Potential Biases Introduced by Conflating Screening and Diagnostic Testing in Colorectal Cancer Screening Surveillance

Elizabeth A. Becker1, Derek M. Griffith2, Brady T. West3, Nancy K. Janz4, Ken Resnicow4, and Arden M. Morris5

Abstract

Background: Screening and postsymptomatic diagnostic testing are often conflated in cancer screening surveillance research. We examined the error in estimated colorectal cancer screening prevalence due to the conflation of screening and diagnostic testing.

Methods: Using data from the 2008 National Health Interview Survey, we compared weighted prevalence estimates of the use of all testing (screening and diagnostic) and screening in at-risk adults and calculated the overestimation of screening prevalence across sociodemographic groups.

Results: The population screening prevalence was overestimated by 23.3%, and the level of overestimation varied widely across sociodemographic groups (median, 22.6%; mean, 24.8%). The highest levels of overestimation were in non-Hispanic white females (27.4%), adults ages 50–54 years (32.0%), and those with the highest socioeconomic vulnerability [low educational attainment (31.3%), low poverty ratio (32.5%), no usual source of health care (54.4%), and not insured (51.6%); all \( P < 0.001 \)].

Conclusions: When the impetus for testing was not included, colorectal cancer screening prevalence was overestimated, and patterns of overestimation often aligned with social and economic vulnerability. These results are of concern to researchers who use survey data from the Behavioral Risk Factor Surveillance System (BRFSS) to assess cancer screening behaviors, as it is currently not designed to distinguish diagnostic testing from screening.

Impact: Surveillance research in cancer screening that does not consider the impetus for testing risks measurement error of screening prevalence, impeding progress toward improving population health. Ultimately, to craft relevant screening benchmarks and interventions, we must look beyond “what” and “when” and include “why.”

Cancer Epidemiol Biomarkers Prev; 24(12): 1–5. ©2015 AACR.

Introduction

By definition, cancer screening occurs at a presymptomatic stage of disease. Clinical tests such as colonoscopy and mammography are used for both presymptomatic screening and postsymptomatic diagnostic testing. Consequently, researchers and practitioners often collect and analyze surveillance data in a way that conflates these two distinct behavioral outcomes. Potential challenges of this practice include overestimation of screening prevalence, misclassification of screening trends over time, and mischaracterization of within- and between-group disparities.

A distinguishing feature of two of the most cited national surveys in cancer screening research, the National Health Interview Survey (NHIS) and the Behavioral Risk Factor Surveillance System (BRFSS), is their varying ability to differentiate screening from diagnostic testing. NHIS affords researchers the opportunity to distinguish between screening and diagnostic testing by asking respondents not only what test they had and when, but why they had it. In contrast, conflation of screening and diagnostic testing is unavoidable for users of BRFSS data, as BRFSS does not collect data on the impetus for testing (1–3).

The impact of conflating screening and diagnostic testing on the quality of cancer screening surveillance data has not been explored. We used data from the NHIS dataset and the methods of BRFSS to estimate (i) the potential measurement error in colorectal cancer screening prevalence due to conflation of screening and diagnostic testing and (ii) the variation in degree of measurement error in colorectal cancer screening prevalence across sociodemographic characteristics.

Materials and Methods

Study population and data source

Our study population included non-Hispanic white and non-Hispanic black respondents ages 50–80 years with no history of colorectal cancer, as colonoscopic testing is used as disease surveillance and not preventative screening care. We analyzed nationally representative data from the 2008 NHIS and Cancer Screening and Sun Protection Supplement (CSSPS). The NHIS is administered annually by the National Center for Health Statistics (NCHS) of the Centers for Disease Control and Prevention (CDC) and uses a stratified multistage probability sample design (4). The CSSPS includes questions on physician recommendations for...
screening, screening behaviors, and reasons for screening tests. The NHIS oversamples self-identified non-Hispanic blacks, Hispanics and Asian Americans. Weights constructed for the NHIS respondents reflect the resulting unequal probabilities of selection and also incorporate adjustments for nonresponse and poststratification procedures designed to align survey estimates with population distributions from the 2000 Census. The NHIS data are therefore nationally representative of the adult civilian noninstitutionalized population of the United States (5). The annual NHIS response rate averages close to 90% of the eligible households in the sample (4).

This study received an exempt status designation from the University of Michigan Institutional Review Board (Study HUM00062074).

