Methodological Issues in Colorectal Cancer Screening Research: Implications for Researchers, Practitioners and Policymakers

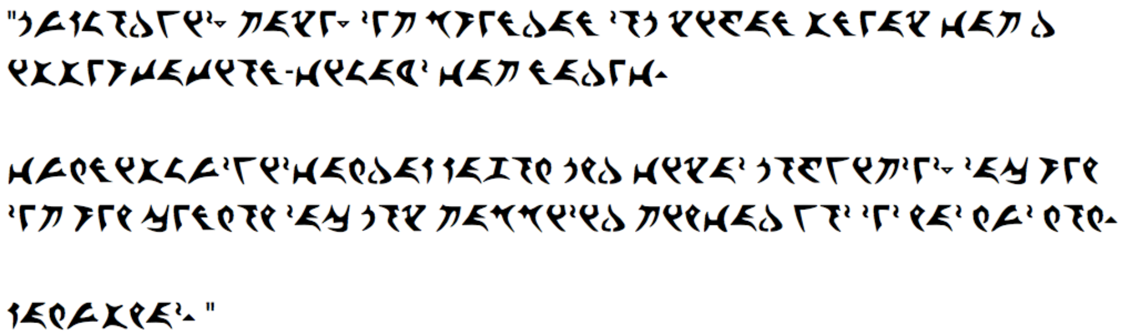
by

Elizabeth A. Becker

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Health Behavior & Health Education)
in the University of Michigan
2014

Doctoral Committee:

Associate Professor Derek M. Griffith, Vanderbilt University, Co-Chair
Associate Professor Arden M. Morris, Co-Chair
Professor Nancy K. Janz
Professor Kenneth A. Resnicow
Research Assistant Professor Brady T. West



Translation for non-Trekkies:

“For my sons, Brendan and Jake, I promised to sneak some Klingon into my dissertation - so here it is.

Sure, it's nice to have PhD after my name, but forever and always the most important letters to me make up the word 'Mom.'

I love you.”

Dedication

To DMG

For never allowing me to give up.
For never, ever giving up on me.

ACKNOWLEDGEMENTS

This dissertation would not be possible without the support of my committee. Derek, you have been calling me The Future Doctor Becker ("TFDB") longer than anyone. I am a better scholar, researcher, and future mentor for knowing you. Your dedication to my success is unmatched – thank you. Arden, your tenacity and scholarship have inspired me since 2010. I am so grateful for our chance meeting in Matrix Algebra that led to a long and fulfilling professional collaboration. You inspire me not only as a scholar but also as someone who never forgets her priorities and always supports the priorities of those around you. Brady, I am forever grateful that you enthusiastically accepted my invitation to serve on this committee. You have been a strong source of support for me. Nancy, you may never know how pivotal our conversation was back in 2010. I'm here because you were there – listening, empathizing, supporting, and guiding. Thank you. Your presence on my committee meant the world to me, not only for your expertise but also for your open door and open heart. Ken, I have learned an incredible amount from you, both inside and outside the classroom. I know my questions sometimes irked you, but you always hung in there and recognized their value. Thank you for your service on the committee and for bringing a unique, comprehensive lens to my work.

To my friends and colleagues at SPH, ICPSR, ISR, CRECH, CSCAR and Sweetland Writing Center – thank you. To my dear "work wife" Natalie – words can't express my

gratitude for our friendship. To Pete, Andre, Liz, Alana, Becca, Allison, Lori, Debbie, Amy, Katrina, Emily, Sarah, Jackie, Ann, Annie, Dana, Bill & Jorge and so many more, thank you for supporting me through thick and thin. To my dedicated advisors and mentors – Woody Neighbors and Linda Chatters – I couldn't have done this without you. And lastly, to Michelle, Felicia, Melissa, Saima, Carol and Scott – thank you.

To "Coach" -- Dr. Pat Eldred began as my writing professor at St. Kate's, and I'm now lucky to count her as a dear friend. She has been waving her pompoms from afar since my move to Michigan. I couldn't have done this without her cheering, coaching, and love. Thank you. And to my undergrad advisor and friend, Dr. Anita Ho, for never, ever telling me I couldn't do what I set out to do.

Four teachers from long ago deserve mention here. In sixth and seventh grade, Sandra Fedyniak and Elizabeth Usher artfully allowed the arts and learning to coexist in their classrooms (and allowed me to impersonate Elvis several times). In high school, Herr Fahnert always encouraged my learning of other languages and cultures, resulting in two terms abroad when I was a teenager. Finally, Jane Blackie demystified public speaking and encouraged students, particularly girls, to express themselves in the classroom. Each, in his or her own way, helped me grow into the scholar that I am today.

I am truly grateful for the financial support of the Center for Research on Ethnicity, Culture & Health, the Rackham Merit Fellowship, the Rackham Graduate Student Emergency Fund, the Center for the Education of Women, and the Interuniversity Consortium for Political and Social Research.

And finally, above all, I want to thank my family. In their own unique ways, my four siblings, Kevin, Debbie, Rebecca and David, and their families have also fostered that love of questions. Though they tortured me as their youngest sibling, they are also strong cheerleaders. They have been with me through the ups and downs of my personal, professional and academic lives, and for that I am grateful.

My parents, Dick and Charline Watts, have supported me from the moment I was born, and especially since the questions about the world started coming from my mouth. When I was very young, my father nicknamed me The Question Girl. If I ever felt like I didn't have an audience for all of those questions growing up, I'm guessing my educational aspirations and achievement would've been very different. Somehow they have never tired of the guidance, encouragement, unconditional love, care packages, daddy tucking-ins, FaceTime pity parties, FaceTime cheering sessions, and other beautiful shows of their affection. I know that I was labeled as "disruptive in class" on elementary school report cards, but the tenacity and persistent questioning that annoyed my teachers is what got me to this point. Thank you for not squashing it!

And saving the best for last, my husband, Geoff, has remained committed to my goals and dreams and has been the fiercest cheerleader one could ever want or need. My children, Brendan and Jake, were 9 and 6 years old when I started my undergraduate degree. Three degrees later, they're now 22 and 19. They have supported their mama through the giant peaks and deep valleys of this journey. I love them more than anything.

And to all who have supported me but whom I have missed naming, ***thank you***.

TABLE OF CONTENTS

List of Tables

List of Figures

ABSTRACT

This dissertation explores three distinct, but interrelated, areas of research that provide important additions to the current colorectal cancer (CRC) screening literature. The first study reveals the problematic nature of the common practice of ignoring the impetus for CRC testing when specifying outcomes in survey design and analysis. The results show significant overestimation of true screening rates, high rates of diagnostic testing and disproportionate rates of screening among those with less socioeconomic privilege when compared with those with more socioeconomic resources. The second study finds that survey measurements often mistake diagnostic testing referrals for true screening recommendations, resulting in significant misperception about who is likely to receive a screening recommendation. Finally, the third study showed that although individuals with multiple chronic conditions are more likely to engage in diagnostic testing, their overall rates of true screening are higher than the general population.

Based on our findings, we recommend immediate changes to survey methodology, specifically collecting data that allow researchers to clearly differentiate diagnostic testing and true screening. Furthermore, behavioral researchers need to reanalyze additional NHIS data to determine the extent to which diagnostic testing has misestimated improvements in CRC screening rates across time. Researchers exploring

psychosocial and instrumental barriers and facilitators of CRC screening need to actively engage with individuals from many sociodemographic groups. Our results show troubling variation in screening uptake and recommendations across groups. We need to understand the source of this variation in order to design and target interventions appropriately. We further recommend that policymakers reconsider their approach to federal screening goals. Specifically, a one-size-fits-all goal that averages very high and very low rates across sociodemographic groups fails to acknowledge the challenges we face in improving public health across levels of privilege. Significant gaps in screening rates will continue to persist if resources are not allocated to addressing these disparities. Finally, our results show we must continue to explore contextual factors such as MCC as we seek to identify influential factors at multiple levels that influence the recommendation and delivery of CRC screening services.

CHAPTER ONE

Introduction: What are we missing?

Whether we are undergraduate students studying a public health challenge for the first time, doctoral students crafting their dissertation research agenda, or well-established researchers, we all strive to do great research. Our shared goal is to reduce suffering, to improve quality of life, and to broaden our level of understanding of the complex challenges that we face in the field of public health. Sometimes our goals are met by following established convention in the approach that we take, whether we are doing research, practice or policy work. Other times, we have to think outside the proverbial box to discover something that we might be missing.

Despite decades of research on strategies to reduce morbidity and mortality from colorectal cancer (CRC), the disease remains the third most commonly diagnosed cancer and third leading cause of cancer-related death in the United States (American Cancer Society, 2011). Compared to other tumor-related cancers, CRC is unique in that it is nearly completely preventable with the detection and removal of pre-cancerous polyps through screening. Unlike other cancer screenings that can involve patient-directed behaviors at home such as self-exams of the breast, testicles, and skin, all types of CRC screening require active participation of both patient and their physician.

Physician referrals are necessary to initiate tests ranging from non-invasive fecal occult blood tests (FOBT) to invasive tests such as colonoscopies and sigmoidoscopies. Over ten percent of cancer deaths in this country are attributable to CRC. This statistic reflects the troubling patterns of low population screening rates and low rates of physician recommendations to screen, resulting in later, less survivable stage at diagnosis (American Cancer Society, 2010; Coughlin & Thompson, 2005). Considering the tools at our disposal to prevent CRC morbidity and mortality, the degree to which the disease impacts population health is simply unacceptable, and we need novel approaches to understanding

a) patterns in screening and recommendation rates;
b) factors that contribute to patients' persistent underutilization of CRC screening and
c) factors that contribute to physicians' lack of communication of CRC screening recommendations to their patients.

This introduction will provide an overview of features of CRC screening, my entry into this line of research and the research questions explored throughout the dissertation.

Most clinical and advocacy organizations recommend initiation of CRC screening in average risk adults at age 50, with earlier screening recommended for individuals with gastrointestinal comorbidities or a family history of CRC (American Cancer Society, 2011; US Preventive Services Task Force, 2008). This age group is growing: by the end of the year 2015, an estimated one in five Americans will be ages 50-64 (Holden, Jonas, Porterfield, Reuland, & Harris, 2010). Accurate assessment of behavioral trends and predictors will be crucial to effective design of interventions to engage this population and their physicians in CRC screening.

Over the past decades, researchers have developed methods to understand CRC screening in many ways including analysis of large-scale surveys, electronic health record reviews, state-level surveys, and more (Breen, Wagener, Brown, Davis, & Ballard-Barbash, 2001; Centers for Disease Control, 2010; Doubeni et al., 2010; Guerra, Katrina Armstrong MD, & Brown, 2007). This research is mainly in two areas of focus: first, on patients and their engagement with CRC screening. Researchers have assessed rates of screening, sociodemographic and psychosocial predictors of screening, trends over time and disparities across groups. Second is a focus on physicians and patterns and predictors of their behaviors with their patients. Here, too, researcher have looked at rates of recommendations, predictors of the physician's recommendation patterns (both in terms of demographic characteristics of the physician and his or her patient), trends over time and disparities in who reports the receipt of a recommendation to screen for CRC from their physician or other health care provider.

Overall, we have seen some improvements in CRC screening rates at both the patient and physician level over time, but both rates remain far below federal goals set by Healthy People 2010 and Healthy People 2020 (US Department of Health & Human Services, 2013). Furthermore, there is evidence of disparities existing between males and females, across race/ethnicity, age, education, and income categories and between states and wider geographic areas (Ahmed, Pelletier, Winter, & Albatineh, 2013; Ananthakrishnan, Schellhase, Sparapani, Laud, & Neuner, 2007; Cole, Jackson, & Doescher, 2012; James, 2006).

My dissertation research seeks to answer three important sets of questions that will be valuable additions to the current literature on CRC screening. The research aims were driven in large part by my involvement with a qualitative study led by Dr. Arden Morris examining treatment decisions of CRC patients. Although the primary aims of the study were about decisions related to treatment of their disease and not directly related to CRC screening, I noticed three themes while analyzing the interview and focus group data:

- participants were diagnosed most often by diagnostic testing and not pre-symptomatic screening;
- participants' discussions with their physician were related to referrals for diagnostic testing and not recommendations for pre-symptomatic screening; and
- acute and multiple chronic conditions appear to be associated with delay or omission of CRC screening.

Answering the specific questions that arise from each theme will complement the existing body of literature on CRC screening, and will have unique implications for researchers, practitioners, and policymakers. For researchers, it will be important to answer these questions so that we know whether we are measuring what we have intended to measure. This applies not only to rates of screening, but also to trends in screening over time, disparities in screening, and predictors of screening. For practitioners, answering these questions will clarify the impact of conflation of screening versus diagnostic testing, which has the potential to interfere with accurate measurement of intervention efficacy. Specifically, our study will reveal whether simply counting procedures (colonoscopy) provides adequate assessment of the success or failure of an intervention or whether we must also ask why the procedure was ordered.

And finally, policymakers will gain evidence supporting or refuting longstanding perceptions of the gap between rates of screening and established Healthy People goals. Answering these questions will also enable policymakers to better understand disparities in screening, which is essential to effective program planning at federal and state levels.

Are We Really Measuring Screening?

The first theme from the preliminary qualitative data was that most patients delayed screening and few were diagnosed through testing at an asymptomatic stage, what I will call "true screening". Instead, over 70% of the patients were diagnosed as the result of post-symptomatic testing, what I will call "diagnostic testing" (Becker, E., Elliott, H., Griffith, D., Alexander, G., & Morris, A., 2010). I began to wonder how cancer researchers specified behavioral outcomes. If they did not distinguish the two behaviors, respondents like these study participants engaging in post-symptomatic testing will be categorized as true screeners, which was certainly not reflective of their behaviors, nor is it consistent with the behavioral outcome that will improve public health.

As I began to explore the cancer screening literature and CRC screening data sources in particular to see if these questions had been addressed, I found that there was persistent conflation of diagnostic testing and true screening behaviors. In fact, I found only three studies that did not conflate the two behaviors. As I reviewed the survey protocol, I found that the National Health Interview Survey inquired about the impetus for testing, leaving me wondering why so few survey analysts used the available

data on impetus for testing in their research analyses. And perhaps the most often cited resource for cancer screening data, the Behavioral Risk Factor Surveillance Survey (BRFSS), did not even ask respondents why they engaged in CRC testing or why their physician discussed it with them. In fact, the conflation of diagnostic testing and true screening was a consistently cited limitation across studies using BRFSS data. Furthermore, I found no examination of the potential impact of conflation of these two behaviors when examining published screening rates.

Considering the paucity of available information on these issues, the first research area that I explore in this dissertation is patient-focused; namely, in Chapter 2, I explore whether or not researchers have been overestimating population rates of screening by conflating diagnostic testing and true screening in cancer screening research.

Are We Really Measuring Physician Recommendations to Screen?

This theme of diagnostic testing versus true screening is highly relevant for the second theme that emerged from the qualitative study, the lack of clarity around the physician recommendation for colonoscopy for colorectal cancer testing. The study participants often spoke of the interactions with their physicians and how that impacted their engagement (or lack thereof) with CRC testing. In addition, many study participants reported scenarios in which diagnostic testing was the clear reason for the referral, rather than a recommendation for true screening. Improving public health mandates high rates of recommendations for true screening, whereas high rates of diagnostic testing indicate that not enough patients are being screened.

The second research area I explore this dissertation is physician-focused; namely, in Chapter Three, I seek to answer the question, have we been overestimating rates of physician screening recommendations by conflating referrals for diagnostic testing with recommendations for true screening in our surveys and analyses of electronic health records?

As mentioned earlier, CRC screening is a behavior that requires action by both patient and physician. Prior research has consistently demonstrated that a top predictor of patients' engagement with screening is a recommendation from their physician (Coughlin & Thompson, 2005; Klabunde, Breen, & Meissner, 2005). The implications of these questions are important to researchers as they seek to accurately assess rates, predictors and trends of physician CRC screening behaviors. Practitioners can use this evidence to determine if healthcare quality benchmarks in cancer screening recommendations are met. However, conflating referrals for diagnostic testing and recommendations for true screening may give a false impression of recommendation rates and may misrepresent success of reaching these benchmarks. Finally, policymakers will find these analyses useful as they develop goals for physician recommendation rates, such as the goals now in the developmental phase for Healthy People 2020 (US Department of Health & Human Services, 2013).

Exploring the Influence of MCC on CRC Screening and Recommendations to Screen

In the preliminary qualitative study , another important discovery was the third theme explored in this dissertation: the impact of acute and multiple chronic conditions (MCC) on participants' delays in engaging in true screening and instead engaging in post-

symptomatic diagnostic testing. While study participants with acute health challenges (e.g. knee replacement) described delays in screening, respondents with multiple chronic conditions, defined as two or more co-occurring chronic conditions (e.g. diabetes and kidney disease), described other barriers such as physiological symptoms (e.g. chronic pain) that caused them to be apprehensive about going through another medical test unrelated to their conditions. Furthermore, patients with MCC described feeling fearful of yet another diagnosis, and therefore avoiding cancer screening altogether (Becker, Elliott, Griffith, Alexander, & Morris, 2010). The third research area explored in this dissertation is focused on the role of multiple chronic conditions in screening and recommendations for CRC screening; namely, in Chapter 4, I explore the association between MCC and screening and recommendations to screen.

To date, much of the CRC screening literature has focused on sociodemographic and psychosocial characteristics and their influence on screening behaviors of study participants and their physicians. This is helpful as we consider differences between genders, race/ethnicity groups, and socioeconomic groups, but little attention has been paid to other factors such as MCC that create context at both individual and institutional levels. At the individual level, patients with MCC face increasing physical, emotional, and financial burdens that may reduce an individual's likelihood to engage in screening. At the institutional level, MCC can influence the priorities of care. Instead of preventive care, physicians of patients with MCC may instead focus on more serious illnesses, conditions that require immediate attention, disease management and flares. Finally, MCC may explain some of the sociodemographic disparities in CRC screening literature.

More than 25% of Americans live with MCC, but MCC is unequally distributed across racial and socioeconomic groups, with greater burden in those with less social and economic privilege. If MCC interferes with CRC screening, it may account for part of the screening disparities between groups as well as the predictive value of those sociodemographic categories in multivariate models.

Answering the questions of MCC's association with screening and physician recommendations will be critical to enriching the existing literature that is largely focused on sociodemographic and psychosocial predictors of screening. To my knowledge, this is the first study to approach MCC as a predictor of CRC screening using a nationally representative dataset.

In summary, in an effort to add valuable information to the conversation on CRC screening, this dissertation explores the following research questions:

> **Chapter 2 – Is this really screening?**: Are rates of screening behaviors overestimated by conflation of diagnostic testing and true screening? If so, does the degree of overestimation vary by sociodemographic group? How do predictors of screening change when the outcome is correctly specified and does not include diagnostic testing?
>
> **Chapter 3 – Is this really a physician recommendation for screening?**: Are rates of screening recommendation behaviors overestimated by conflating referrals for diagnostic testing and recommendations for true screening? If so, does the degree of overestimation vary by sociodemographic group? How do predictors of screening recommendations change when the outcome is correctly specified and does not include referrals for diagnostic testing?
>
> **Chapter 4 – The influence of MCC on screening:** Is the presence of multiple chronic conditions associated with patients' and physicians' CRC screening behaviors? How does accounting for MCC in predictive models impact the associations between sociodemographic characteristics and the outcomes of interest?

This research will add valuable information to the CRC screening literature and assist researchers, practitioners and policymakers as we seek to reduce the burden of CRC morbidity and mortality.

