrjournals.org/).

Corresponding Author: Sandi L. Pruitt, 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; E-mail: Sandi.Pruitt@utsouthwestern.edu

doi: 10.1158/1055-9965.EPI-13-1130

©2014 American Association for Cancer Research.
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 2 may arise from independent causal processes.

Although multilevel conceptual frameworks acknowledge numerous levels of influence (3), traditional multilevel analyses of CRC screening typically include 2, or at most, 3, strictly hierarchical levels—an oversimplification of the true complexity present in the CRC screening continuum. For example, Fig. 1A–C depict hierarchical data structures assumed in traditional multilevel models: patients are assumed to be nested in nonoverlapping census tracts (Fig. 1A), or assigned to single physicians (Fig. 1B) or clinics (Fig. 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 Fig. 1D, wherein patients are simultaneously cross-classified across multiple overlapping, nonhierarchical levels. For example, patients from the same neighborhood may attend different clinics and physicians in some healthcare systems practice in more than 1 clinic.

Figure 1. A–D, hypothetical multilevel data structures.

Legend:
- Patient
- Clinic
- Physician
Neglecting data cross-classification can result in misspecified statistical models and potentially spurious conclusions (14, 15). For example, model misspecification can bias estimates of fixed effects, standard errors, and variance components. Thus, observed variation 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, although multilevel factors associated with CRC screening have been identified in descriptive research, these results have not 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 2 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 2 multilevel modeling techniques. To our knowledge, we are the first study to simultaneously examine CRC test use uptake across multiple, overlapping, nonhierarchical 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 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). Furthermore, 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 to 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 this 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.).

Measures
We measured 2 outcomes because we hypothesized that the influence of different "levels" (e.g., clinic, physician) varied by test modality. Outcomes included completion of either a (i) simple noninvasive stool blood test (FOBT) or (ii) more complicated, facility-based tests (colonoscopy, barium enema, or sigmoidoscopy) within 1 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 (i) primary-care clinic and (ii) 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
multiple, 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; mea-
sures were moderately correlated ($r = 0.39$). 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 ran-
dom effects logistic models, with patients nested within (i)
primary care physicians (Fig. 1A); (ii) primary care clinics
(Fig. 1B); or (iii) census tracts (Fig. 1C). Next, we fitted one
Bayesian, cross-classified random effects logistic model
(25) allowing patients to be cross-classified within multi-
ple separate, nonhierarchical grouping "levels" (Fig. 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 indepen-
dent 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:

$$\log \left( \frac{P_i}{1 - P_i} \right) = \beta^T x_i + \theta_{(j)} + \delta_{(d)} \quad \text{for } i = 1, \ldots, I.$$  \hspace{1cm} (1)

where $Y_i$, generically denotes the binary outcome of the $i$th
subject (1 for yes, 0 for no). We use $x_i = (x_{i1}, \ldots, x_{id})^T$, a vector of length $p$, to denote the vector
of covariates from subject $i$. The vector of regression coeffi-
cients is defined as $\beta = (\beta_1, \ldots, \beta_p)^T$. Note that we set
$x_{i1} = 1$ for $i = 1, \ldots, I$, which implies that $\beta_1$ is the
intercept. Suppose that subject $i$ is nested in the $j$th ($j = 1, \ldots, J$)
physician; the $k$th ($k = 1, \ldots, K$) clinic, and the $l$th ($l = 1, \ldots, L$) census tract. We define $\theta_j$, $\theta_k$, and $\delta_l$ to be the
random effect for physician $j$, clinic $k$, and census tract $l$, respectively. All analyses are based on logistic regression
models. We defined $P_i = P(Y_i = 1)$. Here subscript $(j(i), k(i)), l(i)$ 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
predictors, see Supplemenary 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 ($V$):

$$\text{MOR} = \exp(0.95\sqrt{V})$$. It is interpreted as the median value of the ratio of predicted odds of the outcome for 2
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 Mar-
kov Chain Monte Carlo methods.

To evaluate methods for assessing variation, we com-
pared traditional 2-level and novel cross-classified mod-
els 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 indi-
cate 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 1-year
follow-up period; only 8.1% ($n = 299$), received colonos-
cy 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
Fig. 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 combina-
tions 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 FOBT testing. Across all 2-level
models, physicians demonstrated the largest variability
followed by clinics, and finally, neighborhoods. In 2-level
hierarchical models, the adjusted model MORs were 2.31
(95% CI, 1.83−3.03) for physicians, 1.77 (95% CI, 1.46−2.46)
for clinics, and 1.56 (95% CI, 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 and in
separate 2-level models (Table 2 and Fig. 3).