Measures

**Dependent variables.** American Cancer Society (ACS) guideline adherent testing We created a dichotomous variable using respondents’ self-reports of the type of testing they underwent and the timeframe in which the procedure occurred to establish whether the testing was ACS guideline adherent. At the time of the survey’s administration, the ACS recommendations included tests and time intervals as follows: flexible sigmoidoscopy (previously noted as proctoscopy) every 5 years, colonoscopy every 10 years, or fecal occult blood test (FOBT) every year (6). Additional testing modalities were also part of the ACS recommendations in 2008 but were not readily available to the general population and were not part of the NHIS questionnaire.

To assess engagement with noninvasive testing, respondents were given a brief explanation and asked, “Have you EVER HAD a blood stool test, using a HOME test kit?” Responses of “yes” or “no” were coded as such, whereas responses of “refused” or “don’t know” were coded as missing data and excluded from analysis (missing, n = 52). Respondents reported testing timelines in one of several formats: month/year, number of days, weeks, months or years since testing, or by using years since testing (a year ago or less/more than 1 year but not more than 2 years/more than 2 years but not more than 3 years/more than 3 years but not more than 5 years/more than 5 years but not more than 10 years/over 10 years ago). The dependent variable was coded as 1 if FOBT was reported within oni year of survey administration, 0 if no testing or FOBT testing was earlier than 1 year from survey administration, and missing. All FOBT testing was coded as screening, as it is used exclusively for that purpose.

To assess engagement with invasive testing, respondents were given a brief explanation and asked, “Have you ever had a sigmoidoscopy, colonoscopy, or proctoscopy?” Responses of “yes” or “no” were coded as such, whereas responses of “refused” or “don’t know” were coded as missing data and excluded from analysis (missing, n = 23). Invasive testing (sigmoidoscopy, proctoscopy, colonoscopy) was coded as ACS guideline adherent if the testing was reported to be within the recommended timeframe. One dependent variable was created for each mode of testing, coded as 1 if ACS guideline adherent, 0 if no testing or nonadherent, and missing otherwise.

ACS guideline adherent screening To measure reports of screening (vs. diagnostic testing) among those participants who reported testing, we created a dichotomous variable combining data on ACS guideline adherent screening behavior variable described above and NHIS data on the impetus for the reported testing. To determine the impetus for testing among individuals who reported invasive tests (sigmoidoscopy, proctoscopy, or colonoscopy), we assessed responses to the question “What was the MAIN reason you had this exam—was it part of a routine exam, because of a problem, or some other reason?” The dependent variable was coded as “screening” for responses of “part of a routine exam” and as “diagnostic testing” for responses of “because of a problem, or some other reason.” Responses of “refused” or “don’t know” were coded as missing data and excluded from analysis if noninvasive testing was not a “yes” (missing, n = 4).

Data from these 5 variables (FOBT, three invasive testing variables, and screening variable) informed the combined dependent variable measuring ACS guideline adherent screening. If any mode of testing was ACS guideline adherent screening, the variable was coded as 1, and if no mode of testing was ACS guideline adherent screening or no testing was reported, the variable was coded as 0. The variable was coded missing otherwise.

**Demographic variables.** We used NHIS data directly for variables indicating race/ethnicity (non-Hispanic white/non-Hispanic black), sex (male/female), age (continuous 50–80, categorical 50–54/55–59/60–64/65–69/70–74/75–80, and categorical 50–64/65–80), educational attainment (less than a high school diploma/high school graduate or GED/some college no degree or associate degree/bachelor’s degree/master professional or doctoral), poverty ratio [tertiles (measured as ratio of family income to poverty threshold): low (under 0.50 to 2.49)/medium (2.50 to 4.99)/high (5.00 and over]), insurance status (covered/not covered), and usual source of health care (yes/no). The poverty ratio variable came from NHIS with no missing values as the result of multiple imputation done by the survey administrator. Rates of missing values for all other variables of interest were below 3%.

**Statistical analysis**

We computed weighted and unweighted estimates of the prevalence of all testing (i.e., both screening and diagnostic testing that is consistent with BRFSS methodology), across levels of each of the independent variables of interest in the study population using data from the NHIS and CSSPS. We then repeated the analyses using the variable that accurately characterized presymptomatic screening.