References

Ahmed, N. U., Pelletier, V., Winter, K., & Albatineh, A. N. (2013). Factors Explaining Racial/Ethnic Disparities in Rates of Physician Recommendation for Colorectal Cancer Screening. *American Journal of Public Health*, *103*(7), e91–e99.

American Cancer Society. (2010). Colorectal cancer and polyps. Retrieved March 21, 2005, from

American Cancer Society. (2011). Colorectal Cancer Facts & Figures 2011-2013. Retrieved from http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-028323.pdf

Ananthakrishnan, A., Schellhase, K., Sparapani, R., Laud, P., & Neuner, J. (2007). Disparities in colon cancer screening in the Medicare population. *Archives of Internal Medicine*, *167*, 258.

Becker, E., Elliott, H., Griffith, D., Alexander, G., & Morris, A. (2010). Qualitative Research and Racial Disparities in Screening Behaviors: Understanding Barriers and Facilitators of Colorectal Cancer Screening. Presented at the Academy Health Annual Research Meeting, Boston, MA.

Breen, N., Wagener, D., Brown, M., Davis, W., & Ballard-Barbash, R. (2001). Progress in cancer screening over a decade: results of cancer screening from the 1987, 1992, and 1998 National Health Interview Surveys. *JNCI Journal of the National Cancer Institute*, *93*, 1704.

Centers for Disease Control. (2010). Vital signs: colorectal cancer screening among adults aged 50-75 years-United States, 2008. *MMWR. Morbidity and Mortality Weekly Report*, *59*(26), 808.

Cole, A. M., Jackson, J. E., & Doescher, M. (2012). Urban–rural disparities in colorectal cancer screening: cross-sectional analysis of 1998–2005 data from the Centers for Disease Control's Behavioral Risk Factor Surveillance Study. *Cancer Medicine*, *1*(3), 350–356.

Coughlin, S., & Thompson, T. (2005). Physician recommendation for colorectal cancer screening by race, ethnicity, and health insurance status among men and women in the United States, 2000. *Health Promotion Practice*, *6*, 369.

Doubeni, C. A., Laiyemo, A. O., Klabunde, C. N., Young, A. C., Field, T. S., & Fletcher, R. H. (2010). Racial and ethnic trends of colorectal cancer screening among Medicare enrollees. *American Journal of Preventive Medicine*, *38*(2), 184–191.

Guerra, C. E., Katrina Armstrong MD, M., & Brown, J. S. (2007). Barriers of and facilitators to physician recommendation of colorectal cancer screening. *Journal of General Internal Medicine*, *22*(12), 1681–1688.

Holden, D. J., Jonas, D. E., Porterfield, D. S., Reuland, D., & Harris, R. (2010). Systematic review: enhancing the use and quality of colorectal cancer screening. *Annals of Internal Medicine*, *152*, 668.

James, T. M. (2006). Disparities in colorectal cancer screening: a guideline-based analysis of adherence. *Ethnicity & Disease*, *16*, 228.

Klabunde, C., Breen, N., & Meissner, H. (2005). Use of Colonoscopy for Colorectal Cancer Screening. *Cancer Epidemiology Biomarkers & Prevention*, *14*, 2279.

US Department of Health & Human Services. (2013, March 13). Leading Health Indicators - Healthy People 2020. Retrieved from http://www.healthypeople.gov/2020/LHI/clinicalPreventive.aspx

US Preventive Services Task Force. (2008, October). Screening for Colorectal Cancer. *Screening for Colorectal Cancer*. Retrieved April 16, 2014, from http://www.uspreventiveservicestaskforce.org/uspstf/uspscolo.htm

CHAPTER TWO

Methodological Issues in Colorectal Cancer Research: The Dangers of Conflating Diagnostic Testing and True Screening

Colorectal cancer (CRC) is the third leading cause of cancer-related death in the United States. It is estimated that 143,460 new cases of the disease and 51,690 deaths (~10% of all cancer deaths) will be attributable to CRC in 2012 (American Cancer Society, 2012). Colorectal cancer mortality is highly correlated with stage at diagnosis, with five-year survival rates ranging from 90% for local-stage disease to 68% for regional-stage disease and only 11% for distant-stage disease (American Cancer Society, 2010; Doubeni et al., 2006; Emmons et al., 2008). Unlike other tumor cancers, CRC is largely preventable with early detection and removal of pre-cancerous polyps (American Cancer Society, 2012). Research suggests that nine of ten deaths from CRC could be prevented with early detection through use of one or more screening modalities (Subramanian, Klosterman, Amonkar, & Hunt, 2004). By definition, cancer screening occurs at a stage of disease at which no symptoms are present (we will refer to this as 'true screening') (National Cancer Institute, 2013). Tests for disease after symptoms appear are diagnostic in nature, and are associated with diagnosis at a later stage of disease and a greater morbidity and mortality burden. In order to improve public health

by reducing CRC morbidity and mortality, we want high rates of true screening and low rates of diagnostic testing.

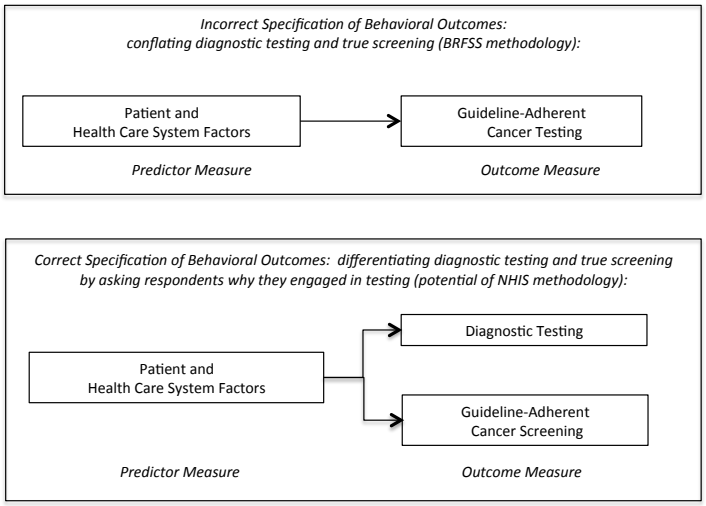
Historically, methods to measure rates of screening have focused on self-reported data from national surveys. National organizations including the Centers for Disease Control and Prevention (CDC), the American Cancer Society (ACS), the Agency for Healthcare Research Quality (AHRQ), and Healthy People 2020 rely on these surveys to provide comprehensive data on cancer screening rates and trends. These surveys ask respondents about their engagement with screening for cancers of the colon/rectum, cervix, breast or prostate including if their physician or other health care provider recommended screening.

A distinguishing feature of two of the most cited national surveys, the Behavioral Risk Factor Surveillance Survey (BRFSS) and the National Health Interview Survey (NHIS), is their varying ability to differentiate true screening from diagnostic testing. Survey researchers utilizing BRFSS data often cite an inability to distinguish between these behaviors as a limitation to their studies (Centers for Disease Control, 2010; D. A. Joseph, King, Miller, & Richardson, 2012; Soneji, Armstrong, & Asch, 2012). BRFSS does not ask respondents why they engaged in testing. In contrast to BRFSS, NHIS affords researchers the opportunity to distinguish between the two by asking respondents not only *what* test they had and *when*, but *why* they had it. NHIS methodology allows us to better differentiate screening from diagnostic testing, as depicted in **Figure 1**.

A review of the cancer screening literature reveals a potential problem in interpreting the results of cancer screening studies due to the irregular measurement of

screening. Specifically, methodological choices in survey design and/or analysis often result in a conflation of true (pre-symptomatic) screening with diagnostic testing when specifying the behavioral outcome of interest.

Figure 2.1. Methodological Differences Resulting in Incorrect and Correct Specification of Behavioral Outcomes



Empirical research has explored several facets of screening including predictors of screening, rates of screening across sociodemographic and geographic characteristics, trends in rates over time and disparities in screening. In addition to these areas, intervention evaluation and health services research focusing on adherence to a physician recommendation to screen are also subject to survey design and analysis decisions. Finally, we need to be cautious when comparing and contrasting rates from behavioral studies cited in meta-analyses and review articles when the behavioral outcome is specified differently across studies (see Pruitt et al. (2009) and Beydoun & Beydoun (2008)). In addition to publishing inaccurate screening data by incorrectly estimating rates of these outcomes by conflating diagnostic testing and screening, multivariate models from these and other studies are likely misestimating a) predictors

of screening; and b) the degree to which screening rates are improving.

Even when the data are available (as in NHIS), researchers often choose to conflate diagnostic testing and true screening. Coups and colleagues (2007) utilize NHIS data in their analysis of behavioral risk factors for screening, but fail to utilize available data to account for the impetus for testing when defining their outcome variable. James et al. (2006) utilized NHIS data to examine disparities in screening but failed to utilize available data on the impetus for testing, noting,

> "Both diagnostic and screening exams were included in adherence calculations to create the most lenient definition of up-to-date with CRC screening recommendations and to be consistent with Healthy People 2010 goal measurement methods."

There are very few examples of research that accurately specifies the outcome of interest. Of many studies utilizing NHIS data, we found three that do not conflate diagnostic testing and true screening. In their work examining screening rate trends across time, Breen and colleagues (2001) specify their analyses across three cycles of NHIS (spanning eleven years) to tests reported as part of a routine exam. Meissner and colleagues (2006) exclude respondents who reported CRC testing for diagnostic purposes, while Subramanian and colleagues (2004) reported separate rates for any type of testing and true screening, but did not measure the impact of conflating the two.

The majority of studies that fail to account for the impetus for testing are important because they are representative of a broader issue. In CRC screening literature, researchers often misspecify behavioral outcomes in a way that fails to recognize the most fundamental characteristic of screening: namely, that cancer

screening occurs at the pre-symptomatic stage of disease.

The goals of this study are to understand a) the amount of error that may exist in screening rates by conflating diagnostic testing and true screening; b) whether the degree of error in screening rates varies across sociodemographic categories and c) how predictive models vary when the outcome is correctly specified. We explore these questions through an analysis of colorectal cancer screening behaviors reported in the 2008 NHIS and Cancer Control and Sun Protection Supplement (CCSPS).

Methods

Data source. We examined cross-sectional survey data from the 2008 NHIS and CSSPS. Although there are other options for collecting preventative care data (including other nationally representative datasets and electronic health record data), we are using NHIS for the specific aims and broader goals of this study for several reasons. First, it allows us to consider multiple factors simultaneously as we consider influences on screening recommendation and screening behaviors. Second, NHIS is the only nationally representative dataset that allows us to capture data on the impetus for testing, a critical component of our analyses. And finally, due to its frequent use in survey research, we can compare and contrast results from this study with a large body of literature in CRC and other cancer screening behaviors. Since its inception in 1957, the NHIS has been the principal source of information on the health and health behaviors of civilian, non-institutionalized households in the United States. The survey 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 (Centers for Disease Control and Prevention, 2012). The 2008 Cancer Screening and Sun Protection Supplement (CSSPS) was sponsored by the National Cancer Institute (NCI) and includes questions on physician recommendations for screening, screening behaviors, and reasons for screening tests. The 2008 NHIS/CSSPS was selected over the more recent 2010 NHIS due to the fact that questions on physician recommendations for screening were linked to respondents' screening behaviors rather than in a more general manner. This allowed us to identify the impetus for the physician recommendation as well as CRC testing.

The NHIS sampling strategies result in an oversample of self-identified Black, Hispanic and Asian Americans. Weights constructed for the NHIS respondents reflect the resulting unequal probabilities of selection and also incorporate adjustments for non-response and post-stratification procedures designed to align survey estimates with population distributions from the 2000 Census. The application of these weights in secondary analyses of the NHIS data results in estimates that are representative of the adult civilian non-institutionalized population of the U.S. (Inter-University Consortium for Political and Social Research, 2012). The annual NHIS response rate averages close to ninety percent of the eligible households in the sample (Centers for Disease Control and Prevention, 2012).

Study population. Our study population included non-Hispanic White and non-Hispanic Black respondents ages 50-80 with no history of colorectal cancer. We chose these two groups because Non-Hispanic Whites and Non-Hispanic Blacks suffer the highest rates of cancers of the colon and rectum in the United States (American Cancer

Society, 2011) and are the most common population groups studied in CRC screening literature (Ananthakrishnan, Schellhase, Sparapani, Laud, & Neuner, 2007; Bellizzi, Breslau, Burness, & Waldron, 2011; Breen, Wagener, Brown, Davis, & Ballard-Barbash, 2001; Centers for Disease Control and Prevention (CDC), 2001; James, 2006; Klabunde et al., 2005). Individuals with a history of colorectal cancer were excluded since colonoscopic testing is used as disease surveillance and not preventative care. We defined a subpopulation for our analysis of true screening to individuals who reported ACS guideline-adherent screening since the question regarding impetus for testing was limited to those respondents.

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

Measures

Dependent variables.

Any ACS guideline adherent testing. To assess ACS guideline adherence, 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 assess engagement with non-invasive testing, respondents were asked, "The following questions are about the blood stool or occult blood test, a test to determine whether you have blood in your stool or bowel movement. The blood stool test can be done at home using a kit. You use a stick or brush to obtain a small amount of stool at home and send it back to the doctor or lab. Have you EVER HAD a blood stool test, using a HOME test kit?" Responses of "yes" or "no" were coded as such, while responses of "refused"

or “don’t know” were coded as missing data and excluded from analysis. However, in the case of respondents whose response was coded as missing for non-invasive testing but “yes” or “no” for invasive testing (see following paragraph), missing data from the non-invasive testing question were ignored[1] (missing n = 52). Respondents reporting testing also reported testing timelines in one of the following 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 testing was reported within one year of survey administration, 0 if no testing or FOBT testing was earlier than one year from survey administration, and missing otherwise (see note above regarding handling of missing data). All FOBT testing was coded as true screening, as it is used exclusively for that purpose.

To assess engagement with invasive testing, respondents were asked, “Have you ever had a sigmoidoscopy, colonoscopy, or proctoscopy? These are exams in which a health care professional inserts a tube into the rectum to look for signs of cancer or other problems.” Responses of “yes” or “no” were coded as such, while responses of “refused” or “don’t know” were coded as missing data and excluded from analysis. However, in the case of respondents whose response was coded as missing for invasive testing but “yes” or “no” for non-invasive testing (see preceding paragraph), missing

[1] As an example, a respondent did not provide an answer to the question concerning whether they had engaged in FOBT testing but indicated that ACS guideline adherent testing was done via colonoscopy. This respondent’s testing status was coded as ‘yes’ despite the missing data for the FOBT question.

data from the invasive testing question were ignored[1] (missing n = 23).

Respondents reporting testing also reported testing timelines in one of the following 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. Invasive testing (sigmoidoscopy, proctoscopy, colonoscopy) was coded as ACS guideline adherent if the testing was reported to be within five, five and ten years, respectively. One dependent variable was created for each mode of testing, coded as 1 if ACS guideline adherent, 0 if no testing or non-adherent, and missing otherwise.

Data from these four variables (FOBT and three invasive testing variables) informed the combined dependent variable measuring ACS guideline adherence. If any mode of testing was ACS guideline adherent, the variable was coded as 1, if no mode of testing was ACS guideline adherent, the variable was coded as 0, and missing otherwise. Alone, this variable reflects the methodology of BRFSS as it conflates diagnostic testing and true screening.

True screening. To measure reports of true screening versus diagnostic 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. In order to determine the impetus for the reported testing for 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 'true 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 non-invasive testing was not a "yes" (missing n = 4).

If respondents reported both non-invasive and invasive modes of ACS guideline adherent testing, their answer to the question above with respect to the impetus for the invasive tests determined whether or not it was 'true screening' or 'diagnostic testing.' A total of 559 respondents fit these criteria: 131 were coded as diagnostic and 428 were coded as true screeners.

These processes resulted in two dependent variables: CREHADACS, measuring any ACS guideline adherent testing (1-yes, 0-no, missing), and TRUSCRACS, measuring ACS guideline adherent true screening (1-yes (true screening) 2-no (diagnostic testing), missing).

Demographic and control variables.

We utilized 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 doctorate), poverty ratio (tertiles (measured as ratio of family income to poverty threshold): low (under .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. Missing values for all variables except poverty ratio were below 3%, a threshold determined to be acceptable with this sample size.

Statistical Analysis

All data analyses for this chapter were performed with the SAS/STAT statistical software, version 9.3_M1 of the SAS System for Windows, Copyright © 2002-2010 SAS Institute Inc. SAS employs procedures capable of computing appropriate variance estimates for survey estimates generated from analyses of complex sample survey data sets such as the 2008 NHIS (Heeringa et al., 2010; SAS Institute Inc., 2014). Due to the fact that SAS estimates variance in a way that excludes strata with only one primary sampling unit (PSU) from its variance estimates, we ran selected models in Stata 13.0. Stata enables multiple ad-hoc variance estimation methods for dealing with "singleton" PSUs to compare estimated standard errors between the two program's procedures (StataCorp, 2013). For the logistic regression models including main effects, of 268 estimates only ten had greater than or equal to a 10% difference in standard errors, with the greatest difference at 17%. In most cases, Stata produced higher standard errors; none of these differences resulted in changes in statistical inferences.

We began our analysis with computation of weighted and unweighted frequencies of all testing (including diagnostic testing and true screening[2]) for each of

[2] 'True screening' refers to pre-symptomatic screening (and not diagnostic testing).

the independent variables of interest in the full subpopulation as defined above. We then obtained weighted and unweighted frequencies of true screening for each of the independent variables of interest in the smaller subpopulation of ACS guideline-adherent screeners. We calculated estimates of true screening rates for the full study population by applying the rates of true recommendations from the smaller subpopulation to the unadjusted[3] testing rates of the full study population. For example, the unadjusted testing rate for the full study population was 55.46%. In the smaller subpopulation of respondents who engaged in guideline adherent CRC testing, the true screening rate was 74.12%. We applied this rate to the unadjusted testing rate (55.46 x 74.12%) to estimate an adjusted true screening rate for the full population of 41.10%. This same process was repeated for each subgroup of the full study population. Since this calculation reflects the diagnostic testing rates of only respondents reporting ACS guideline adherent testing, the adjusted true screening rate for the full population is likely a conservative estimate. We calculated estimates of the confidence intervals and resulting tests of significance for the adjusted true screening rate of the full population. In this process, we were unable to account for the fact that we derived the true screening rate for the full population from two estimates. This resulted in narrower confidence intervals and inferences that were more likely to result in statistically significant values. However, given the precise nature of the subpopulation rate

[3] 'Unadjusted testing rate' refers to a rate that conflates diagnostic testing and true screening (and fails to account for the impetus for testing, and is comparable to BRFSS methodology), while 'adjusted true screening rate' refers to a rate that properly accounts for the impetus for testing (and excludes diagnostic testing, utilizing NHIS methodology).

estimate, this issue is likely negligible. Finally, we calculated the overestimation of true screening recommendation rate by dividing the difference between the unadjusted and adjusted rates by the adjusted rate. For example, the unadjusted testing rate for the general population is 55.46% and the adjusted true screening rate for the same population is 45.06%. We calculated the overestimation of the true screening rate as (.5546-.4506)/.4506 = 23.08%.

In preparation for fitting multivariate logistic regression models to the dependent variables of interest, we examined weighted bivariate associations between independent and dependent variables for each aim using the Rao-Scott chi-square test to account for the features of the complex NHIS sampling design. Predictors that had an association at $p < 0.25$ were included in the initial multivariate logistic regression model. In models including two-way interaction terms, both variables defining the interactions will be included in each model, irrespective of bivariate associations (Hosmer & Lemeshow, 2000).