**Multilevel variation in colonoscopy**

Both empty and adjusted models demonstrated statisti-
cally 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 (95% CI, 1.39–2.06) for clinics, 1.57 (95% CI, 1.38–1.83) for physicians, and 1.53 (95% CI, 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 (Fig. 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 with 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 with 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

| Table 1. Patient characteristics by uptake of FOBT and colonoscopy test use among usual care patients (n = 3,195) |

<table>
<thead>
<tr>
<th></th>
<th>FOBT</th>
<th>Colonoscopy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N(%)/mean (SD)</td>
<td>N(%)/mean (SD)</td>
</tr>
<tr>
<td>Age</td>
<td>59.0 (2.9)</td>
<td>58.9 (3.0)</td>
</tr>
<tr>
<td>Sex</td>
<td>Male 1,002 (33.0)</td>
<td>46 (29.3)</td>
</tr>
<tr>
<td></td>
<td>Female 2036 (67.0)</td>
<td>1,11 (67.7)</td>
</tr>
<tr>
<td>Race</td>
<td>Non-Hispanic white 1,203 (39.6)</td>
<td>58 (36.9)</td>
</tr>
<tr>
<td></td>
<td>Non-Hispanic black 724 (23.8)</td>
<td>40 (45.5)</td>
</tr>
<tr>
<td></td>
<td>Hispanic 931 (30.7)</td>
<td>47 (29.9)</td>
</tr>
<tr>
<td></td>
<td>Other 180 (5.9)</td>
<td>12 (7.6)</td>
</tr>
<tr>
<td>Primary language</td>
<td>English 2,513 (82.7)</td>
<td>129 (82.2)</td>
</tr>
<tr>
<td></td>
<td>Spanish 525 (17.3)</td>
<td>28 (17.8)</td>
</tr>
<tr>
<td>Driving time (minutes) to colonoscopy clinic</td>
<td>14.8 (6.5)</td>
<td>14.5 (5.8)</td>
</tr>
<tr>
<td>Driving time (minutes) to primary care clinic</td>
<td>10.9 (6.3)</td>
<td>11.1 (6.1)</td>
</tr>
<tr>
<td>Neighborhood poverty rate</td>
<td>20.2 (13.3)</td>
<td>19.4 (13.3)</td>
</tr>
</tbody>
</table>

CI, 1.38–1.83 for physicians, and 1.53 (95% CI, 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 (Fig. 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 with 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 with 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

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.
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 ($n = 3,195$)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>95% CI</th>
<th>Parameter</th>
<th>95% CI</th>
<th>Parameter</th>
<th>95% CI</th>
<th>Parameter</th>
<th>95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FOBT test use</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hierarchical 2-level models</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>PCP variability</td>
<td>0.735</td>
<td>0.399–1.306</td>
<td>0.334</td>
<td>0.153–0.850</td>
<td>0.213</td>
<td>0.110–0.418</td>
</tr>
<tr>
<td>Model 2</td>
<td>Clinic variability</td>
<td>0.334</td>
<td>0.153–0.850</td>
<td>0.213</td>
<td>0.110–0.418</td>
<td>0.110–0.418</td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td>Neighborhood variability</td>
<td>0.110–0.418</td>
<td>0.213</td>
<td>0.110–0.418</td>
<td>0.110–0.418</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOR</td>
<td>2.27</td>
<td>1.83–3.97</td>
<td>1.74</td>
<td>1.45–2.41</td>
<td>1.55</td>
<td>1.37–1.85</td>
<td></td>
</tr>
<tr>
<td>Cross-classified model 4</td>
<td>0.773</td>
<td>0.404–1.350</td>
<td>0.356</td>
<td>0.159–0.889</td>
<td>0.216</td>
<td>0.11–0.417</td>
<td></td>
</tr>
<tr>
<td>MOR</td>
<td>2.31</td>
<td>1.83–3.03</td>
<td>1.77</td>
<td>1.46–2.46</td>
<td>1.56</td>
<td>1.37–1.85</td>
<td></td>
</tr>
<tr>
<td><strong>Colonoscopy test use</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hierarchical 2-level models</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 1</td>
<td>PCP variability</td>
<td>0.716</td>
<td>0.387–1.33</td>
<td>0.332</td>
<td>0.146–0.863</td>
<td>0.224</td>
<td>0.112–0.445</td>
</tr>
<tr>
<td>Model 2</td>
<td>Clinic variability</td>
<td>0.332</td>
<td>0.146–0.863</td>
<td>0.224</td>
<td>0.112–0.445</td>
<td>0.112–0.445</td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td>Neighborhood variability</td>
<td>0.112–0.445</td>
<td>0.224</td>
<td>0.112–0.445</td>
<td>0.112–0.445</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOR</td>
<td>2.24</td>
<td>1.81–3.01</td>
<td>1.73</td>
<td>1.44–2.43</td>
<td>1.57</td>
<td>1.38–1.89</td>
<td></td>
</tr>
<tr>
<td>Cross-classified model 4</td>
<td>0.765</td>
<td>0.413–1.40</td>
<td>0.359</td>
<td>0.149–0.989</td>
<td>0.227</td>
<td>0.114–0.452</td>
<td></td>
</tr>
<tr>
<td>MOR</td>
<td>2.30</td>
<td>1.85–3.09</td>
<td>1.77</td>
<td>1.45–2.58</td>
<td>1.57</td>
<td>1.38–1.89</td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CI, credible interval; MOR, median odds ratio; PCP, primary care provider.