We then calculated the overestimation of screening prevalence estimates by dividing the difference between the prevalence estimates of all testing (BRFSS methodology) and screening by the prevalence estimate of screening. For example, the prevalence estimate for testing for all participants is 55.5% and the screening prevalence estimate is 45.0%. Therefore, we calculated the overestimation of the screening prevalence as:

\[
\frac{(\text{all testing estimate} - \text{screening estimate})}{\text{screening estimate}}
\]

\[
\frac{(0.555 - 0.450)}{0.450} = 23.3%.
\]

Most data analyses for this study were performed with the SAS/STAT statistical software (version 9.3 M1, SAS Institute), which
enables weighted estimation and design-based variance estimation. We ran selected models in Stata (version 13.0, StataCorp), which enables multiple ad-hoc variance estimation methods for dealing with "singleton" PSUs, as the SAS SURVEY procedures use Taylor Series Linearization for variance estimation by default and exclude strata with only one primary sampling unit (PSU) from all variance estimates. We then compared estimated standard errors between the two program’s procedures.

### Results

#### Weighted frequencies

Table 1 shows estimated demographic characteristics for the full study population. In Table 2, we present estimated prevalence of all testing (conflating diagnostic testing and screening) and estimated prevalence of screening. Table 2 also displays the extent to which the screening prevalence is overestimated when the outcome is misspecified.

#### Colorectal cancer testing prevalence estimates

The colorectal cancer testing prevalence estimate for the NHIS study population was 55.5%. With the exception of differences across sex [males (M: 55.8%) and females (F: 55.2%), \( P = 0.6812 \)], all within-group differences were statistically significant at or below \( P = 0.0001 \) level. Colorectal cancer testing prevalence estimates for non-Hispanic whites [M (56.8%) and F (56.2%)] exceeded those of non-Hispanic blacks [M (47.9%) and F (48.2%)] and estimates increased with age (42.1% to 64.4%). Across measures of socioeconomic status, we saw estimates increase across levels of education (44.9% to 69.5%), poverty ratio (47.7% to 62.8%), and insurance coverage (24.1% to 57.8%). Finally, we found variation in colorectal cancer testing prevalence estimates across aspects of the health care delivery system including having a usual source of care (19.3% to 59.0%) and whether or not a physician recommendation to screen was reported (11.5% to 84.4%).

#### Colorectal cancer screening prevalence estimates

The colorectal cancer screening prevalence estimate for the NHIS study population was 45.0%. All within-group differences were statistically significant at or below \( P = 0.020 \) level. We found similar patterns to screening prevalence estimates that emerged in testing prevalence estimates, with colorectal cancer screening prevalence estimates for non-Hispanic whites [M (47.6%) and F (44.1%)] exceeding those of non-Hispanic blacks [M (41.0) and F (40.2%)] and screening prevalence estimates increasing with age (31.9% to 54.1%). Across measures of socioeconomic status, we saw estimates increase across levels of education (34.2% to 59.0%), poverty ratio (36.0% to 53.5%), and insurance coverage (15.9% to 47.2%). And again, we found variation in colorectal cancer testing prevalence estimates across aspects of the health care delivery system including having a usual source of care (12.5% to 48.1%) and whether or not a physician recommendation to screen was reported (9.0% to 68.7%).

#### Rate overestimation

The degree to which colorectal cancer screening prevalence estimates were overestimated when the impetus for testing was not taken into account varied, ranging from 16.8% (non-Hispanic black males) to 54.4% (those with no usual source of care). The median rate of prevalence overestimation was 22.6% and the mean rate of prevalence overestimation was 24.8%.

The colorectal cancer screening prevalence estimate for the full study population was overestimated by 23.3%. The estimate for non-Hispanic whites was overestimated more than for non-Hispanic blacks (23.6% and 18.5%, respectively). The estimate for females was overestimated more than males (26.6% and 19.0%), and we found that this is largely due to the high rate of overestimation in non-Hispanic white females. Compared with the other sex/race/ethnicity groups, their rate of overestimation (27.4%) exceeded others [non-Hispanic white male (19.3%), non-Hispanic black male (16.8%), and non-Hispanic black female (19.9%)]. The degree to which colorectal cancer screening prevalence rates are overestimated decreases with age (32.0% to 19.0%), education (31.3% to 17.8%), poverty ratio (32.5% to 17.4%), and insurance status (51.6% to 22.5%). Across measures of health care delivery, we saw overestimation rates higher in those without a usual source of care (54.4% vs. 22.7% without) and relatively flat rates of overestimation across categories of physician recommendation reports [without a recommendation (27.8%) and with a recommendation (22.9%)].