In the second stage of analysis, we performed a four-step fitting of weighted multivariate logistic regression models to the odds of reporting true screening, adjusting for sociodemographic factors including sex, race, age, education, poverty ratio, health insurance status, and usual source of health care (Breen et al., 2001; O'Malley, Forrest, Feng, & Mandelblatt, 2005; Rosen & Schneider, 2004). First, we fit a main effects model with all covariates. Second, we fit an interaction model with demographic interactions. Third, we fit an interaction model with socioeconomic interactions. Finally, we fit a full model with statistically significant main effects and interaction terms. We estimated

the regression parameters using pseudo-maximum likelihood estimation, due to the complex sampling nature of the NHIS (Heeringa, West, & Berglund, 2010), and we used Taylor Series Linearization to compute design-based estimates of standard errors for the estimated regression parameters. Hypothesis tests were performed using design-adjusted Wald X^2 tests. We report estimated odds ratios and their 95% confidence intervals.

Results

Please refer to **Table 2.1** for estimated demographic characteristics for the full study population and subpopulation. Distribution of NHB in the full population was 17% higher than in the subpopulation, and the full population was weighted more heavily to the younger end of the age range, with 22.18% ages 50-54 versus 16.92% in the subpopulation. Overall, the full population is distributed more heavily in the lower socioeconomic categories. The full population has 24% more respondents in the lowest level of education (13.97% versus 11.27%) and 20% less in the highest (11.45% versus 14.36%), and across poverty ratio tertiles, the full population has 17% more respondents in the low tertile than the subpopulation, and 12% fewer in the highest tertile.

Weighted frequencies. Table 2.2 displays estimated rates of true screening and diagnostic testing for the subpopulation of respondents who reported ACS guideline adherent testing. **Table 2.3** displays estimated rates of true screening received by the full study population, with unadjusted estimates that conflate diagnostic testing and true screening as well as adjusted estimates that properly specify true screening as the outcome.

Study Population. The estimated unadjusted testing rate for the study population was 55.46%. Adjusting for the impetus for testing resulted in a true screening estimate of 45.06%, resulting in an 23.08% overestimation of the true screening rate.

***Race/ethnicity*.** Estimates of unadjusted rates of testing in NHW were higher than in NHB. Adjusting for the impetus for the test reduced both the overall rates and between-group estimates. True screening rates were significantly different for NHW and NHB, at 45.69% and 40.51% (p = 0.001), respectively. Compared to NHB, the rate of diagnostic testing was greater in NHW (19.11% versus 15.68%, p = 0.0957), leading to overestimation of true screening rates in NHW by 23.62% and NHB by 18.60%.

***Sex*.** Unadjusted testing rates were slightly higher in males than females but the difference was not statistically significant (p = 0.6812). After adjustment for the impetus for testing, this advantage increased and became statistically significant (true screening rates of 46.87% versus 43.63%, p = 0.004). Females reported higher rates of diagnostic testing than males (20.98% versus 15.96%), yielding an overestimation of true screening rates in 26.55% for females and 18.99% for males.

Race/ethnicity and sex. In addition to race/ethnicity and sex, we also examined patterns of testing rates by these variables together. NHWF and NHWM reported the highest unadjusted rates of testing. Once adjusted, the advantage of NHW over NHB persisted for each sex. However, NHWF reported the highest rates of diagnostic testing (21.53%), followed by NHBF (16.61%), NHWM (16.14%), and the lowest rate of diagnostic testing was in NHBM (14.34%). Adjusted rates of true screening were

substantially overestimated for all groups, ranging from 16.74% (NHBM) to 27.44% (NHWF).

***Age*.** With one exception, both unadjusted testing rates and adjusted true screening rates increased with age. The lowest unadjusted testing rates were in 50-54 year olds and the highest unadjusted testing rates in 70-74 year olds. Differences in unadjusted and adjusted rates were statistically significant at the $p < .0001$ level. The greatest overestimation of true screening rates was in 50-54 year-old respondents who reported a diagnostic testing rate of 24.24%, resulting in overestimation of true screening rates by 32.00%. Respondents who were 70-74 years old reported the lowest rate of diagnostic testing (16.05%), resulting in the least overestimation of true screening rates (19.12%).

In order to provide data on group differences by age that would be comparable with existing literature, we also analyzed pre- and post-Medicare populations and found that respondents aged 50-64 years reported lower unadjusted testing rates and lower true screening rates compared with their older counterparts ($p < .0001$ in both analyses). Respondents aged 50-64 years reported a higher rate of diagnostic testing than those aged 65-80 years (20.06% versus 17.01%), resulting in an overestimation of true screening rates of 25.09% and 20.50%, respectively.

***Educational attainment*.** Both unadjusted testing rates and adjusted true screening rates increased as educational attainment increased. Unadjusted rates ranged from 44.91% to 69.53% ($p < .0001$) while adjusted rates ranged from 34.23% to 58.99% ($p < .0001$). Individuals at the lowest level of educational attainment reported the

highest rate of diagnostic testing (23.77%) while those at the highest level reported the lowest rate (15.16%). Overestimation of true screening rates was inversely associated with educational attainment level and ranged from 31.18% to 17.87%.

***Poverty ratio*.** Both unadjusted testing rates and adjusted true screening rates increased as the ratio of family income to the poverty threshold increased across tertiles, and adjusted true screening rates ranged from 36.09% in the lowest tertile to 46.18% in the medium tertile and 53.51% in the highest tertile ($p < .0001$). Individuals in the lowest tertile reported the highest rate of diagnostic testing (24.33%), while those in the highest tertile reported the lowest (14.75%). The greatest rate of screening rate overestimation was in the lowest poverty ratio tertile at 17.30%.

Insurance. Respondents reporting insurance coverage reported higher unadjusted testing rates than those without insurance coverage (57.84% versus 24.10%) as well as higher adjusted rates of true screening (47.27% versus 15.91%). Respondents reporting no coverage were more likely to receive diagnostic testing, resulting in a 51.52% overestimation of true screening rates for the uninsured population compared to 22.37% for those with insurance.

Physician recommendation. Respondents reporting a physician recommendation reported much higher unadjusted testing rates than those who did not (84.35% versus 11.54%) as well as higher adjusted rates of true screening (62.70% versus 8.22%). Respondents reporting no recommendation were slightly more likely to receive diagnostic testing, resulting in a 40.32% overestimation of true screening rates for those reporting no recommendation.

Table 2.1. Weighted Estimates of Demographic Characteristics of Full Population and Subpopulation[A]

	Full Population					Subpopulation				
	n	%	CI of %	Rao-Scott Chi-Square	Pr > ChiSq	n	%	CI of %	Rao-Scott Chi-Square	Pr > ChiSq
Non-Hispanic White	5780	87.83	(86.96, 88.70)	3132	<.0001	3214	89.57	(88.65, 90.49	2670.14	<.0001
Non-Hispanic Black	1278	12.17	(11.30, 13.04)			583	10.43	(9.51, 11.36)		
Male	3094	44.20	(43.04, 45.37)	94.08	<.0001	1669	44.50	(42.83, 46.18)	41.28	<.0001
Female	3964	55.80	(54.63, 56.96)			2128	55.50	(53.82, 57.17)		
Non-Hispanic White Male	2575	39.23	(38.04, 40.43)	3836.50	<.0001	1433	40.25	(38.57, 41.93)	2391.45	<.0001
Non-Hispanic Black Male	519	4.97	(4.49, 5.46)			236	4.26	(3.64, 4.88)		
Non-Hispanic White Female	3205	48.60	(47.37, 49.83)			1781	49.32	(47.60, 51.04)		
Non-Hispanic Black Female	759	7.20	(6.52, 7.87)			347	6.17	(5.42, 6.93)		
Ages 50-54	1565	22.18	(21.18, 23.17)	382.70	<.0001	629	16.92	(15.57, 18.26)	98.49	<.0001
Ages 55-59	1511	20.97	(19.92, 22.02)			796	20.41	(19.02, 21.80)		
Ages 60-64	1279	18.41	(17.40, 19.42)			745	19.77	(18.34, 21.20)		
Ages 65-69	1070	15.07	(14.20, 15.95)			643	16.72	(15.46, 17.98)		
Ages 70-74	794	11.25	(10.42, 12.08)			511	13.11	(11.97, 14.26)		
Ages 75-80	839	12.12	(11.31, 12.93)			473	13.08	(12.04, 14.11)		
Ages 50-64	4355	61.56	(60.32, 62.79)	322.23	<.0001	2170	57.09	(55.38, 58.80)	65.13	<.0001
Ages 65-80	2703	38.44	(37.21, 39.68)			1627	42.91	(41.20, 44.62)		
Less than High School Diploma	1017	13.97	(12.89, 15.04)	820.36	<.0001	433	11.27	(10.12, 12.42)	433.73	<.0001
High School Degree or GED	2184	30.39	(29.15, 31.63)			1101	28.52	(26.93, 30.12)		
Some College No Degree or Associate Degree	2004	28.51	(27.34, 29.67)			1100	29.32	(27.69, 30.94)		
Bachelor's Degree	1047	15.13	(14.23, 16.03)			620	16.21	(14.96, 17.45)		
Master's, Professional or Doctorate Degree	761	11.45	(10.59, 12.31)			527	14.36	(13.09, 15.64)		
Poverty Ratio - Low Tertile	2587	35.15	(33.56, 36.74)			1172	30.09	(28.12, 32.06)		
Poverty Ratio - Medium Tertile	2227	31.65	(30.31, 32.99)			1231	32.27	(30.42, 34.13)		
Poverty Ratio - High Tertile	2243	33.20	(31.69, 34.71)			1390	37.64	(35.74, 39.54)		
Insurance - Not Covered	525	7.02	(6.35, 7.69)	4146.62	<.0001	123	3.03	(2.36, 3.71)	2193.76	<.0001
Insurance - Covered	6526	92.98	(92.31, 93.65)			3672	96.97	(96.29, 97.65)		
Marital Spouse not in HH or Separated	274	3.71	(3.23, 4.19)	3412.65	<.0001	121	3.03	(2.42, 3.64)	2202.24	<.0001
Widowed or Divorced	2471	34.68	(33.28, 36.08)			1248	32.60	(30.89, 34.31)		
Never Married / Unknown Status	632	8.44	(7.73, 9.15)			283	7.27	(6.33, 8.21)		
Living with Marital Spouse or Unmarried Partner	3681	53.17	(51.68, 54.66)			2145	57.11	(55.20, 59.01)		
Usual Source of Health Care - No	477	6.80	(6.17, 7.43)	4564.11	<.0001	90	2.32	(1.80, 2.84)	2957.08	<.0001
Usual Source of Health Care - Yes	6476	93.20	(92.57, 93.83)			3707	97.68	(97.16, 98.20)		
Physician Recommendation - No	2534	36.40	(34.95, 37.84)	317.08	<.0001	283	7.22	(6.31, 8.13)	2313.67	<.0001
Physician Recommendation - Yes	4206	63.60	(62.16, 65.05)			3514	92.78	(91.87, 93.69)		

[A] Full population defined as adults ages 50-80 with no history of colorectal cancer. Subpopulation defined as adults ages 50-80 with no history of colorectal cancer and reported any ACS guideline adherent testing.

Note: unweighted n, weighted % and CI of %

Table 2.2. Weighted Estimates of True Screening Rates and Diagnostic Testing Rates for Subpopulation

	n	True Screening %	CI of %	Diagnostic Testing %	CI of %	Rao-Scott Chi-Square[1]	Pr > ChiSq
Subpopulation	3793	81.25	(79.76, 82.74)	18.75	(17.26, 20.24)		
Non-Hispanic White	3210	80.89	(79.29, 82.50)	19.11	(17.50, 20.71)	2.78	0.0957
Non-Hispanic Black	583	84.32	(80.80, 87.84)	15.68	(12.16, 19.20)		
Female	2125	79.02	(76.95, 81.08)	20.98	(18.92, 23.05)	13.49	0.0003
Male	1668	84.04	(82.13, 85.94)	15.96	(14.06, 17.87)		
Non-Hispanic White Male	1432	83.86	(81.84, 85.89)	16.14	(14.11, 18.16)	17.37	0.0006
Non-Hispanic Black Male	236	85.66	(79.51, 91.82)	14.34	(8.18, 20.49)		
Non-Hispanic White Female	1778	78.47	(76.26, 80.67)	21.53	(19.33, 23.74)		
Non-Hispanic Black Female	347	83.39	(78.95, 87.84)	16.61	(12.16, 21.05)		
Ages 50-54	629	75.76	(72.13, 79.39)	24.24	(20.61, 27.87)	15.84	0.0073
Ages 55-59	796	80.79	(77.66, 83.91)	19.21	(16.09, 22.34)		
Ages 60-64	743	82.66	(79.77, 85.56)	17.34	(14.44, 20.23)		
Ages 65-69	641	82.22	(79.33, 85.11)	17.78	(14.89, 20.67)		
Ages 70-74	511	83.95	(80.58, 87.32)	16.05	(12.68, 19.42)		
Ages 75-80	473	83.01	(79.10, 86.93)	16.99	(13.07, 20.90)		
Ages 50-64	2168	79.94	(77.94, 81.94)	20.06	(18.06, 22.06)	5.07	0.0243
Ages 65-80	1625	82.99	(81.02, 84.96)	17.01	(15.04, 18.98)		
Less than High School Diploma	433	76.23	(71.76, 80.70)	23.77	(19.30, 28.24)	11.24	0.024
High School Degree or GED	1098	80.30	(77.62, 82.98)	19.70	(17.02, 22.38)		
Some College No Degree or Associate Degree	1100	81.71	(79.06, 84.37)	18.29	(15.63, 20.94)		
Bachelor's Degree	620	82.56	(79.50, 85.62)	17.44	(14.38, 20.50)		
Master's, Professional or Doctorate Degree	526	84.84	(81.46, 88.22)	15.16	(11.78, 18.54)		
Poverty Ratio - Low Tertile	1172	75.67	(72.77, 78.35)	24.33	(21.65, 27.23)	f(2, 591.31) = 17.61 P>F <.0001	
Poverty Ratio - Medium Tertile	1231	81.77	(79.30, 84.01)	18.23	(15.99, 20.70)		
Poverty Ratio - High Tertile	1390	85.25	(83.10, 87.17)	14.75	(79.71, 82.70)		
Insurance - Not Covered	123	66.00	(56.14, 75.86)	34.00	(24.14, 43.86)	14.63	0.0001
Insurance - Covered	3668	81.72	(80.25, 83.20)	18.28	(16.80, 19.75)		
Living with Marital Spouse or Unmarried Partner	2143	82.51	(80.58, 84.43)	17.50	(15.57, 19.42)	12.53	0.0058
Never Married / Unknown Status	283	84.82	(80.33, 89.31)	15.18	(10.69, 19.67)		
Marital Spouse not in HH or Separated	120	71.37	(62.11, 80.64)	28.63	(19.36, 37.89)		
Widowed or Divorced	1247	79.16	(76.55, 81.78)	20.84	(18.22, 23.45)		
Usual Source of Health Care - No	90	64.62	(53.59, 75.64)	35.38	(24.36, 46.41)	13.02	0.0003
Usual Source of Health Care - Yes	3703	81.65	(81.13, 83.17)	18.35	(16.83, 19.87)		
Physician Recommendation - No	283	78.34	(73.11, 83.56)	21.67	(16.44, 26.89)	1.42	0.2333
Physician Recommendation - Yes	3510	81.48	(79.92, 83.03)	18.52	(16.97, 20.08)		

Note: unweighted n, weighted % and CI of %

[1] Exception: test of significance for Poverty Ratio is Global F Test, computed with Stata.

Table 2.3. Weighted Estimates of Unadjusted Testing Rates and Adjusted True Screening Rates for Full Study Population

	n	Unadjusted Testing Rate %	CI of %	Tests of Significance[1]		Adjusted True Screening Rate %*	CI of %	Tests of Significance[2]	Overestimation of True Screening Rate (%)*
Full Study Population	6984	55.46	(54.04, 56.88)			45.06	(43.91, 46.22)		23.08
Non-Hispanic White	5727	56.48	(54.98, 57.98)	23.47	<.0001	45.69	(44.47, 46.90)	t = -3.50 P>|t| = 0.001	23.62
Non-Hispanic Black	1257	48.04	(44.77, 51.30)			40.51	(37.75, 43.26)		18.60
Female	3922	55.21	(53.43, 56.99)	0.1688	0.6812	43.63	(42.22, 45.04)	t = 2.89 P>|t| = 0.004	26.55
Male	3062	55.77	(53.63, 57.92)			46.87	(45.07, 48.68)		18.99
Non-Hispanic White Male	2551	56.77	(54.49, 59.05)	20.39	0.0001	47.61	(36.91, 45.08)	F(3,296) = 6.72 P>F =.0002	19.25
Non-Hispanic Black Male	511	47.86	(43.09, 52.63)			41.00	(36.43, 43.90)		16.74
Non-Hispanic White Female	3176	56.24	(54.33, 58.15)			44.13	(42.64, 45.63)		27.44
Non-Hispanic Black Female	746	48.16	(43.69, 52.63)			40.16	(36.43, 43.90)		19.92
Ages 50-54	1556	42.12	(39.14, 45.09)	148.98	<.0001	31.91	(29.65, 34.16)	F(5,294) = 39.08 P>F <.0001	32.00
Ages 55-59	1497	53.92	(51.19, 56.65)			43.56	(41.35, 45.77)		23.78
Ages 60-64	1262	59.74	(56.76, 62.73)			49.38	(46.91, 51.85)		20.98
Ages 65-69	1058	61.59	(58.45, 64.73)			50.64	(48.06, 53.22)		21.62
Ages 70-74	787	64.40	(60.68, 68.12)			54.06	(50.94, 57.18)		19.12
Ages 75-80	824	60.27	(58.82, 63.72)			50.03	(47.16, 52.89)		20.47
Ages 50-64	4315	51.39	(49.63, 53.15)	64.27	<.0001	41.08	(49.73, 53.18)	t = -9.71 P>|t| <.0001	25.09
Ages 65-80	2669	62.00	(59.93, 64.08)			51.45	(39.67, 42.49)		20.50
Less than High School Diploma	1002	44.91	(41.66, 48.16)	103.31	<.0001	34.23	(31.76, 36.72)	F(4,295) = 48.78 P>F <.0001	31.18
High School Degree or GED	2162	52.00	(49.65, 54.36)			41.76	(39.87, 43.65)		24.53
Some College No Degree or Associate Degree	1987	56.92	(54.38, 59.46)			46.51	(44.43, 48.59)		22.38
Bachelor's Degree	1039	59.35	(55.73, 62.98)			49.00	(46.01, 52.00)		21.12
Master's, Professional or Doctorate Degree	753	69.53	(65.95, 73.10)			58.99	(55.95, 62.02)		17.87
Poverty Ratio - Low Tertile	2549	47.69	(45.49, 49.89)	F(1.98, 590.51) = 52.47 P>F <.0001		36.09	(34.48, 37.70)	F(2,297) = 103.18 P>F<.0001	32.15
Poverty Ratio - Medium Tertile	2211	56.47	(54.01, 58.73)			46.18	(44.25, 47.94)		22.29
Poverty Ratio - High Tertile	2224	62.77	(60.54, 64.99)			53.51	(51.61, 55.37)		17.30
Insurance - Not Covered	517	24.10	(19.34, 28.85)	131.2	<.0001	15.91	(12.76, 19.04)	t = 18.21 P>|t| <.0001	51.52
Insurance - Covered	6461	57.84	(56.36, 59.32)			47.27	(46.05, 48.47)		22.37
Living with Marital Spouse or Unmarried Partner	3646	59.49	(57.64, 61.34)	47.14	<.0001	49.09	(47.56, 50.61)	F(3,296) = 24.09 P>F <.0001	21.20
Never Married / Unknown Status	622	47.96	(43.40, 52.52)			40.68	(36.81, 44.55)		17.90
Marital Spouse not in HH or Separated	271	45.34	(38.24, 52.43)			32.36	(27.29, 37.42)		40.11
Widowed or Divorced	2445	52.17	(49.94, 54.41)			41.30	(39.53, 43.07)		26.33
Usual Source of Health Care - No	470	19.29	(15.38, 23.20)	229.29	<.0001	12.47	(9.94, 14.99)	t = 24.88 P>|t| <.0001	54.75
Usual Source of Health Care - Yes	6409	58.97	(57.57, 60.44)			48.15	(46.96, 49.35)		22.47
Physician Recommendation - No	2534	11.54	(10.14, 12.94)	3298.18	<.0001	9.04	(7.95, 10.14)	t = 77.60 P>|t| <.0001	27.65
Physician Recommendation - Yes	4206	84.35	(83.13, 85.58)			68.73	(67.73, 69.73)		22.73

Note: unweighted n, weighted % and CI of %

[1] Rao-Scott Chi-Square and Pr > ChiSq calculated for all variables with SAS except Poverty Ratio, where Design-Based Pearson F test was used with Stata.