*Adjusted model covariates include: sex, race/ethnicity, age, primary language, neighborhood poverty, and driving time to endoscopy clinic and primary care clinic.
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 2 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 2 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 versus multiple traditional

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

<table>
<thead>
<tr>
<th></th>
<th>Cross-classified model</th>
<th>Physician model</th>
<th>Clinic model</th>
<th>Neighborhood model</th>
</tr>
</thead>
<tbody>
<tr>
<td>FOBT models</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empty</td>
<td>1186.43$^a$</td>
<td>1187.52$^a$</td>
<td>1239.77</td>
<td>1262.70</td>
</tr>
<tr>
<td>Adjusted</td>
<td>1195.38$^a$</td>
<td>1196.87$^a$</td>
<td>1250.64</td>
<td>1274.06</td>
</tr>
<tr>
<td>Colonoscopy models</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Empty</td>
<td>1953.42$^a$</td>
<td>1946.88$^a$</td>
<td>1947.37$^a$</td>
<td>1961.79</td>
</tr>
<tr>
<td>Adjusted</td>
<td>1944.90$^a$</td>
<td>1938.27$^a$</td>
<td>1939.24$^a$</td>
<td>1948.81</td>
</tr>
</tbody>
</table>

$^a$Indicates best fitting model(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).
Multiple Levels of Influence

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. Furthermore, 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 about 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 with colonoscopy across all levels. Similarly, a Missouri study demonstrated larger area-level variation in FOBT testing as compared with 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 used provider reminders and only 15% used patient reminders in 2006 to 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 to 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. Furthermore, 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. Finally, 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, nonhierarchical levels. In doing so, we provide early empirical evidence supporting the
simultaneous consideration of variation across multiple, nonhierarchical 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 multi-level analyses can bias standard errors and alter statistical inference (25). 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. Multi-level 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 2 different multilevel modeling techniques, we found that multiple levels of influence exert influence on CRC testing. Our conclusions were similar when using cross-classified versus 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 article 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.

Disclosure of Potential Conflicts of Interest
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.

Authors’ Contributions
Conception and design: S.L. Pruitt, M. Schootman, E.A. Halm, S. Gupta Development of methodology: S.L. Pruitt, M. Schootman, S. Gupta Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): S. Zhang, S. Gupta Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): S.L. Pruitt, T. Leonard, S. Zhang, M. Schootman, E.A. Halm, S. Gupta Writing, review, and/or revision of the manuscript: S.L. Pruitt, T. Leonard, E.A. Halm, S. Gupta Study supervision: S.L. Pruitt, T. Leonard

Acknowledgments
The authors thank A. Hughes for her assistance obtaining American Community Survey data and calculating driving times and K. McCallister and D.B. Grinsfelder for their assistance developing the figures.

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 U54 CA163308 (S. Gupta). Contents of this article are solely the responsibility of the authors and do not necessarily represent the official view of the NIH.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received October 29, 2013; revised March 5, 2014; accepted March 21, 2014; published OnlineFirst April 14, 2014.

Multiple Levels of Influence