### Discussion

In this study, we found that ignoring the impetus for colorectal cancer testing resulted in pervasive overestimation of colorectal cancer screening prevalence estimates and patterns of overestimation that most often align with social and economic vulnerability. A notable exception to this patterning of overestimation is in non-Hispanic white females, who had the highest rate overestimation when compared with their peers.
Physician recommendation

Full study population 6,984 55.5 (54.0–56.9)

Ages 65–69 1,002 44.9 (41.7–48.2) <0.0001 45.7 (44.1–47.2) 0.005 23.6

Female 3,922 55.2 (53.4–57.0) 0.6812 46.9 (44.7–49.0) 0.020 19.0

Non-Hispanic white male 2,551 56.8 (54.5–59.1) 0.0001 47.6 (45.3–49.9) 0.0029 19.3

Non-Hispanic black male 511 47.9 (43.1–52.6) 0.0001 41.0 (36.2–46.0) 16.8

Non-Hispanic white female 3,176 56.2 (54.3–58.2) 0.0001 44.1 (42.1–46.1) 27.4

Non-Hispanic black female 746 48.2 (43.7–52.6) 0.0001 40.2 (35.9–44.6) 19.9

Ages 50–54 1,556 42.1 (39.1–45.1) <0.0001 31.9 (29.2–34.7) <0.0001 13.0

Ages 55–59 1,407 53.9 (51.2–56.7) 0.0001 43.6 (40.8–46.4) 12.3

Ages 60–64 1,262 59.7 (56.8–62.7) 0.0001 49.3 (46.2–52.4) 21.1

Ages 70–74 787 64.4 (60.7–68.1) — —

Ages 75–80 824 60.3 (58.8–63.7) — —

Ages 50–64 4,315 59.3 (57.6–61.0) 0.0001 — —

Ages 65–69 2,669 59.4 (57.6–61.0) 0.0001 — —

Less than high school diploma 1,002 44.9 (41.7–48.2) <0.0001 54.2 (50.9–57.7) <0.0001 31.3

High school degree or GED 2,162 52.0 (49.7–54.4) 0.0001 41.7 (39.4–44.4) 24.7

Some college no degree or associate degree 1,987 56.9 (54.4–59.5) 0.0001 46.5 (44.0–49.0) 22.4

Bachelor’s degree 1,039 59.4 (55.7–63.0) 0.0001 49.0 (45.7–52.4) 21.2

Master’s, professional, or doctorate degree 753 69.5 (66.0–73.1) 0.0001 59.0 (55.1–62.7) 17.8

Poverty ratio—low tertile 2,549 47.7 (45.5–49.9) <0.0001 36.0 (33.9–38.2) <0.0001 32.5

Poverty ratio—medium tertile 2,271 56.5 (54.0–58.7) 0.0001 46.1 (43.8–48.4) 22.6

Poverty ratio—high tertile 2,224 62.8 (60.5–65.0) 0.0001 53.5 (51.2–55.7) 17.4

Insurance—not covered 517 24.1 (19.3–28.9) <0.0001 15.9 (12.6–19.8) <0.0001 51.6

Insurance—covered 6,461 57.8 (56.4–59.3) 0.0001 47.2 (45.7–48.8) 22.5

Usual source of health care—no 470 19.3 (15.4–23.2) <0.0001 12.5 (9.7–15.9) <0.0001 54.4

Usual source of health care—yes 6,409 59.0 (57.6–60.4) 0.0001 48.1 (46.6–49.7) 22.7

Physician recommendation—no 2,534 11.5 (10.1–12.9) <0.0001 9.0 (7.9–10.4) <0.0001 27.8

Physician recommendation—yes 4,206 84.4 (83.1–85.6) 0.0001 68.7 (67.1–70.3) 22.9

Note: unweighted n, weighted % and CI of %.

aP values are based on Rao–Scott χ² tests and calculated for all variables with SAS except poverty ratio, where design-based Pearson F.

bFor groups with 2 categories, P values are based on t test and P > |t| calculated for overall effects with Stata; for groups with more than 2 categories, design-based Pearson F test was used with Stata.