[2] For groups with 2 categories, 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.

* applying true screening % from subpopulation (ACS guideline-adherent screeners who reported a recommendation) to the unadjusted population rate (which is based on the full population of 50-80 year olds with no history of colorectal disease).

Table 2.4. Multivariate Models Estimating Predictors of Any ACS Guideline Adherent Testing (Full Population) and True Screening (Subpopulation)

Full Study Population - Outcome: Any ACS guideline adherent testing n = 7058

Predictor*	Category	Odds Ratio (OR)	Confidence Limits (OR)		*Wald Chi-Square*	Pr > ChiSq
Intercept		0.075	0.051	0.110	177.65	<.0001
Race/Ethnicity	Non-Hispanic Black	0.584	0.373	0.837	6.86	0.0088
Age	Continuous (Centered (65))	1.034	1.030	1.053	17.01	<.0001
Education	High School Graduate or GED	1.219	0.990	1.500	3.49	0.0619
	Some College No Degree or Associate Degree	1.388	1.111	1.735	8.34	0.0039
	Bachelors Debree	1.651	1.260	2.164	13.24	0.0003
	Master Professional or Doctorate	2.250	1.650	3.070	26.23	<.0001
Usual Source of Health Care	Yes	1.586	1.127	2.230	7.02	0.0081
Physician Recommendation	Yes	38.122	31.328	46.395	1321.41	<.0001
Race * Age	Non-Hispanic Black * Age (Continuous (Centered 65))	0.972	0.946	0.999	4.18	0.041
Race * Physician Recommendation	Non-Hispanic Black * Yes	1.768	1.106	2.824	5.67	0.0172
Age * Physician Recommendation	Age (Continuous (Centered 65)) * Yes	1.020	1.001	1.040	4.11	0.0427

* Referent group for Race/Ethnicity is Non-Hispanic White, for Education, Less than High School Diploma; for Usual Source of Health Care, No; for Physician Recommendation, No.

Subpopulation -Outcome: ACS guideline adherent True Screening n - 3791

Predictor*	Category	Odds Ratio (OR)	Confidence Limits (OR)		*t*	Pr > |t|
Intercept		1.273	0.768	2.111	0.94	0.3489
Race/Ethnicity	Non-Hispanic Black	1.412	1.046	1.906	2.25	0.0242
Age	Continuous (Centered (65))	1.031	1.019	1.044	5	<.0001
Sex	Male	1.346	1.120	1.617	3.17	0.0015
Poverty Ratio	Medium Tertile	1.529	1.184	1.975	3.31	0.0015
Poverty Ratio	High Tertile	2.100	1.666	2.647	6.28	<.0001
Usual Source of Health Care	Yes	2.112	1.276	3.496	2.91	0.0036
Race * Age	Non-Hispanic Black * Age (Continuous (Centered 65))	0.970	0.940	1.000	-1.96	0.0496

* Referent group for Race/Ethnicity is Non-Hispanic White, for Sex is Female; for Poverty Ratio, Low Tertile; for Usual Source of Health Care, No.

Multivariate logistic regression modeling. Table 2.4 shows the final logistic regression models predicting any ACS guideline adherent testing (including diagnostic testing and true screening when specifying the behavioral outcome) in the full population and true screening (excluding diagnostic testing when specifying the behavioral outcome) in the subpopulation of respondents reporting ACS guideline adherent screening.

The estimates from multivariate models of any ACS guideline adherent testing versus true screening reveal important information about how the same predictor varies across the two outcomes. The most startling change in predictors of any ACS guideline adherent testing versus true screening is the physician recommendation to screen, which is highly predictive of any ACS guideline adherent testing, but is not statistically

significant (and was removed from the model) when predicting true screening.

Level of educational attainment is predictive of any ACS guideline adherent testing, but not predictive of true screening. In contrast to educational attainment, poverty ratio is not predictive of any ACS guideline adherent testing, but is only statistically significant in the model predicting true screening (overall test of poverty ratio $p < .0001$, not in table). Increases in poverty ratio result in increased adjusted odds ratios for medium and high tertile categories (1.529, 95% CI [1.184, 1.975] and 2.100, 95% CI [1.666, 2.647], respectively.

Discussion

Colorectal cancer is a disease that is largely preventable with screening, yet screening rates remain low across population groups, risking a high burden of the disease over the next decades. By choice or by data limitations, researchers often conflate diagnostic testing and true screening when defining their behavioral outcome as they measure screening rates, disparities, predictors, and trends over time. In addition, intervention evaluators and policymakers routinely conflate diagnostic testing with true screening in their assessments of both intervention effectiveness and their own success in reaching federal screening benchmarks. In this study, we investigated whether this practice of conflating diagnostic testing and true screening overestimates screening rates, whether that overestimation varies across sociodemographic groups, and whether the predictors of any ACS guideline adherent testing (including diagnostic testing and true screening) vary from those predicting true screening.

Using data from the National Health Interview Survey, we showed that the practice of conflating diagnostic testing and true screening results in substantial overestimation of screening rates. This overestimation varied considerably across sociodemographic characteristics. Finally, we found that correctly specifying the behavioral outcome of interest as true screening changed predictors of the behavior when compared to predictors of a behavioral outcome that conflates diagnostic testing and true screening. These results inform the CRC literature in some unexpected and expected ways.

Before discussing implications for researchers, practitioners, and policymakers more broadly, we would like to highlight two unexpected findings in our study, namely the pre- and post-adjustment rates of male and female respondents, and the extent to which females (and NHW females in particular) engage in diagnostic testing and true screening. We found that although the pre-adjustment rates are nearly identical between sexes (consistent with data from the National Center for Health Statistics, 2011, which ignores the impetus for testing), a statistically significant difference emerges after adjusting for the impetus for testing, with males having a statistically significant advantage in true screening rates over their female counterparts (46.87% vs. 43.63%, $p = 0.001$). This difference in true screening rates may result from several factors. Males may have a higher level of risk perception than females, they may be underutilizing diagnostic testing compared to females, or there may be contributing factors at the health care level that influence engagement with true screening. This may

also result from gendered differences in the interpretation of what the survey instrument is asking when it differentiates true screening from diagnostic testing.

We can contrast this finding with both data on disease incidence as well as the broader literature of sex differences in help seeking. Males suffer significantly higher CRC incidence and mortality rates than females. Although some gender differences are theorized to originate in risk factor and hormone exposures, lower rates of screening are broadly hypothesized to contribute to higher disease burden (American Cancer Society, 2012). Considering these data, we would expect to find lower true screening rates in males; we found the opposite. If the results of this study are not due to differences in self report, this suggests that there may be alternate explanations for higher incidence and mortality differences between genders including biological differences in tumor occurrence, growth, and location.

We can also position this gender gap finding in a broader help seeking literature that consistently demonstrates that females seek help and see health care practitioners at a higher rate than their male counterparts. This is theorized to originate in psychosocial differences between sexes, with males being more likely to delay or avoid help seeking for illnesses. It is widely reported that men in the United States have fewer contacts with physicians across the life span (see Mansfield, Addis & Mahalik, 2003). In 2008 (close temporally to the time at which this study's data was collected), over 20% of men in the United States had no visits to a health care provider in the preceding 12 months, compared to 10.8% of women (National Center for Health Statistics, 2010). Considering that CRC screening requires the coordination between physician and

patient, we expected to find lower rates of true screening in males, as they have fewer opportunities to discuss screening with their physicians; again, our results contradict our hypotheses and show that males engage in true screening at a higher rate than females. This suggests that quantity of health care practitioner contact may be less predictive of preventative health behaviors such as CRC screening. Instead, quality of care and adherence to established risk-based behavior recommendations (including CRC screening) may prove to be more predictive of these behaviors. Although females see physicians more frequently than males, this may be less relevant for CRC screening, as many of the invasive options are repeated only every five or ten years.

Although our study sample is made of individuals with no history of CRC, we expected to see patterns in diagnostic testing that reflect disease incidence, with higher rates of diagnostic testing in males and in NHB. Instead, we found the highest rates of diagnostic testing by females. Our findings contradict our hypotheses as well as the findings in the CRC and help seeking literature. A possible explanation for this may be that females are responding more aggressively to changes in bowel habits and are more likely to be referred for diagnostic testing than males, who are more likely to delay help seeking until a condition interferes with their daily life (Galdas, Cheater & Marsall, 2005). This can explain the higher rate of diagnostic testing in females, but the origin of the higher rate of true screening in males is still unclear. Unfortunately much of the help seeking literature focuses on curative, not preventative care, and it is unclear how much of the gendered patterning of help seeking for curative care relates to preventative care

including CRC screening. Further exploration both qualitatively and quantitatively is necessary to understand these unexpected patterns in diagnostic testing and screening.

These findings are of great concern to users of NHIS data. Despite the methodological advantage of NHIS through its ability to differentiate true screening from diagnostic testing, many researchers fail to use all available data in their analyses, or use it differently across studies. For example, in a study of colorectal cancer screening disparities, researchers note,

> "…both diagnostic and screening exams were included in adherence calculations to create the most lenient definition of up-to-date with CRC screening recommendations and to be consistent with Healthy People 2010 goal measurement methods" (James, 2006).

Klabunde et al. (2011) and Shapiro et al. (2008) use similar language, noting that

> "we included…tests done for any reason, not just as part of a routine exam, because the reported reason for having the test may not be accurate, and having the test within the recommended time interval, regardless of the reason, essentially means that the individual has been screened" (Klabunde et al. (2011), p. 1612).

Articles such as these inform the CRC literature (including research, interventions, and policy) in a potentially damaging way, overestimating rates, miscalculating trends over time, misspecifying behavioral predictors and more.

Our results provide evidence that is of most concern to users of BRFSS data, as our results demonstrate that ignoring the impetus for testing and the resulting misspecification of the outcome of interest result in grossly overestimated rates of true CRC screening. Misspecification of the outcome is especially problematic for users of BRFSS data, as the study's methodology results in data that cannot account for impetus for testing. BRFSS data are used for reports in Morbidity and Mortality Weekly Reports

by the Centers for Disease Control and Prevention as well as state-level data for Healthy People goals and benchmarks. In this study sample, overall population rates were inflated by 23.08% without accounting for the impetus for testing, with the least advantaged socioeconomic groups' rate overestimation at 31.18%, 32.15% and 51.52%. This alone has the potential for undermining decades of data from MMWR, Healthy People and AHRQ.

Although the limitations of BRFSS data are frequently acknowledged by researchers (including users of data for MMWR reports -- see Joseph, Rim, & Seeff, 2008), that is simply not enough. This is a preventable limitation with the addition of one simple question – "*Why* did you engage in the CRC testing?" Our study demonstrates that without that one question, we risk overestimation of screening rates and failure to accurately assess predictors of the optimal behavioral outcome. We showed evidence of gross overestimation rates ranging from 16.74% (NHBM) to 54.75% (no usual source of health care).

Inaccurate rate estimation has implications for disparities research as well. One of the most common disparities discussed in the CRC literature is between NHW and NHB. In this sample, failure to account for the impetus for testing resulted in an overestimation of disparities between NHW and NHB of 37.40% (17.57% NHW advantage to 12.79% NHW advantage), and other sociodemographic disparities were mildly and grossly under- or overestimated as well.

Finally, there was substantial change in the nature of the predictors of true screening versus any ACS guideline adherent testing, with both education and physician

recommendation no longer significantly associated with the correctly specified outcome, and sex and poverty ratio predictive of true screening but not an outcome that conflated diagnostic testing and true screening. In sum, correct specification of the behavioral outcome contributes to a more accurate assessment of the behavior that we need for improved public health – true screening – and allows us to better assess predictors of that behavior.

As we consider the design and evaluation of CRC screening interventions, the results from multivariate models predicting incorrectly and correctly specified outcomes provide evidence that longstanding predictors of 'screening' may, in fact, not predict true screening. In particular, the physician recommendation to screen is not, in fact, predictive of true screening in the subpopulation of respondents who reported ACS guideline-adherent testing. This evidence suggests that focusing interventions at the health services level may not, in fact, increase individuals' odds of engaging in CRC screening.

Another interesting finding in the multivariate models is that among the subpopulation who engaged in ACS guideline adherent testing, NHB were at greater odds of engaging in true screening than their NHW counterparts. This highlights the critical importance of correctly specifying the outcome of interest through inclusion of the impetus for testing—what we thought was a strong disadvantage in NHB in the full population was instead a strong advantage in the subpopulation. As discussed earlier, NHW/NHB disparities pre- and post-adjustment decreased by over 37% after adjusting for the impetus for testing. This finding suggests that the perceptions of racial

disparities discussed in the literature to date (showing a strong advantage for NHW) may be the result of higher diagnostic testing in NHW, specifically NHW women. Interventions targeting NHB populations may still be useful, as even with a lower rate of diagnostic testing, they are still engaging in CRC screening at a lower rate, but the degree to which they differ is substantially less once we account for the impetus for testing.

In order to accurately assess whether or not we are meeting federally established benchmarks for CRC screening at individual and institutional levels, we need to derive estimates from data that properly differentiate diagnostic testing from true screening. However, we are falling short of that by using NHIS data without considering the impetus for testing and by using BRFSS data that is unable to differentiate the two behaviors. Estimated population rates from BRFSS are already below federal goals for many sociodemographic groups, but are likely far lower since BRFSS includes diagnostic testing in their screening measures. As James et al. (2006) allude to, Healthy People utilizes NHIS for national-level data, but there is no evidence that the impetus for testing is accounted for in their estimates. And we must remember that this applies not only to overall population rates but to rates by demographic and socioeconomic group as well. Examining variation within or between groups using BRFSS data, especially across levels of socioeconomic status indicators of education, poverty ratio and insurance, is problematic due to differential rates of diagnostic testing. In our sample, diagnostic testing accounted for as much as 35.38% of screening behaviors (in those with no usual source of health care) and varied considerably across sociodemographic and age groups.

This significant variability in diagnostic testing and true screening across sociodemographic groups calls into question the relevance and effectiveness of a one-size-fits-all federal goal such as Healthy People 2020. Consider, for example, the variability in true screening rate estimates across levels of education: from 34.23% (the lowest level) to 58.99% (the highest level). This is a 72% difference comparing the higher rate to the lower. Does a federal benchmark that averages these types of rates across the entire population serve the public most effectively? We argue that instead of one goal, we seek to improve rates across sociodemographic categories, putting resources into assessing and addressing the barriers to CRC screening to each of these groups. Averaging the population as a whole ignores significant variation across population groups and allows us to perceive progress when, in fact, only those with great socioeconomic privilege enjoy the highest rates of screening.

The final policy implication resulting from this data is the need for data collection by SEER to capture data on how individuals with CRC disease were diagnosed – from diagnostic testing or true screening. Ideally, we could place our findings in the conversation of stage and survival of CRC disease. However, this is difficult to do since we would be comparing individuals without cancer (as defined by our population parameters) with those with cancer. In order to determine if there is a relationship between diagnostic testing and stage of disease, we must collect data at the time of diagnosis on how patients were diagnosed (diagnostic testing vs. true screening) and their disease stage to determine if there is an association between timeliness of screening and stage.

This study is subject to several unavoidable limitations. First, we analyzed data collected in 2008, as analyses began as the 2010 survey was being released. 2008 data remained the best option, however, due to its specificity with its questions pertaining to physician recommendations. A general limitation of using NHIS data is the inability to conduct analyses by geographic region with much certainty; other data collected by state-level entities such as the Department of Public Health or data that are representative to each state's population (including BRFSS data) would be more appropriate. CRC screening rates vary considerably by state, and within state by county; this variation should be considered in future studies of CRC screening (Rim et al., 2011; Weir et al., 2003).

Common to survey data, NHIS data are subject to recall bias on the part of the respondents for both dependent and independent variables. Prior studies have found that self-report of CRC screening are moderate to high in both sensitivity and specificity with no clear patterns in differences in the accuracy of self-report by age, sex, race or family history (Baier et al., 2000). 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 (Centers for Disease Control and Prevention, 2012).

In addition, there may be differential bias in comprehension of survey questions. Differential interpretation of questions by race, gender, and socioeconomic status may explain some of our findings. However, the survey questions of interest were follow-up questions to screening responses. These questions received very little attention in

testing, as the majority of the cognitive testing subjects (which numbered only 9 over age 40) indicated they did not undergo CRC testing, resulting in little to no cognitive testing on the follow-up questions (B. Taylor, personal communication, June 17, 2014). Finally, our practice of deriving full population rates using unadjusted testing rates and true screening rate estimates from the subpopulation resulted in estimates that may have confidence intervals narrower than what they should be. This is due to the fact that we are unable to account for the fact that we are treating the rate used in estimation as fixed when it is, in fact, an estimate. The confidence interval for the true screening rate of the full population excludes the variance estimates of the true screening rate of the study's subpopulation. Therefore, the resulting tests of statistical significance will be more likely to show statistical significance than they would have had the results not been derived from two estimates. In the future, researchers should explore the use of jackknifing as a tool to incorporate the variance of both estimates when calculating confidence intervals and resulting statistical tests for the full population.

In sum, what has been called 'screening' in the majority of the CRC literature is erroneous and misleading as it conflates diagnostic testing and true screening. This practice ignores the most fundamental aspect of CRC screening, namely that it occurs at a pre-symptomatic, or asymptomatic, stage of disease. The evidence from this study shows the importance of immediate changes in survey methodology to correctly specify the behavioral outcome that promotes public health: true screening. Quick action to resolve avoidable methodological limitations of BRFSS and other surveys is required to

help ensure collection of accurate screening data. Furthermore, we need to reexamine data across the past decades. Research has shown that increased use of CRC screening over time is attributable to increased use of colonoscopies, which are the most frequent test for diagnostic testing. Research showing increased screening over time may, in fact, be erroneous and represent increases in diagnostic testing instead. As we consider intervention and policy evaluation approaches, it will be critical to accurately measure the behavioral outcome. Excessive diagnostic testing does not promote public health goals and should not be counted as true screening. Finally, policymakers should use these results as a starting point to evaluate the benefits and drawbacks of one-size-fits-all goals. The dangers of conflating diagnostic testing and CRC screening are clear. Researchers, practitioners and policymakers must take steps now to ensure accurate assessment of rates, trends, disparities and other facets of CRC screening so that we may design interventions to address the persistent underutilization of CRC screening.

References

American Cancer Society. (2010, March 10). Cancer facts & figures 2009. Retrieved from http://www.cancer.org/downloads/STT/500809web.pdf

American Cancer Society. (2012). Colorectal Cancer. Retrieved January 2, 2013, from http://www.cancer.org/acs/groups/cid/documents/webcontent/003096-pdf.pdf

Baier, M., Calonge, N., Cutter, G., McClatchey, M., Schoentgen, S., Hines, S., … Ahnen, D. (2000). Validity of self-reported colorectal cancer screening behavior. *Cancer Epidemiology Biomarkers & Prevention*, *9*(2), 229–232.

Beydoun, H. A., & Beydoun, M. A. (2008). Predictors of colorectal cancer screening behaviors among average-risk older adults in the United States. *Cancer Causes & Control*, *19*(4), 339–359.

Breen, N., Wagener, D., Brown, M., Davis, W., & Ballard-Barbash, R. (2001). Progress in cancer screening over a decade: results of cancer screening from the 1987, 1992, and 1998 National Health Interview Surveys. *JNCI Journal of the National Cancer Institute*, *93*, 1704.

Centers for Disease Control. (2010). Vital signs: colorectal cancer screening among adults aged 50-75 years-United States, 2008. *MMWR. Morbidity and Mortality Weekly Report*, *59*(26), 808.

Centers for Disease Control and Prevention. (2012, June 13). NHIS - About the National Health Interview Survey. Retrieved March 14, 2014, from http://www.cdc.gov/nchs/nhis/about_nhis.htm

Coups, E. J., Manne, S. L., Meropol, N. J., & Weinberg, D. S. (2007). Multiple behavioral risk factors for colorectal cancer and colorectal cancer screening status. *Cancer Epidemiology Biomarkers & Prevention*, *16*(3), 510–516.

Doubeni, C., Field, T., Buist, D., Korner, E., Bigelow, C., Lamerato, L., … Hornbrook, M. (2006). Racial differences in tumor stage and survival for colorectal cancer in an insured population. *Cancer*, *109*, 612–620.

Emmons, K., Puleo, E., McNeill, L., Bennett, G., Chan, S., & Syngal, S. (2008). Colorectal cancer screening awareness and intentions among low income, sociodemographically diverse adults under age 50. *Cancer Causes and Control*, *19*, 1031.

Galdas, P. M., Cheater, F., & Marshall, P. (2005). Men and health help‐seeking behaviour: literature review. *Journal of advanced nursing*, *49*(6), 616-623.

Hall, H. I., Van Den Eeden, S. K., Tolsma, D. D., Rardin, K., Thompson, T., Hughes Sinclair, A., … Nadel, M. (2004). Testing for prostate and colorectal cancer: comparison of self-report and medical record audit. *Preventive Medicine*, *39*(1), 27–35.

Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). Applied survey data analysis.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. New York: Wiley. Retrieved from http://dx.doi.org/10.1002/0471722146 http://mirlyn.lib.umich.edu/Record/004133598

Inter-University Consortium for Political and Social Research. (2012). National Health Interview Survey, 2008 (ICPSR 27341). Retrieved from

http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/27341?q=27341&searchSource=icpsr-landing

James, T. M. (2006). Disparities in colorectal cancer screening: a guideline-based analysis of adherence. *Ethnicity & Disease*, *16*, 228.

Joseph, D. A., King, J. B., Miller, J. W., & Richardson, L. C. (2012). Prevalence of colorectal cancer screening among adults–behavioral risk factor surveillance system, United States, 2010. *MMWR Morb Mortal Wkly Rep*, *61*(Suppl), 51–56.

Joseph, D., Rim, S., & Seeff, L. (2008). Use of colorectal cancer tests-United States, 2002, 2004, and 2006. *MMWR*, *57*, 253–8.

Klabunde, C. N., Cronin, K. A., Breen, N., Waldron, W. R., Ambs, A. H., & Nadel, M. R. (2011). Trends in colorectal cancer test use among vulnerable populations in the United States. *Cancer Epidemiol Biomarkers Prev*, *20*, 1611–21. doi:10.1158/1055-9965.EPI-11-0220

Meissner, H. I., Breen, N., Klabunde, C. N., & Vernon, S. W. (2006). Patterns of colorectal cancer screening uptake among men and women in the United States. *Cancer Epidemiol Biomarkers Prev*, *15*, 389–94. doi:10.1158/1055-9965.EPI-05-0678

National Cancer Institute. (2013, June 6). Cancer Screening Overview (PDQ). Retrieved from http://www.cancer.gov/cancertopics/pdq/screening/overview/patient

National Center for Health Statistics (2011). Health, United States, 2010: With Special Feature on Death and Dying. Hyattsville, MD.

O'Malley, A., Forrest, C., Feng, S., & Mandelblatt, J. (2005). Disparities despite coverage: gaps in colorectal cancer screening among Medicare beneficiaries. *Archives of Internal Medicine*, *165*, 2129.

Pruitt, S. L., Shim, M. J., Mullen, P. D., Vernon, S. W., & Amick, B. C. (2009). Association of area socioeconomic status and breast, cervical, and colorectal cancer screening: a systematic review. *Cancer Epidemiology Biomarkers & Prevention*, *18*(10), 2579–2599.

Richards, C. A., Kerker, B. D., Thorpe, L., Olson, C., Krauskopf, M. S., Silver, L. S., … Winawer, S. J. (2011). Increased screening colonoscopy rates and reduced racial disparities in the New York Citywide campaign: an urban model. *The American Journal of Gastroenterology*, *106*(11), 1880–1886.

Rim, S. H., Joseph, D. A., Steele, C. B., Thompson, T. D., Seeff, L. C., & (CDC), C. for D. C. and P. (2011). Colorectal cancer screening—United States, 2002, 2004, 2006, and 2008. *MMWR Surveill Summ*, *60*(suppl 15), 42–46.

Rosen, A., & Schneider, E. (2004). Colorectal cancer screening disparities related to obesity and gender. *Journal of General Internal Medicine*, *19*, 332–338.

SAS Institute Inc. (2014). SAS (Version 9.3). Cary, NC: SAS Institute Inc.

Shapiro, J. A., Seeff, L. C., Thompson, T. D., Nadel, M. R., Klabunde, C. N., & Vernon, S. W. (2008). Colorectal cancer test use from the 2005 National Health Interview Survey. *Cancer Epidemiology Biomarkers & Prevention*, *17*(7), 1623–1630.

Soneji, S., Armstrong, K., & Asch, D. A. (2012). Socioeconomic and physician supply determinants of racial disparities in colorectal cancer screening. *Journal of Oncology Practice*, *8*(5), e125–e134.

StataCorp. (2013). Stata Version 13 (Version Stata Statistical Software: Release 13). College Station, TX: StateCorp LP.

Subramanian, S., Klosterman, M., Amonkar, M., & Hunt, T. (2004). Adherence with colorectal cancer screening guidelines: a review. *Preventive Medicine*, *38*, 536–550.

Trivers, K. F., Shaw, K. M., Sabatino, S. A., Shapiro, J. A., & Coates, R. J. (2008). Trends in Colorectal Cancer Screening Disparities in People Aged 50–64 Years, 2000–2005. *American Journal of Preventive Medicine*, *35*, 185–193. doi:10.1016/j.amepre.2008.05.021

Weir, H. K., Thun, M. J., Hankey, B. F., Ries, L. A. G., Howe, H. L., Wingo, P. A., … Edwards, B. K. (2003). Annual Report to the Nation on the Status of Cancer, 1975-2000, Featuring the Uses of Surveillance Data for Cancer Prevention and Control. *J. Natl. Cancer Inst.*, *95*, 1276–1299. doi:10.1093/jnci/djg040

CHAPTER THREE

Methodological Issues in Colorectal Cancer Research: When a Physician Recommendation Isn't a Physician Recommendation

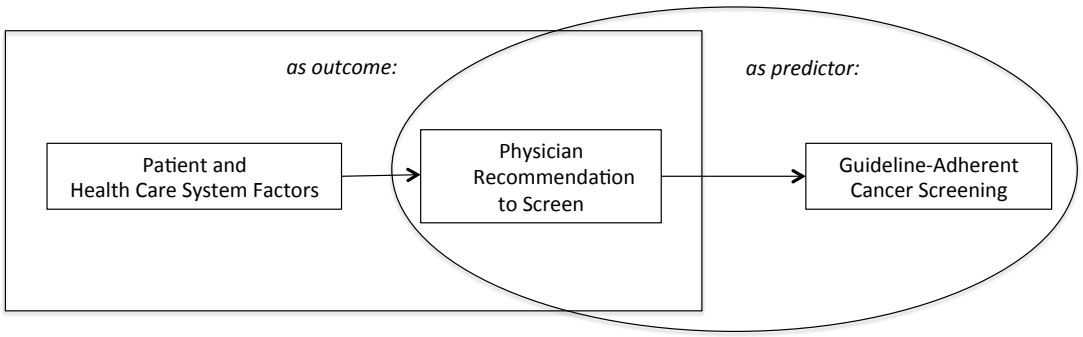
Colorectal cancer (CRC) screening reduces the burden of CRC morbidity and mortality. In population surveys of patients who have engaged in CRC screening, the receipt of a physician recommendation to screen[4] has been shown to be a strong predictor of screening (Coughlin & Thompson, 2005; Klabunde, Breen, & Meissner, 2005). Historically, methods to measure recommendation rates have focused on self-reported data from national surveys. National organizations including the Centers for Disease Control and Prevention (CDC), the American Cancer Society (ACS), and Healthy People 2020 rely on these surveys to provide comprehensive data on cancer screening. The data collected allows researchers to assess rates of screening engagement, rates of recommendations from physicians to patients, disparities in screening and recommendation patterns between groups, trends in screening and recommendations over time, and adherence to recommendations. The Behavioral Risk Factor Surveillance Survey (BRFSS) and the National Health Interview Survey (NHIS) are among the most cited surveys in cancer screening research.

In cancer screening research, we are interested in physician recommendations to

[4] With few exceptions, from this point forward a physician or other health care provider's recommendation to screen will be referred to as simply a 'recommendation.'

screen in two contexts (see **Figure 3.1**). First, we are concerned with recommendations as a behavioral outcome: whether or not physicians are recommending cancer screening to their patients. Second, we are concerned with physician recommendations to screen as a predictor of guideline-adherent cancer screening: whether or not the recommendation to screen is associated with, and predictive of, screening engagement of their patients.

Figure 3.1. Physician Recommendation to Screen as Outcome and Predictor



A review of the literature in both of these contexts reveals irregular measurement of physician recommendations. Specifically, methodological choices and limitations of surveys result in a conflation of referrals for diagnostic testing with recommendations for pre-symptomatic screening (which we will refer to as true screening recommendation). The resulting literature on physician recommendations to screen results in inaccurate estimates of the behavior, both as an outcome and predictor. This conflation of diagnostic referrals and true screening recommendations is not surprising given that the some of the actual tests, including colonoscopy, serve as an important screening test option as well as a diagnostic tool. However, conflating diagnostic testing recommendations with true screening recommendations may lead to incorrect estimations of recommendation rates, trends over time, disparities, and

effectiveness of health services interventions.

Most often, empirical research on these outcomes in CRC screening fall victim to the methodological choices of national surveys. In their work exploring factors associated with racial and ethnic differences in screening recommendations, Ahmed & colleagues (2013) failed to account for the impetus for the recommendation. Though their findings included associations between respondents' socioeconomic status and health care access with differences between racial and ethnic groups, these relationships are likely misestimated due to the conflation of referrals for diagnostic testing and true screening recommendations. James et al. (2006) utilized NHIS data to examine disparities in screening and recommendations, but failed to utilize available data from NHIS that asks respondents to report the impetus for testing (as part of a routine exam or due to a problem), noting,

> "Both diagnostic and screening exams were included in adherence calculations to create the most lenient definition of up-to-date with CRC screening recommendations and to be consistent with Healthy People 2010 goal measurement methods."

Finally, in an analysis of the 2005 NHIS, Shapiro and colleagues (2008) examine physician recommendations as a predictor of screening and note,

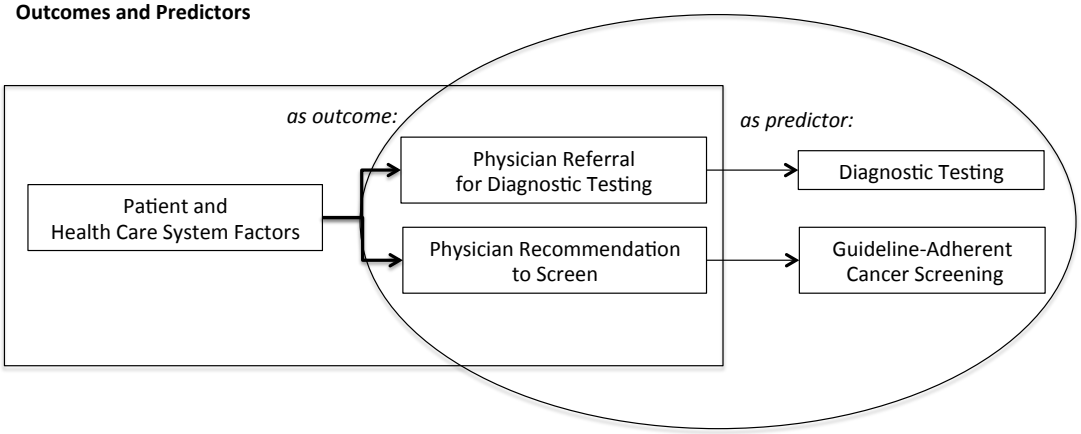
> "Colorectal cancer tests done for any indication were included in the analysis because … even if a test were conducted for nonscreening purposes, a person would have been considered effectively screened."

These studies are important because they are representative of a broader issue: the persistent misspecification of behavioral outcomes in cancer screening research that often fails to recognize the most fundamental characteristic of screening; namely, that

true screening occurs at the pre-symptomatic stage of disease. A physician referral for diagnostic testing is not a physician recommendation to screen. Therefore, multivariate models from these and other studies are likely misestimating a) predictors of physician recommendations and b) the degree to which a recommendation is associated with screening behaviors.

A distinguishing feature of two of the most cited national surveys, the Behavioral Risk Factor Surveillance Survey (BRFSS) and the National Health Interview Survey (NHIS), is their varying ability to differentiate referrals for diagnostic testing and true screening recommendations. Survey researchers utilizing BRFSS data often cite an inability to distinguish between these behaviors as a limitation to their studies (Centers for Disease Control, 2010; Joseph, King, Miller, & Richardson, 2012; Soneji, Armstrong, & Asch, 2012). BRFSS does not ask respondents why they engaged in testing (Centers for Disease Control and Prevention, 2013). In contrast to BRFSS, NHIS affords researchers the opportunity to distinguish between the two by asking respondents not only *what* test they had and *when*, but *why* they had it (Centers for Disease Control and Prevention, 2012). NHIS methodology allows us to better differentiate physician recommendations to screen versus physician referrals for diagnostic testing, as depicted in **Figure 3.2**.

Figure 3.2. Physician Referral for Diagnostic Testing and Physician Recommendation to Screen as Outcomes and Predictors



The goals of the current study are a) to quantify the potential error in CRC screening recommendation rates that results from conflating referrals for diagnostic testing and true screening recommendations; b) to determine whether the quantity of error in CRC recommendation rates varies across sociodemographic categories; and c) to examine how predictive models of true screening recommendations for CRC differ from predictive models of an outcome variable that conflates referrals for diagnostic testing with true screening recommendations. We explore these questions through an analysis of colorectal cancer (CRC) screening behaviors reported in the 2008 NHIS and Cancer Control and Sun Protection Supplement (CCSPS).

Methods

Data source. We examined cross-sectional survey data from the 2008 NHIS and CCSPS. Although there are other options for collecting preventative care data (including other nationally representative datasets and electronic health record data), we are using NHIS for the specific aims and broader goals of this study for several reasons. First, it allows us to consider multiple factors simultaneously as we consider influences on screening recommendation and screening behaviors. Second, NHIS is the only

nationally representative dataset that allows us to capture data on the impetus for testing, a critical component of our analyses. And finally, due to its frequent use in survey research, we can compare and contrast results from this study with a large body of literature in CRC and other cancer screening behaviors. Since its inception in 1957, the NHIS has been the principal source of information on the health and health behaviors of civilian, non-institutionalized households in the United States (Centers for Disease Control and Prevention, 2012). The survey 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 (Centers for Disease Control and Prevention, 2012). The 2008 CCSPS was sponsored by the National Cancer Institute (NCI) and includes questions on physician recommendations for screening, screening behaviors, and reasons for screening tests. The 2008 NHIS/CSSPS was selected over the more recent 2010 NHIS due to the fact that questions on physician recommendations for screening were linked to respondents' screening behaviors rather than in a more general manner. This allowed us to identify the impetus for the physician recommendation as well as CRC testing.

The NHIS sampling strategies result in an oversample of self-identified Black, Hispanic and Asian Americans. Weights constructed for the NHIS respondents reflect the resulting unequal probabilities of selection and also incorporate adjustments for non-response and post-stratification procedures designed to align survey estimates with population distributions from the 2000 Census. The application of these weights in secondary analyses of the NHIS data results in estimates that are representative of the

adult civilian non-institutionalized population of the U.S. (Inter-University Consortium for Political and Social Research, 2012). The annual NHIS response rate averages close to ninety percent of the eligible households in the sample (Centers for Disease Control and Prevention, 2012).

Study population. We calculated estimates of overall population recommendation rates using the full subpopulation, defined as non-Hispanic White and non-Hispanic Black respondents ages 50-80 with no history of colorectal cancer. We chose these two groups because Non-Hispanic Whites and Non-Hispanic Blacks suffer the highest rates of cancers of the colon and rectum in the United States (American Cancer Society, 2011) and are the most common population groups studied in CRC screening literature (Ananthakrishnan, Schellhase, Sparapani, Laud, & Neuner, 2007; Bellizzi, Breslau, Burness, & Waldron, 2011; Breen, Wagener, Brown, Davis, & Ballard-Barbash, 2001; Centers for Disease Control and Prevention (CDC), 2001; James, 2006; Klabunde et al., 2005). Individuals with a history of colorectal cancer were excluded since testing is used as disease management or control and not preventative care. We defined a subpopulation for our analysis of true screening recommendation rates of individuals who reported a recommendation to screen and ACS guideline-adherent screening since only they were asked about the impetus for the testing.

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

Measures

Dependent variables.

Physician recommendation. To measure rates of recommendations that would reflect the methodology of BRFSS, we used questions derived from the NHIS. Respondents who reported any type of testing (including FOBT) in the past ten years were asked, "Was your most recent [screening] test recommended by a doctor or other health professional?" Respondents who did not report testing in the past ten years were asked, "In the past twelve months, has a doctor or other health professional recommended that you be tested to look for problems in your colon or rectum?" A response of "yes" or "no" was coded as such, while a response of "did not see a doctor in the past 12 months" was coded as "no." A response of "refused" or "don't know" was coded as missing data and excluded from analysis. However, in the case of respondents who were asked the question twice due to affirmative responses to both non-invasive and invasive testing, an answer resulting in a code of 'yes' or 'no' for any question resulted in missing data from the other being ignored[5] (total missing n = 91).