Our results should give researchers and organizations (e.g., CDC, ACS, Agency for Healthcare Research Quality) pause as they consider the utilization of BRFSS and other survey data in cancer screening. Currently, BRFSS is unable to distinguish past diagnostic testing from screening, and analyses of BRFSS data that confound diagnostic testing and screening inform the CDC’s Morbidity and Mortality Weekly Reports as well as state-level data for Healthy People goals and benchmarks. These organizations’ documents and guidelines shape research, public health practice funding, and priorities. Therefore, accurate data are critical to ensuring that benchmarks for success and tailored interventions are relevant, timely, and effective. However, a review of the literature reveals that this limitation is not exclusive to use of BRFSS data, as researchers often underestimate the available data on the impetus for testing in NHIS. In our review, we found only three published articles on colorectal cancer screening using the additional NHIS data that do not confound screening and diagnostic testing (7–9).

These results suggest that closer examination of methodologies in surveillance research on cancer screening modalities that are used for both diagnostic testing and screening is warranted. The types of measurement error we found in screening prevalence estimates could be corrected easily at the survey level for future research by collecting and analyzing data on the impetus for colorectal cancer testing. Although this is an additional burden for survey methodologists and survey respondents, it is clear that ignoring the impetus for testing risks gross measurement error. This measurement error is not easily adjusted when reviewing past literature, as the rate of overestimation due to confounding varies greatly across subgroups and likely varies over time as modalities used for both diagnostic testing and screening have become more widely available.

The implications of this study extend far beyond surveillance research, as our understanding of disparities is also affected. We found that the degree of misestimation varies widely across sociodemographic groups. The most common disparity discussed in the colorectal cancer screening literature is between non-Hispanic white and non-Hispanic black. In the NHIS sample, using methods that fail to account for the impetus for testing, and therefore confound diagnostic testing and screening, resulted in a 37.9% overestimation of disparities between non-Hispanic white and non-Hispanic black (from a 17.7% non-Hispanic white advantage to a 12.8% non-Hispanic white advantage), whereas other sociodemographic disparities were mildly and grossly underestimated as well. Furthermore, behavioral research that confutes diagnostic testing and screening when specifying the outcome of interest risks inaccurate assessment of predictors of colorectal cancer screening.

This study is subject to several unavoidable limitations. Common to survey data, NHIS data are subject to recall bias on the part of the respondents for both dependent and independent variables.
independent variables. However, this is not likely to affect our prevalence estimates differentially, as there are no clear patterns in differences in the accuracy of self-report by age, sex, race, or family history (10). Another potential source of bias in this study is selection bias due to differential nonresponse among specific subgroups of individuals, which is addressed via nonresponse adjustments of the weights provided by NHIS to survey analysts (4). In addition, there may be differential bias in comprehension of survey questions by race, gender, and socioeconomic status; however, the survey questions of interest were follow-up questions to screening responses. These follow-up questions received very little attention, as the majority of the cognitive testing subjects (which numbered only 9 older than 40 years) indicated they did not undergo colorectal cancer testing (B. Taylor; personal communication).

Despite these limitations, our analyses demonstrated that surveillance research in colorectal cancer screening that ignores the impetus for testing risks measurement error in screening prevalence estimates. This measurement error hinders our ability to systematically improve population health. Ultimately, to reduce potential bias in surveillance research in colorectal cancer screening, we must look beyond "what" and "when" and include "why."

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

References

Authors’ Contributions
Conception and design: E.A. Becker, D.M. Griffith, B.T. West, N.K. Janz, K. Resnicow, A.M. Morris
Development of methodology: E.A. Becker, B.T. West, A.M. Morris
Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): E.A. Becker, D.M. Griffith, B.T. West
Writing, review, and/or revision of the manuscript: E.A. Becker, D.M. Griffith, B.T. West, N.K. Janz, K. Resnicow, A.M. Morris
Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): E.A. Becker
Study supervision: N.K. Janz, A.M. Morris

Disclaimer
The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the NCI.

Grant Support
E.A. Becker was partially supported by the University of Michigan Promoting Ethnic Diversity in Public Health Training Grant R25-GM058641-13 and Grant Number R25-CA057699 from the NCI.

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 April 6, 2015; revised September 2, 2015; accepted September 29, 2015; published OnlineFirst October 21, 2015.
Potential Biases Introduced by Conflating Screening and Diagnostic Testing in Colorectal Cancer Screening Surveillance

Elizabeth A. Becker, Derek M. Griffith, Brady T. West, et al.

Cancer Epidemiol Biomarkers Prev  Published OnlineFirst October 21, 2015.

Updated version  Access the most recent version of this article at: doi:10.1158/1055-9965.EPI-15-0359

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.