Physician referral for diagnostic testing vs. true screening recommendations. To measure rates of referrals for diagnostic testing and true screening recommendations that would reflect the methodology of NHIS, we created a dichotomous variable combining data on reports of recommendations for testing (see above) and the impetus for the reported testing. Individuals who reported invasive tests (sigmoidoscopy, proctoscopy, or colonoscopy) were asked, "What was the MAIN reason

[5] As an example, a respondent did not provide an answer to the question concerning whether or not a recommendation was received for FOBT testing but indicated that a physician had recommended a form of invasive testing. This respondent's recommendation status was coded as 'yes' despite the missing data for the FOBT question.

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 'true 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 this step. If the respondent's answer to this question was coded as missing data but the respondent did report engagement in ACS guideline-adherent FOBT testing, the missing data were ignored[2] (total missing n = 4). Since non-invasive FOBT testing is used exclusively for screening, most reports of FOBT testing were coded as 'true screening.' However, if respondents reported both FOBT and invasive modes of ACS guideline adherent testing, the impetus for testing was categorized solely on the impetus for the invasive testing. A total of 559 respondents fit this criterion: 131 were coded as diagnostic and 428 were coded as true screeners. The resulting dependent variable is a binary measure indicating whether or not the recommendation was a referral for diagnostic testing or a true screening recommendation.

Demographic and control variables. We utilized NHIS data directly for variables of 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 doctorate), poverty ratio (tertiles (measured as ratio of family income to poverty threshold): low (under .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). There were no missing values for poverty ratio as the data are the result of multiple imputations done by the survey administrator. Missing value rates for all variables except poverty ratio were below 3%, a threshold determined to be acceptable with this sample size.

Statistical analysis.

Data analysis. All data analyses for this chapter were performed with the SAS/STAT statistical software, version 9.3_M1 of the SAS System for Windows, Copyright © 2002-2010 SAS Institute Inc. SAS employs procedures capable of computing appropriate variance estimates for survey estimates generated from analyses of complex sample survey data sets such as the 2008 NHIS (Heeringa, West, & Berglund, 2010; SAS Institute Inc., 2014). Due to the fact that SAS excludes strata with only one primary sampling unit (PSU) including sample from the subpopulation of interest from its variance estimates, we ran selected models in a different software package, Stata 13.0, which enables multiple ad-hoc variance estimation methods for dealing with "singleton" PSU to compare estimated standard errors (StataCorp, 2013). For the logistic regression models including main effects, of 268 estimates only ten had greater than or equal to a 10% difference in standard errors, with the greatest difference at 17%. In most cases, Stata produced higher standard errors; none of these differences resulted in changes in statistical inferences.

We began our analysis with weighted and unweighted frequencies of all recommendations (including true screening recommendations and referrals for

diagnostic) for each of the independent variables of interest in the full subpopulation as defined above. We then obtained weighted and unweighted frequencies of true screening recommendations for each of the independent variables of interest in the smaller subpopulation of ACS guideline-adherent screeners. We calculated estimated rates of true screening recommendations for the full population by applying the rates of true screening recommendations from the smaller subpopulation to the unadjusted recommendation rates of the full study population. For example, the unadjusted recommendation rate for the full study population was 63.60%. In the smaller subpopulation of respondents reporting ACS guideline adherent testing, the true screening recommendation rate was 74.24%. We applied this estimate to the unadjusted testing rate (63.60 x 74.24%) to estimate an adjusted true screening recommendation rate for the full population of 47.21%. This same process was repeated for each subgroup of the full study population. Since this calculation reflects the diagnostic testing rates of only respondents reporting ACS guideline adherent testing, it is likely a conservative estimate of the true estimate for the full study population. We calculated estimates of the confidence intervals and resulting tests of significance for the adjusted true screening recommendation rate of the full population. In this process, we were unable to account for the fact that we derived the true screening recommendation rate for the full population from two estimates. This resulted in the possibility of narrower confidence intervals and inferences that were more likely to result in statistically significant values. However, given the precise nature of the subpopulation rate estimate, this issue is likely negligible. Finally, we calculated

the overestimation of true screening recommendation rate by dividing the difference between the unadjusted and adjusted rates by the adjusted rate. For example, the unadjusted recommendation rate for the general population is 63.60% and the adjusted true screening recommendation rate for the same population is 47.21%. We calculated the overestimation of the true screening recommendation rate as (.6360-.4721)/.4721 = 34.70%.

In preparation for fitting multivariate logistic regression models, we examined weighted bivariate associations between independent and dependent variables using the Rao-Scott chi-square test to account for the features of the complex sampling design. Predictors that had a bivariate association with $p < 0.25$ were included in the initial multivariate logistic regression model. In models including two-way interaction terms, both variables defining the interactions will be included in each model, irrespective of bivariate associations (Hosmer & Lemeshow, 2000).

In the second stage of analysis, we used a four-step design-based logistic regression approach to model the odds of receiving a true screening recommendation, adjusting for sociodemographic factors including sex, race, age, education, poverty ratio, health insurance status, and usual source of health care (Breen et al., 2001; O'Malley, Forrest, Feng, & Mandelblatt, 2005; Rosen & Schneider, 2004). First, we fit a main effects model with all covariates. Second, we fit an interaction model with demographic interactions. Third, we fit an interaction model with socioeconomic interactions. Finally, we fit a full model with statistically significant main effects and interaction terms. We estimated the regression parameters using pseudo-maximum likelihood

estimation, due to the complex sampling nature of the NHIS sampled design (Heeringa et al., 2010), and we used Taylor Series Linearization to compute design-based estimates of standard errors for the estimated regression parameters. Hypothesis tests were performed using design-adjusted Wald X^2 tests. We report estimated odds ratios and their 95% confidence intervals.

Results

Please refer to **Table 3.1** for estimated demographic characteristics for the full study population and subpopulation. Overall, the distributions of demographic characteristics are similar between the populations. The full population is slightly younger (22.18% versus 16.75% in ages 50-54) and is weighted more heavily toward the lower ends of educational attainment and poverty ratio. The full population's distribution in the top two tiers of educational attainment totaled 26.58% versus 30.00% in the subpopulation. The highest poverty ratio tertile represents 33.20% of the full population distribution, versus 37.81% of the subpopulation.

Weighted frequencies. Table 3.2 displays estimated rates of true screening recommendations and diagnostic testing referrals for the subpopulation of respondents who reported both ACS guideline adherent testing and a physician recommendation. **Table 3.3** displays estimated rates of true screening recommendations received by the full study population, with unadjusted estimates that conflate diagnostic testing recommendations and true screening recommendations as well as adjusted estimates that properly specify true screening recommendations as the outcome.

Study Population. The estimated unadjusted testing referral rate for the study

population was 63.60%. Adjusting for the impetus for the recommendation reduced this rate estimate to 47.21%, resulting in an overestimation of the true screening recommendation rate of 34.70%.

***Race/ethnicity*.** Rate estimates of recommendations for Non-Hispanic Whites (NHW) were significantly higher than for non-Hispanic Blacks (NHB) both pre- and post-adjustment. True screening recommendation rates for NHW were 47.89% and 42.29% for NHB. Compared to NHB, the rate of referrals for diagnostic testing was greater for NHW, leading to a greater overestimation of true screening recommendation rates for NHW than NHB.

***Sex*.** Prior to adjustment for the impetus for the recommendation, estimates of unadjusted testing referral rates were higher for females than for males ($p = 0.22$). However, after adjustment this advantage changed to a disadvantage, and adjusted true screening recommendation rates were significantly higher for males than for females ($p = 0.004$). Female respondents reported higher rates of referrals for diagnostic testing than males, leading to an overestimation of true screening recommendation rates of 40.07% for females and 28.53% for males.

Race/ethnicity and sex. In addition to examining race/ethnicity and sex independently, we also examined patterns of recommendation rates by combinations of these variables. Non-Hispanic White females (NHWF) and non-Hispanic White males (NHWM) reported the highest unadjusted testing referral rates. Once adjusted, the advantage of NHW over NHB persisted for each sex. However, NHWF reported the highest rates of referrals for diagnostic testing (29.06%), and the lowest rates of

referrals for diagnostic testing were for non-Hispanic Black males (NHBM) (21.42%). Overestimation of true screening recommendation rates was substantial for all groups, ranging from 27.27% (NHBM) to 40.96% (NHWF).

***Age*.** With one exception, estimates of both unadjusted and adjusted rates increase with age. The lowest unadjusted testing referral rates are for 50-54 year olds (54.05%) and the highest overall rates for 70-74 year olds (69.86%). The oldest age group, ages 75-80, reported slightly lower unadjusted rates of testing referrals. The greatest overestimation of true screening recommendation rates is for 50-54 year-old respondents, who report a rate of referrals for diagnostic testing of 32.83%, resulting in overestimation of true screening recommendation rates by 48.88%. Respondents between the ages of 75 and 80 report a diagnostic testing rate of only 21.75% resulting in the least overestimation of true screening recommendation rates at 27.77%.

In order to provide data on group differences by age that would be comparable with existing literature focusing on pre- and post-Medicare populations, we also analyzed these groups and found that those ages 50-64 reported the lowest rates before and after adjustment, and differences between groups were statistically significant ($p < .0001$) in both estimates. Those ages 50-64 reported a higher rate of referrals for diagnostic testing than those ages 65-80, resulting in an overestimation of true screening recommendation rates of 39.26% and 29.08%, respectively.

***Educational attainment*.** Both unadjusted and adjusted rates increased as educational attainment increased. Estimates of unadjusted testing referral rates ranged from 51.69% to 77.20% while adjusted true screening recommendation rates ranged

from 34.14% to 61.44%. Individuals at the lowest level of educational attainment reported the highest rate of referrals for diagnostic testing while those at the highest level reported the lowest rate. Overestimations of true screening recommendation rates were inversely associated with educational attainment level and ranged from 51.39% to 25.66%.

***Poverty ratio*.** Estimates of both unadjusted and adjusted rates of recommendations increased as level of poverty ratio increased. Unadjusted rates ranged from 54.83% to 72.98% while adjusted rates ranged from 36.41% to 58.65%. The highest rates of referrals for diagnostic testing were in the lowest tertile (33.59%). Overestimation of true screening recommendation rates was inversely associated with poverty ratio level and ranged from 24.42% (high poverty ratio tertile) to 50.58% (low poverty ratio tertile), indicating that screening was most likely to be overestimated among the most impoverished.

Insurance. Respondents reporting insurance coverage reported higher unadjusted testing referral rates than those without insurance coverage as well as higher adjusted rates of true screening recommendations. Respondents reporting no coverage were significantly more likely to receive a referral for diagnostic testing (47.68% versus 25.15%, $p < .0001$); overestimation of rates of true screening recommendation are 91.13% for the uninsured population compared to 33.6% for those with insurance.

Marital status. Respondents who reported living with a marital spouse or unmarried partner reported the lowest rates of referrals for diagnostic testing (24.14%),

while respondents with a marital spouse not in the household or separated from their spouse reported the highest rates at 39.57%. Overestimation of true screening recommendation rates by marital status ranged from 31.81% for those living with their spouse or unmarried partner to 65.48% for those not living with a marital spouse or separated.

Table 3.1. Weighted Estimates of Demographic Characteristics of Full Population and Subpopulation[A]

	Full population					Subpopulation				
	n	%	CI of %	Rao-Scott Chi-Square	Pr > ChiSq	n	%	CI of %	Rao-Scott Chi-Square	Pr > ChiSq
Non-Hispanic White	5780	87.83	(86.96, 88.70)	3132.20	<.0001	2971	89.50	(88.57, 90.43)	2632.84	<.0001
Non-Hispanic Black	1278	12.17	(11.30, 13.04)			543	10.50	(9.57, 11.43)		
Male	3094	44.20	(43.04, 45.37)	94.08	<.0001	1542	44.41	(42.72, 46.10)	41.77	<.0001
Female	3964	55.80	(54.63, 56.96)			1972	55.59	(53.90, 57.28)		
Non-Hispanic White Male	2575	39.23	(38.04, 40.43)	3836.50	<.0001	1325	40.18	(38.47, 41.89)	2270.89	<.0001
Non-Hispanic Black Male	519	4.97	(4.49, 5.46)			217	4.23	(3.58, 4.87)		
Non-Hispanic White Female	3205	48.60	(47.37, 49.83)			1646	49.32	(47.60, 51.05)		
Non-Hispanic Black Female	759	7.20	(6.52, 7.87)			326	6.27	(5.49, 7.05)		
Ages 50-54	1565	22.18	(21.18, 23.17)	382.70	<.0001	575	16.75	(15.36, 18.15)	94.85	<.0001
Ages 55-59	1511	20.97	(19.92, 22.02)			739	20.49	(19.06, 21.93)		
Ages 60-64	1279	18.41	(17.40, 19.42)			694	19.85	(18.40, 21.31)		
Ages 65-69	1070	15.07	(14.20, 15.95)			591	16.66	(15.36, 17.96)		
Ages 70-74	794	11.25	(10.42, 12.08)			475	13.08	(11.92, 14.24)		
Ages 75-80	839	12.12	(11.31, 12.93)			440	13.16	(12.10, 14.22)		
Ages 50-64	4355	61.56	(60.32, 62.79)	322.23	<.0001	2008	57.10	(55.37, 58.84)	63.61	<.0001
Ages 65-80	2703	38.44	(37.21, 39.68)			1506	42.90	(41.16, 44.64)		
Less than High School Diploma	1017	13.97	(12.89, 15.04)	820.36	<.0001	403	11.31	(10.10, 12.53)	411.27	<.0001
High School Degree or GED	2184	30.39	(29.15, 31.63)			1026	28.84	(27.17, 30.50)		
Some College No Degree or Associate Degree	2004	28.51	(27.34, 29.67)			1026	29.51	(27.78, 31.24)		
Bachelor's Degree	1047	15.13	(14.23, 16.03)			558	15.60	(14.31, 16.89)		
Master's, Professional or Doctorate Degree	761	11.45	(10.59, 12.31)			486	14.40	(13.07, 15.73)		
Poverty Ratio - Low Tertile	2587	35.15	(33.56, 36.74)			1087	30.08	(28.06, 32.10)		
Poverty Ratio - Medium Tertile	2227	31.65	(30.31, 32.99)			1135	32.11	(30.21, 34.02)		
Poverty Ratio - High Tertile	2243	33.20	(31.69, 34.71)			1292	37.81	(35.85, 39.77)		
Insurance - Not Covered	525	7.02	(6.35, 7.69)	4146.62	<.0001	104	2.76	(2.07, 3.46)	1938.64	<.0001
Insurance - Covered	6526	92.98	(92.31, 93.65)			3408	97.24	(96.54, 97.93)		
Living with Marital Spouse or Unmarried Partner	3681	53.17	(51.68, 54.66)	3412.65	<.0001	1984	57.11	(55.15, 59.08)	2088.84	<.0001
Never Married / Unknown Status	632	8.44	(7.73, 9.15)			256	7.11	(6.14, 8.08)		
Marital Spouse not in HH or Separated	274	3.71	(3.23, 4.19)			109	2.95	(2.34, 3.57)		
Widowed or Divorced	2471	34.68	(33.28, 36.08)			1165	32.82	(31.06, 34.59)		
Usual Source of Health Care - No	477	6.80	(6.17, 7.43)	4564.11	<.0001	70	1.91	(1.45, 2.37)	3185.33	<.0001
Usual Source of Health Care - Yes	6476	93.20	(92.57, 93.83)			3444	98.09	(97.63, 98.55)		
Physician Recommendation - No	2534	36.40	(34.95, 37.84)	317.08	<.0001			(not applicable)		
Physician Recommendation - Yes	4206	63.60	(62.16, 65.05)					(not applicable)		

[A] Full population defined as adults ages 50-80 with no history of colorectal cancer. Subpopulation defined as adults ages 50-80 with no history of colorectal cancer and reported both a physician recommendation to screen and any ACS guideline adherent testing.
Note: unweighted n, weighted % and CI of %. Poverty Ratio chi-square tests calculated with Stata.

Table 3.2. Weighted Estimates of True Screening Recommendation Rates and Diagnostic Testing Referral Rates for Subpopulation

	n	True Screening Recommendation (%)	CI of %	Diagnostic Testing Referral (%)	CI of %	Rao-Scott Chi-Square[1]	Pr > ChiSq
Subpopulation	3514	74.24	(72.65, 75.82)	25.76	(24.18, 27.35)		
Non-Hispanic White	2971	73.99	(72.29, 75.68)	26.01	(24.32, 27.71)	1.29	0.2579
Non-Hispanic Black	543	76.38	(72.60, 80.17)	23.62	(19.83, 27.40)		
Male	1542	77.80	(75.65, 79.95)	22.20	(20.05, 24.35)	17.57	<.0001
Female	1972	71.39	(69.21, 73.57)	28.61	(26.43, 30.79)		
Non-Hispanic White Male	1325	77.72	(75.46, 79.98)	22.28	(20.02, 24.54)	20.76	0.0001
Non-Hispanic Black Male	217	78.58	(71.26, 85.90)	21.42	(14.11, 28.74)		
Non-Hispanic White Female	1646	70.94	(68.61, 73.28)	29.06	(26.72, 31.39)		
Non-Hispanic Black Female	326	74.91	(70.28, 79.53)	25.09	(20.47, 29.72)		
Ages 50-54	575	67.17	(63.24, 71.09)	32.83	(28.91, 36.76)	22.9	0.0004
Ages 55-59	739	73.14	(69.70, 76.59)	26.86	(23.41, 30.30)		
Ages 60-64	694	74.35	(70.98, 77.71)	25.65	(22.29, 29.02)		
Ages 65-69	591	76.58	(73.21, 79.94)	23.42	(20.06, 26.79)		
Ages 70-74	475	77.81	(73.88, 81.73)	22.19	(18.27, 26.12)		
Ages 75-80	440	78.26	(74.00, 82.53)	21.74	(17.47, 26.00)		
Ages 50-64	2008	71.81	(69.60, 74.02)	28.19	(25.98, 30.40)	14.01	0.0002
Ages 65-80	1506	77.47	(75.38, 79.56)	22.53	(20.44, 24.62)		
Less than High School Diploma	403	66.05	(60.74, 71.37)	33.95	(28.63, 39.26)	18.65	0.0009
High School Degree or GED	1026	73.79	(70.86, 76.72)	26.21	(23.28, 29.14)		
Some College No Degree or Associate Degree	1026	74.43	(71.51, 77.34)	25.57	(22.66, 28.49)		
Bachelor's Degree	558	76.09	(72.21, 79.98)	23.91	(20.02, 27.79)		
Master's, Professional or Doctorate Degree	486	79.58	(75.68, 83.48)	20.42	(16.52, 24.32)		
Poverty Ratio - Low Tertile	1087	66.41	(63.13, 69.69)	33.59	(30.31, 36.87)	F(2,588.79) = 24.64	Pr > F <.0001
Poverty Ratio - Medium Tertile	1135	74.34	(71.17, 77.51)	25.66	(22.49, 28.83)		
Poverty Ratio - High Tertile	1292	80.37	(77.92, 82.83)	19.63	(17.17, 22.08)		
Insurance - Not Covered	104	52.32	(41.48, 63.16)	47.68	(36.84, 58.52)	22.14	<.0001
Insurance - Covered	3408	74.85	(73.29, 76.41)	25.15	(23.59, 26.71)		
Living with Marital Spouse or Unmarried Partner	1984	75.86	(73.81, 77.92)	24.14	(22.08, 26.19)	13.82	0.0032
Never Married / Unknown Status	256	75.32	(69.50, 81.14)	24.68	(18.86, 30.50)		
Marital Spouse not in HH or Separated	109	60.43	(50.79, 70.07)	39.57	(29.93, 49.21)		
Widowed or Divorced	1165	72.41	(69.69, 75.13)	27.59	(24.87, 30.31)		
Usual Source of Health Care - No	70	65.84	(54.45, 77.23)	34.16	(22.77, 45.55)	2.48	0.1153
Usual Source of Health Care - Yes	3444	74.40	(72.79, 76.01)	25.60	(23.99, 27.21)		

[1] Exception: test of significance for Poverty Ratio is Global F Test, computed with Stata.
Note: unweighted n, weighted % and CI of %.

Table 3.3. Weighted Estimates of Unadjusted Recommendation Rates and Adjusted True Recommendation Rates for Full Population.

	n	Unadjusted Recommendation Rate (%)	CI of %	Tests of Significance[1]	Adjusted True Recommendation Rate (%)*	CI of %	Tests of Significance[2]	Overestimation of True Screening Recommendation Rate (%)*
Full Population	6740	63.60	(62.16, 65.05)		47.21			34.70
Non-Hispanic White	5537	64.73	(63.21, 66.26)	32.55 <.0001	47.89	(46.77, 49.02)	t = -4.35 P>|t| <.0001	35.16
Non-Hispanic Black	1203	55.37	(52.22, 58.53)		42.29	(39.88, 44.71)		30.92
Male	2965	62.71	(60.64, 64.78)	1.52 0.22	48.79	(47.18, 50.40)	t = 2.94 P>|t| = .004	28.53
Female	3775	64.31	(62.51, 66.12)		45.91	(44.62, 47.20)		40.07
Non-Hispanic White Male	2474	63.79	(61.53, 66.05)	27.59 <.0001	49.58	(47.82, 51.33)	F (3,293) = 8.06 P>F < .0001	28.67
Non-Hispanic Black Male	491	54.18	(49.00, 59.36)		42.57	(38.50, 46.64)		27.27
Non-Hispanic White Female	3063	65.50	(63.60, 67.40)		46.47	(45.12, 47.81)		40.96
Non-Hispanic Black Female	712	56.21	(51.84, 60.57)		42.11	(38.83, 45.38)		33.50
Ages 50-54	1505	54.05	(51.08, 57.02)	79.65 <.0001	36.30	(34.30, 38.30)	F (5,295) = 44.48 P>F < .0001	48.88
Ages 55-59	1433	63.43	(60.65, 66.21)		46.39	(44.36, 48.43)		36.72
Ages 60-64	1220	66.86	(64.07, 69.65)		49.71	(47.64, 51.78)		34.51
Ages 65-69	1025	67.06	(64.01, 70.10)		51.35	(49.02, 53.68)		30.59
Ages 70-74	769	69.86	(66.42, 73.30)		54.36	(51.68, 57.03)		28.52
Ages 75-80	788	66.40	(63.04, 69.75)		51.97	(49.34, 54.59)		27.77
Ages 50-64	4158	61.04	(59.27, 62.82)	27.95 <.0001	43.83	(42.56, 45.11)	t = -9.19 P>|t| <.0001	39.26
Ages 65-80	2582	67.68	(65.67, 69.69)		52.43	(50.88, 53.99)		29.08
Less than High School Diploma	962	51.69	(48.18, 55.20)	118.87 <.0001	34.14	(31.82, 36.46)	F (4,295) = 74.59 P>F < .0001	51.39
High School Degree or GED	2085	60.31	(57.78, 62.84)		44.50	(42.64, 46.37)		35.52
Some College No Degree or Associate Degree	1929	65.60	(63.23, 67.96)		48.82	(47.06, 50.59)		34.36
Bachelor's Degree	1002	67.78	(64.52, 71.05)		51.58	(49.09, 54.06)		31.42
Master's, Professional or Doctorate Degree	724	77.20	(74.02, 80.39)		61.44	(58.90, 63.97)		25.66
Poverty Ratio - Low Tertile	2457	54.83	(52.49, 57.17)	F (2,595) = 75.37 P>F <.0001	36.41	(34.91, 37.92)	F (2,297) = 213.45 P>F <.0001	50.58
Poverty Ratio - Medium Tertile	2143	63.47	(61.00, 65.94)		47.18	(45.51, 48.85)		34.52
Poverty Ratio - High Tertile	2140	72.98	(70.85, 75.10)		58.65	(57.00, 60.30)		24.42
Insurance - Not Covered	500	28.26	(23.50, 33.03)	198.76 <.0001	14.79	(12.30, 17.28)	t = 25.45 P>|t| <.0001	91.13
Insurance - Covered	6233	66.27	(64.78, 67.75)		49.60	(48.49, 50.71)		33.60
Living with Marital Spouse or Unmarried Partner	3533	67.30	(65.45, 69.16)	39.78 <.0001	51.06	(49.65, 52.46)	F(3,298) = 47.11 P>F <.0001	31.81
Never Married / Unknown Status	601	55.90	(51.30, 60.51)		42.11	(38.64, 45.57)		32.76
Marital Spouse not in HH or Separated	262	56.28	(49.19, 63.38)		34.01	(29.73, 38.30)		65.48
Widowed or Divorced	2344	60.54	(58.20, 62.89)		43.84	(42.14, 45.54)		38.10
Usual Source of Health Care - No	458	18.94	(15.11, 22.76)	362.87 <.0001	12.47	(9.95, 14.98)	t = 26.69 P>|t| <.0001	51.88
Usual Source of Health Care - Yes	6282	66.81	(65.33, 68.30)		49.71	(48.60, 50.81)		34.40

Note: unweighted n, weighted % and CI of %

[1] Rao-Scott Chi-Square and Pr > ChiSq calculated for all variables with SAS except Poverty Ratio, where Design-Based Pearson F test was used with Stata.

[2] For groups with 2 categories, 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.

***** applying adjusted recommendation % from subpopulation (ages 50-80 ACS guideline-adherent screeners who reported a recommendation) to the unadjusted population rate (which is based on the full population of 50-80 year olds with no history of colorectal cancer).

Table 3.4. Multivariate Models Estimating Predictors of Any Physician Recommendation (Full Population) and True Physician Recommendation (Subpopulation)

Full Study Population - Outcome: Any Physician Recommendation for Testing

Predictor*	Category	Odds Ratio (OR)	Confidence Limits (OR)		*t*	Pr > |t|
Intercept		0.363	0.243	0.541	-4.97	<.0001
Race/Ethnicity	Non-Hispanic Black	1.369	0.833	2.250	1.24	0.215
Age	Continuous (Centered (65))	1.019	1.010	1.028	4.11	<.0001
Sex	Male	0.605	0.386	0.948	-2.19	0.0283
Education	High School Graduate or GED	1.342	1.116	1.615	3.12	0.0018
	Some College No Degree or Associate Degree	1.655	1.381	1.983	5.46	<.0001
	Bachelor's Degree	1.632	1.295	2.056	4.15	<.0001
	Master's, Professional or Doctorate	2.368	1.845	3.039	6.78	<.0001
Poverty Ratio	Medium Tertile	1.212	1.040	1.412	2.47	0.014
	High Tertile	1.715	1.464	2.009	6.69	<.0001
Insurance	Covered	3.152	2.161	4.597	5.96	<.0001
Age * Sex	Continuous (Centered (65)) * Male	1.015	1.002	1.029	2.19	0.0287
Race * Insurance	Non-Hispanic Black * Covered	0.593	0.354	0.992	-1.99	0.0465
Sex * Insurance	Male * Covered	1.611	1.023	2.535	2.06	0.0395

* Referent group for Race/Ethnicity is Non-Hispanic White, for Sex is Female; for Education, Less than High School Diploma; for Poverty Ratio, Low Tertile; for Insurance, Not Covered.

Subpopulation - Outcome: True Physician Recommendation for Screening

Predictor*	Category	Odds Ratio (OR)	Confidence Limits (OR)		*t*	Pr > |t|
Intercept		1.088	0.687	1.723	0.36	0.7201
Race/Ethnicity	Non-Hispanic Black	1.476	1.163	1.872	3.21	0.0013
Age	Continuous (Centered (65))	1.046	1.032	1.060	6.63	<.0001
Sex	Male	1.251	1.053	1.486	2.54	0.011
Poverty Ratio	Medium Tertile	1.560	1.218	1.998	3.59	0.0006
	High Tertile	2.364	1.873	2.985	7.24	<.0001
Insurance	Covered	1.606	1.008	2.559	1.99	0.0462
Age * Sex	Continuous (Centered (65)) * Male	0.973	0.953	0.992	-2.71	0.0068

* Referent group for Race/Ethnicity is Non-Hispanic White, for Sex is Female; for Poverty Ratio, Low Tertile; for Insurance, Not Covered.

Multivariate logistic regression modeling. Table 3.4 shows the final logistic regression models predicting 1) any referral for diagnostic testing or true screening recommendation in the full study population; and 2) true screening recommendation in the subpopulation of respondents reporting ACS guideline adherent testing.

The multivariate models predicting all recommendations (conflating referrals for diagnostic testing and true screening recommendations) versus a true screening recommendation demonstrate the importance of correctly specifying the behavioral

outcome of interest. We found several important differences in comparing predictors of any recommendation versus a true screening recommendation. Most notably, educational attainment is predictive of the incorrectly specified outcome, but not predictive of true screening recommendations. These results also show that among the subpopulation of those who reported ACS guideline adherent testing and a physician recommendation, respondents are more likely to report a true screening recommendation if they are non-Hispanic Black, older, male, higher poverty ratio, and covered by insurance. There is a significant age-sex interaction in this model at an OR of 0.973 (p=0.0068).

Discussion

Accurate data on physician recommendation rates, disparities, and trends is critical to the continued improvement of CRC morbidity and mortality. The results of this study show that reported rates of CRC screening recommendation are grossly distorted when we fail to account for the impetus for recommendation. By misspecifying diagnostic referrals as true screening recommendations, researchers have published inaccurate screening recommendation rates, trends, and predictors. While the unadjusted estimate of the recommendation rate for the full study population was 63.60%, the adjusted true recommendation rates fell to 47.21%, an overestimation of 34.70%. However, across sociodemographic groups the rate of overestimation of true screening recommendation rates ranged from 24.42% to 91.13% (mean 37.29%, median 34.36%). Especially troubling is the low rate of true screening recommendation and high rates of referrals for diagnostic testing in the youngest respondents, as this is a critical

age of onset of risk. True screening recommendation rates did not exceed 50% of the study population until after age 65. To put both of these findings into context, the Healthy People 2010 goal for provider counseling for CRC screening was 85%, and a developmental goal for Healthy People 2020 is for an increase from Healthy People 2010. Unadjusted and adjusted recommendation rates fall far short of both of these federal goals.

The degree to which screening recommendation rates are distorted varies across socioeconomic subgroups. We found strong evidence of overestimation in groups with lower educational attainment and lower poverty ratio. Surprisingly, we found more diagnostic testing in non-Hispanic Whites compared to non-Hispanic Blacks as well as higher rates in females compared to males. We expected patterns of referrals for diagnostic testing to mirror risk of later stage diagnosis in the SEER data, where we see that irrespective of gender, Blacks suffer higher rates of later stage disease (Surveillance, Epidemiology, and End Results Program, 2014). Although it is difficult to compare our findings with SEER data since our study population is made up of individuals with no history of CRC, it does highlight the need to investigate methods of diagnosis in those patients with CRC disease. This finding is also contrary to empirical research on the patterning of health services utilization by race and sex that consistently demonstrates higher use of preventative care in NHW when compared to NHB and women when compared to men (House & Williams, 2000; Kaplan, Everson, & Lynch, 2000; Smedley, Stith, & Nelson, 2002). These unexpected findings are difficult to fully explain and require further exploration with respect to attribution of impetus of testing

(are there differences in interpretation of 'diagnostic testing' across groups) and psychosocial and instrumental barriers to screening recommendations at the institutional and individual levels.

Two of the most cited predictors of screening recommendations in the literature are patient educational attainment and having a usual source of health care. In prior work by Bao and colleagues (2007), they report that individuals with lower levels of educational attainment were less likely to report physician counseling for CRC testing than their counterparts with higher educational attainment. This effect was statistically significant and persisted in analyses looking at within-physician and between-physician differences. Similar effects of an inverse association between educational attainment and odds of reporting a physician recommendation for CRC testing are demonstrated in research by Ye and colleagues (2009), Brawarsky et al. (2004) and Ahmed et al. (2013). Ahmed and colleagues found a strong and statistically significant relationship between a usual source of health care and odds of reporting a physician recommendation for screening. Individuals with a usual source of health care in that study were 2.83 times more likely than those without to report receiving a recommendation ($p < .001$). All of these examples, however, use models that predict any recommendation and conflate diagnostic referrals with true screening recommendations. In contrast, when we adjusted for the impetus for the recommendation in multivariate models, we found that neither educational attainment nor the presence of a usual source of health care was predictive.

This study is subject to some unavoidable limitations. A key question of rates of adherence is left unanswered due to the methodological barriers present in the data. We utilized data on the impetus for testing to determine the impetus for the physician recommendation. However, we have no data on individuals who did not engage in CRC testing; therefore we are unable to determine rates of adherence to true screening recommendations vs. diagnostic referrals. However, based on the results reported herein, we predict that once both physician recommendations and testing behaviors are correctly specified, adherence rates would likely go down, as much of the testing counting as adherence to recommendations are actually compliance with referrals for diagnostic testing.

First, we analyzed NHIS data collected in 2008, as analyses began as the 2010 survey was being released. 2008 data remained the best option, however, due to its specificity with its questions pertaining to physician recommendations. A general limitation of using NHIS data is the inability to conduct analyses by geographic region with much certainty; other data collected by state-level public health departments or representative to each state's population (including BRFSS data) would be more appropriate. CRC screening rates vary considerably by state, and within state by county; this variation should be considered in future studies of MCC and CRC screening (Rim et al., 2011; Weir et al., 2003).

Common to survey data, NHIS data are subject to recall bias on the part of the respondents for both dependent and independent variables. Prior studies have found that self-report of CRC screening behaviors are demonstrated to be moderate to high in

both sensitivity and specificity with no clear patterns in differences in the accuracy of self-report by age, sex, race or family history (Baier et al., 2000). 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 (Centers for Disease Control and Prevention, 2012). In addition, there may be differential bias in comprehension of survey questions. Differential interpretation of questions by race, gender, and socioeconomic status may explain some of our findings. However, the pivotal questions of interest were follow-up questions to screening responses. These questions received very little attention in cognitive testing, as the majority of the cognitive testing subjects (which numbered only 9 over age 40) indicated they did not undergo CRC testing, resulting in little to no cognitive testing on the follow-up questions (B. Taylor, personal communication, June 17, 2014). NHIS does not collect data on gastrointestinal comorbidities, which may affect an individual's odds of receiving a physician referral for diagnostic testing. Finally, our practice of deriving full population rates using unadjusted testing recommendation rates and estimates from the subpopulation resulted in estimates that may have confidence intervals narrower than what they should be. This is due to the fact that we are unable to account for the fact that we are treating the rate used in estimation as fixed when it is, in fact, an estimate. The confidence interval for the true screening recommendation rate of the full population excludes the variance estimates true screening recommendation rate of the study's subpopulation. Therefore, the resulting tests of statistical significance will be more likely to show statistical

significance than they would have had the results not been derived from two estimates. In the future, researchers should explore the use of jackknifing as a tool to incorporate the variance of both estimates when calculating confidence intervals and resulting statistical tests for the full population.

To our knowledge, this is the first analysis of a nationally representative dataset to measure the degree to which the methodological approach of BRFSS and analytic approach of some researchers using NHIS data impacts our understanding of screening recommendation rates. Researchers using data from BRFSS consistently cite the lack of data on the impetus for testing as a limitation, noting that rates may be over- or underestimated. Our study demonstrates that both overall and across all sociodemographic groups, rates are grossly overestimated when we fail to account for the impetus for the referral or recommendation. This not only has implications for understanding rates and trends, but also disparities by race, sex, and socioeconomic status. In addition to large-scale surveys, the challenges to BRFSS methodology herein extend to practices of evaluating interventions. Simply counting changes in the number of recommendations reported is insufficient, as we need to differentiate between referrals for diagnostic testing and true screening recommendations.

In addition to implications for research, our results suggest important implications for health care services and policy. The patterns of lower true screening recommendations in women and respondents with lower socioeconomic advantages support the need for initiatives that extend the responsibility of cancer screening recommendations. To date, this burden has been with the primary care physician.

However, the source of care for women is often other providers (e.g. gynecologist) and populations of lower socioeconomic advantage often use the emergency room or urgent care as their first point of contact with the health care system (Smedley et al., 2002). Extending the responsibility of the recommendation to all health care providers will broaden the reach of state- and federal-level initiatives.

Interventions directed toward meeting federal goals are critical to reducing the burden of CRC. However, considering the variation in screening recommendation rates that is evident in this nationally representative sample, it is time to consider the usefulness of overall population goals and instead consider goals that are group-specific. Furthermore, quality metrics developed at the policy level by AHRQ use BRFSS data to assess success or failure to meet goals (Agency for Healthcare Research and Quality, 2013). As we demonstrate in this study, the risks of using data that conflates the two behavioral outcomes of interest are substantial, and it paints a much different picture than what is actually occurring at the institutional level.

Finally, our study makes it clear that we must use great caution when reviewing the literature in this area. Significant variation in overestimation of rates by subgroup makes it difficult to apply a "one size fits all" adjustment of overestimation of BRFSS rates. A wide range of research studies and HP 2020 state benchmarks rely on BRFSS data, while other studies and HP 2020 national benchmarks rely on NHIS data. A revised methodological approach of both surveys – both at the data collection and analysis levels – is required in order to accurately assess rates, trends, adherence, and disparities.

References

Agency for Healthcare Research and Quality. (2013, May 7). 2012 National Healthcare Quality Report. Text. Retrieved May 25, 2014, from http://www.ahrq.gov/research/findings/nhqrdr/nhqr12/

Ahmed, N. U., Pelletier, V., Winter, K., & Albatineh, A. N. (2013). Factors Explaining Racial/Ethnic Disparities in Rates of Physician Recommendation for Colorectal Cancer Screening. *American Journal of Public Health*, *103*(7), e91–e99.

American Cancer Society. (2011). Colorectal Cancer Facts & Figures 2011-2013. Retrieved from http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-028323.pdf

Ananthakrishnan, A., Schellhase, K., Sparapani, R., Laud, P., & Neuner, J. (2007). Disparities in colon cancer screening in the Medicare population. *Archives of Internal Medicine*, *167*, 258.

Baier, M., Calonge, N., Cutter, G., McClatchey, M., Schoentgen, S., Hines, S., … Ahnen, D. (2000). Validity of self-reported colorectal cancer screening behavior. *Cancer Epidemiology Biomarkers & Prevention*, *9*(2), 229–232.

Bao, Y., Fox, S. A., & Escarce, J. J. (2007). Socioeconomic and Racial/Ethnic Differences in the Discussion of Cancer Screening:“Between-” versus “Within-” Physician Differences. *Health Services Research*, *42*, 950–970.

Bellizzi, K. M., Breslau, E. S., Burness, A., & Waldron, W. (2011). Prevalence of cancer screening in older, racially diverse adults: still screening after all these years. *Archives of Internal Medicine*, *171*(22), 2031–2037.

Brawarsky, P., Brooks, D., Mucci, L., & Wood, P. (2004). Effect of physician recommendation and patient adherence on rates of colorectal cancer testing. *Cancer Detection and Prevention*, *28*, 260–268.

Breen, N., Wagener, D., Brown, M., Davis, W., & Ballard-Barbash, R. (2001). Progress in cancer screening over a decade: results of cancer screening from the 1987, 1992, and 1998 National Health Interview Surveys. *JNCI Journal of the National Cancer Institute*, *93*, 1704.

Centers for Disease Control. (2010). Vital signs: colorectal cancer screening among adults aged 50-75 years-United States, 2008. *MMWR. Morbidity and Mortality Weekly Report*, *59*(26), 808.

Centers for Disease Control and Prevention. (2012, June 13). NHIS - About the National Health Interview Survey. Retrieved March 14, 2014, from http://www.cdc.gov/nchs/nhis/about_nhis.htm

Centers for Disease Control and Prevention. (2013, July 1). CDC - BRFSS Questionnaires. Retrieved December 20, 2013, from http://www.cdc.gov/brfss/questionnaires.htm

Centers for Disease Control and Prevention (CDC). (2001). Trends in screening for colorectal cancer--United States, 1997 and 1999. *MMWR. Morbidity and Mortality Weekly Report*, *50*(9), 162–166.

Coughlin, S., & Thompson, T. (2005). Physician recommendation for colorectal cancer screening by race, ethnicity, and health insurance status among men and women in the United States, 2000. *Health Promotion Practice*, *6*, 369.

Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). Applied survey data analysis.

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression*. New York: Wiley. Retrieved from http://dx.doi.org/10.1002/0471722146 http://mirlyn.lib.umich.edu/Record/004133598

House, J., & Williams, D. R. (2000). Understanding and Reducing Socioeconomic and Racial/Ethnic Disparities in Health. In B. D. Smedley & S. L. Syme (Eds.), *Promoting Health: Intervention Strategies from Social and Behavioral Research* (pp. 81–124). Washington, DC: National Academy Press.

Inter-University Consortium for Political and Social Research. (2012). National Health Interview Survey, 2008 (ICPSR 27341). Retrieved from http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/27341?q=27341&searchSource=icpsr-landing

James, T. M. (2006). Disparities in colorectal cancer screening: a guideline-based analysis of adherence. *Ethnicity & Disease*, *16*, 228.

Joseph, D. A., King, J. B., Miller, J. W., & Richardson, L. C. (2012). Prevalence of colorectal cancer screening among adults–behavioral risk factor surveillance system, United States, 2010. *MMWR Morb Mortal Wkly Rep*, *61*(Suppl), 51–56.

Kaplan, G., Everson, S., & Lynch, J. (2000). The Contribution of Social and Behavioral Research to an Understanding of the Distribution of Disease: A Multilevel

Approach. In B. D. Smedley & S. L. Syme (Eds.), *Promoting Health: Intervention Strategies from Social and Behavioral Research* (pp. 37–80). Washington, DC: National Academy Press.

Klabunde, C., Breen, N., & Meissner, H. (2005). Use of Colonoscopy for Colorectal Cancer Screening. *Cancer Epidemiology Biomarkers & Prevention*, *14*, 2279.

O'Malley, A., Forrest, C., Feng, S., & Mandelblatt, J. (2005). Disparities despite coverage: gaps in colorectal cancer screening among Medicare beneficiaries. *Archives of Internal Medicine*, *165*, 2129.

Rim, S. H., Joseph, D. A., Steele, C. B., Thompson, T. D., Seeff, L. C., & (CDC), C. for D. C. and P. (2011). Colorectal cancer screening—United States, 2002, 2004, 2006, and 2008. *MMWR Surveill Summ*, *60*(suppl 15), 42–46.

Rosen, A., & Schneider, E. (2004). Colorectal cancer screening disparities related to obesity and gender. *Journal of General Internal Medicine*, *19*, 332–338.

SAS Institute Inc. (2014). SAS (Version 9.3). Cary, NC: SAS Institute Inc.

Shapiro, J. A., Seeff, L. C., Thompson, T. D., Nadel, M. R., Klabunde, C. N., & Vernon, S. W. (2008). Colorectal cancer test use from the 2005 National Health Interview Survey. *Cancer Epidemiology Biomarkers & Prevention*, *17*(7), 1623–1630.

Smedley, B., Stith, A., & Nelson, A. (2002). *Unequal treatment: confronting racial and ethnic disparities in health care*. Natl Academy Pr.

Soneji, S., Armstrong, K., & Asch, D. A. (2012). Socioeconomic and physician supply determinants of racial disparities in colorectal cancer screening. *Journal of Oncology Practice*, *8*(5), e125–e134.

StataCorp. (2013). Stata Version 13 (Version Stata Statistical Software: Release 13). College Station, TX: StateCorp LP.

Surveillance, Epidemiology, and End Results Program. (2014). SEER Cancer Statistics Review: Cancer of the Colon and Rectum (Invasive) 1975-2010. National Cancer Institute. Retrieved from http://seer.cancer.gov/csr/1975_2010/browse_csr.php?sectionSEL=6&pageSEL=sect_06_table.06.html

Weir, H. K., Thun, M. J., Hankey, B. F., Ries, L. A. G., Howe, H. L., Wingo, P. A., … Edwards, B. K. (2003). Annual Report to the Nation on the Status of Cancer, 1975-2000, Featuring the Uses of Surveillance Data for Cancer Prevention and Control. *J. Natl. Cancer Inst.*, *95*, 1276–1299. doi:10.1093/jnci/djg040

Ye, J., Xu, Z., & Aladesanmi, O. (2009). Provider recommendation for colorectal cancer screening: Examining the role of patients' socioeconomic status and health insurance. *Cancer Epidemiology*, *33*(3), 207–211.

CHAPTER FOUR

The Role of Multiple Chronic Conditions in CRC Screening

The segment of the US population at the greatest risk of developing colorectal cancer (CRC) is growing at a rapid rate. By the year 2015, an estimated one in five Americans will be ages 50-64 (Holden, Jonas, Porterfield, Reuland, & Harris, 2010). Despite widespread CRC screening initiatives, studies consistently demonstrate that individuals at the younger end of the at-risk age group (50-64) are less likely to engage in CRC screening compared to their older counterparts (Breen, Wagener, Brown, Davis, & Ballard-Barbash, 2001; Klabunde, Breen, & Meissner, 2005; Schneider, 2009; Subramanian, Klosterman, Amonkar, & Hunt, 2004; Wee, McCarthy, & Phillips, 2005). The difference in adherence between those 50-64 and 65 and older is often attributed to the availability of insurance coverage through Medicare (Gross et al., 2006; Ko, Kreuter, & Baldwin, 2005; O'Malley, Forrest, Feng, & Mandelblatt, 2005).

To date, empirical research exploring predictors of CRC screening has focused on differences in screening patterns across sociodemographic characteristics such as age, race/ethnicity, and socioeconomic status. However, it is possible that other factors including the experience of multiple chronic conditions (MCC), defined as two or more chronic conditions, are associated with CRC screening. MCC creates context for patients and their health care providers that may directly and indirectly influence the two

behavioral outcomes of interest in CRC screening research: the physician recommendation to screen[6] and guideline-adherent screening.

Background

As adults reach the age at which CRC screening should begin, many are experiencing worsening health including the onset or progression of one or more chronic diseases; more than 25% of Americans live with MCC (Singh-Manoux, Ferrie, Chandola, & Marmot, 2004; US Department of Health & Human Services, 2010). It is estimated that by the year 2020, the prevalence of MCC will increase to 81 million Americans, up 42% from the year 2000 (Brody, 2011). Many of the risk factors for MCC also predispose individuals to a higher risk of CRC including sedentary lifestyle, poor diet, smoking and alcohol consumption and environmental exposures (Amos, Bosken, Soliman, & Frazier, 2005; Potter & Hunter, 2002). MCC is unequally distributed across demographic groups. Individuals in lower socioeconomic tiers report poor/fair health status at higher rates than those in higher tiers (AARP, 2010; Cummings & Jackson, 2008). An individual's socioeconomic position is also related to both the timing and severity of chronic disease burden across the life course. Individuals in lower socioeconomic classes experience morbidity earlier and to a greater degree than those in higher socioeconomic classes (House, Kessler, & Herzog, 1990; Kington & Smith, 1997). African Americans suffer a disproportionate burden of MCC, and face diagnosis of chronic disease at a much younger age than their White counterparts, regardless of socioeconomic status (Williams & Jackson, 2005).

[6] With few exceptions, from this point forward, a physician or other health care's recommendation will be referred to as simply a 'recommendation.'

The experience of MCC in older adulthood may influence an individual's likelihood to receive appropriate physician recommendations to screen and engage in CRC screening in a number of ways. Patients with MCC may not receive a recommendation from a physician or other health professional to screen, which is consistently one of the strongest correlates of future screening behaviors across populations (Messina, 2011). A recommendation serves two functions: first, as a cue to action for patients, reminding them of their risk of CRC and recommended screening practices; and second, as a mechanism through which referrals can be made within the health care system to ensure appropriate delivery and payment of screening services. Individuals with MCC and their health care providers face competing demands during the clinical encounter, as the focus of their interactions shifts from wellness care to chronic disease management and treatment. In addition, physicians' perception that the patient will not comply due to his or her medical condition may preclude them from recommending screening (Brawarsky, Brooks, Mucci, & Wood, 2004; Shokar, Carlson, & Weller, 2008). As health care delivery becomes more segmented through use of highly specialized care and the primary responsibility for CRC recommendation remains in the hands of a primary care provider, a recommendation may not be made to individuals with MCC. To date, health care institutions have focused their efforts to increase patient screening rates through initiatives at the primary care level. While placing the burden for recommendations on the primary care provider is relevant for individuals who have access to and use that segment of health care, it misses an increasingly large population that receives care from other sources, including specialty care, emergency

departments and non-clinical care providers, e.g. spiritual leaders (Smedley, Stith, & Nelson, 2002).

In addition to potentially interfering with the receipt of a recommendation to screen, individuals with MCC face unique issues that may interfere with engaging in cancer screening. Results from our earlier qualitative research in racial and ethnic disparities in CRC treatment choices revealed that at the individual level, patients with MCC experienced physiological symptoms of their chronic condition that made them apprehensive to go through another unrelated medical test. In addition, patients with MCC feared another diagnosis, and avoided screening altogether (Becker, E., Elliott, H., Griffith, D., Alexander, G., & Morris, A., 2010). Finally, even at the age of CRC risk, symptoms of early stage CRC disease (change in bowel habits, cramping or abdominal pain, weakness, fatigue, unintended weight loss) may be mistakenly attributed to an individual's chronic disease process or treatment (American Cancer Society, 2012; Gonzalez, Ferrante, Van Durme, Pal, & Roetzheim, 2001). Together, these findings reflect similar patterns found in earlier research across cancer types that shows inverse relationships between an individual's level of chronic disease burden (measured in number of chronic conditions) and the likelihood of screening for breast and cervical cancer as well as being diagnosed at an advanced stage of disease (Fleming, Pursley, Newman, Pavlov, & Chen, 2005; Kiefe, Funkhouser, Ph, Fouad, & May, 1998).

In response to the growing public health threat of CRC, the U.S. Department of Health's Healthy People 2020 has set an aggressive goal of a 30% increase in colorectal screening uptake from 2008, when age-adjusted rates were at 54.2 percent (on

average), to its goal of 70.5% by the year 2020 (US Department of Health & Human Services, 2011). In addition, Healthy People 2020 includes a developmental goal to increase the proportion of adults who are counseled by their providers about colorectal cancer screening. Understanding the association between MCC and CRC screening behaviors of health care providers and patients is an important first step toward the design and implementation of interventions that seek to meet the CRC screening goals of Healthy People 2020 and the needs of the growing segment of the population living with MCC. The goal of this study is to examine the associations between MCC and both behavioral outcomes of interest. We predict that MCC will be associated with reductions in rates of both outcomes, with the greatest reduction of rates in those with the least socioeconomic advantage. We further predict that in multivariate logistic regression models, including MCC in the regression model will attenuate the odds ratio estimates of sociodemographic predictors, including race/ethnicity, age, sex, socioeconomic status measures (including education, poverty ratio, and insurance) and usual source of health care.

Methods

Data source. We examined cross-sectional survey data from the 2008 National Health Interview Survey (NHIS) and Cancer Screening and Sun Protection Supplement (CSSPS). Although there are other options for collecting preventative care data (including other nationally representative datasets and electronic health record data), we are using NHIS for the specific aims and broader goals of this study for several reasons. First, it allows us to consider multiple factors simultaneously as we consider

influences on screening recommendation and screening behaviors. Second, NHIS is the only nationally representative dataset that allows us to capture data on the impetus for testing, a critical component of our analyses. And finally, due to its frequent use in survey research, we can compare and contrast results from this study with a large body of literature in CRC and other cancer screening behaviors. Since its inception in 1957, the NHIS has been the principal source of information on the health and health behaviors of civilian, non-institutionalized households in the United States (Centers for Disease Control and Prevention, 2012). The survey 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 (Inter-University Consortium for Political and Social Research, 2012). The 2008 Cancer Screening and Sun Protection Supplement (CSSPS) was sponsored by the National Cancer Institute (NCI) and includes questions on physician recommendation for screening, screening tests, and reasons for screening tests. The 2008 NHIS/CSSPS was selected over the more recent 2010 NHIS due to the fact that questions on physician recommendations for screening were linked to respondents' screening behaviors rather than in a more general manner. This allowed us to identify the impetus for the physician recommendation as well as CRC testing.

The NHIS sampling strategies result in an oversample of self-identified Black, Hispanic and Asian Americans. Weights constructed for the NHIS respondents reflect the resulting unequal probabilities of selection and also incorporate adjustments for non-response and post-stratification procedures designed to align survey estimates with

population distributions from the 2000 Census. The application of these weights in secondary analyses of the NHIS data results in estimates that are representative of the adult civilian non-institutionalized population of the U.S. (Inter-University Consortium for Political and Social Research, 2012). The annual NHIS response rate averages close to ninety percent of the eligible households in the sample (Centers for Disease Control and Prevention, 2012).

Study population. The full study population consisted of non-Hispanic White and non-Hispanic Black respondents ages 50-64 with no history of colorectal disease. We chose these two groups because Non-Hispanic Whites and Non-Hispanic Blacks suffer the highest rates of cancers of the colon and rectum in the United States (American Cancer Society, 2011) and are the most common population groups studied in CRC screening literature (Ananthakrishnan, Schellhase, Sparapani, Laud, & Neuner, 2007; Bellizzi, Breslau, Burness, & Waldron, 2011; Breen, Wagener, Brown, Davis, & Ballard-Barbash, 2001; Centers for Disease Control and Prevention (CDC), 2001; James, 2006; Klabunde et al., 2005).

Individuals with a history of colorectal cancer were excluded since testing is used as disease management and not preventative care. We are focusing on adults ages 50-64 because CRC screening should be initiated at age 50 for those at average risk of the disease and research consistently shows underutilization of CRC screening modalities in this age group (American Cancer Society, 2011). For the analysis examining patterns of recommendations for any CRC testing, the full study population was analyzed. We created one subpopulation from this full study population for each set of analyses

differentiating a) referrals for diagnostic testing from true screening recommendations[7] and b) true screening[8] from diagnostic testing. Due to the fact that only those who reported testing were asked about the impetus for testing, we restricted both subpopulations for the secondary analyses to individuals who reported American Cancer Society (ACS) guideline adherent testing. Subpopulation A (physician screening recommendation aim) includes only those individuals who reported a recommendation for testing and ACS guideline adherent testing. Subpopulation B (CRC screening aim) includes only those respondents who reported ACS guideline adherent testing.

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

Measures.

Dependent variables. The authors' previous work demonstrated the importance of accounting for the impetus for testing in estimating physician recommendation and screening rates (see Chapters 2 and 3). We established that conflation of diagnostic and screening tests and referrals results in inaccurate screening and recommendation rates. For the current research aims, we provide information on two outcomes – first, an outcome that conflates diagnostic testing referrals and true screening recommendations (or diagnostic testing and true screening), and second we examine true screening recommendations and true screening. This second set of outcomes may not be as easily compared with existing literature in physician recommendation and screening behaviors

[7] "True screening recommendation" refers to a recommendation that results in pre-symptomatic screening (and not diagnostic testing).

[8] "True Screening" refers to CRC testing that is done as part of a routine exam (pre-symptomatic) and not due to a problem (diagnostic).

due to the common practice of conflating true screening with diagnostic testing. However, defining the outcomes in a manner that accounts for the impetus for testing is most consistent with assessing accurate rates and significant predictors of the behaviors.

Physician recommendation. Respondents who reported any type of testing (including fecal occult blood tests (FOBT)) in the past ten years were asked, "Was your most recent [screening] test recommended by a doctor or other health professional?" Respondents who did not report testing in the past ten years were asked, "In the past twelve months, has a doctor or other health professional recommended that you be tested to look for problems in your colon or rectum?" A response of "yes" or "no" was coded as such, while a response of "did not see a doctor in the past 12 months" was coded as "no." A response of "refused" or "don't know" was coded as missing data and excluded from analysis. However, in the case of respondents who were asked the question twice due to affirmative responses to both non-invasive and invasive testing, an answer resulting in a code of 'yes' or 'no' for any question resulted in missing data from the other being ignored[9] (total missing n = 197).

In order to accurately assess whether or not the physician recommendation was for true screening, we created a dichotomous variable combining data on reports of recommendations for testing (as described above) and the impetus for the reported

[9] As an example, a respondent did not provide an answer to the question concerning whether or not a recommendation was received for FOBT testing but indicated that a physician had recommended a form of invasive testing. This respondent's recommendation status was coded as 'yes' despite the missing data for the FOBT question. This same coding was used for CRC testing.

testing. Individuals who reported invasive tests (sigmoidoscopy, proctoscopy, or colonoscopy) were asked, "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 'true 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 this step. If the respondent's answer to this question was coded as missing data but the respondent did report engagement in ACS guideline-adherent FOBT testing, the missing data were ignored[4] (total missing n = 4). Since non-invasive FOBT testing is used exclusively for screening, most reports of FOBT testing was coded as 'true screening.' However, if respondents reported both FOBT and invasive modes of ACS guideline adherent testing, the impetus for testing was categorized solely on the impetus for the invasive testing. A total of 287 respondents fit this criteria: 67 were coded as diagnostic and 220 were coded as true screeners. The resulting dependent variable is a binary measure indicating whether the recommendation was a referral for diagnostic testing or a true screening recommendation.

CRC screening. To assess ACS guideline adherence, we created a dichotomous variable using respondents' self-reports of the type of testing they underwent and the time frame in which the procedure occurred. To assess engagement with non-invasive testing, respondents were asked, "The following questions are about the blood stool or occult blood test, a test to determine whether you have blood in your stool or bowel movement. The blood stool test can be done at home using a kit. You use a stick or

brush to obtain a small amount of stool at home and send it back to the doctor or lab. Have you EVER HAD a blood stool test, using a HOME test kit?" Responses of "yes" or "no" were coded as such, while responses of "refused" or "don't know" were coded as missing data and excluded from analysis. However, in the case of respondents whose response was coded as missing for non-invasive testing but "yes" or "no" for invasive testing (see following paragraph), missing data from the non-invasive testing question were ignored[4] (missing n = 52). Respondents reporting testing also reported testing timelines in one of the following 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). FOBT testing reported to occur within one year of the survey administration was coded as ACS guideline adherent and all FOBT testing was coded as true screening, as it is used exclusively for that purpose.

To assess engagement with invasive testing, respondents were asked, "Have you ever had a sigmoidoscopy, colonoscopy, or proctoscopy? These are exams in which a health care professional inserts a tube into the rectum to look for signs of cancer or other problems." Responses of "yes" or "no" were coded as such, while responses of "refused" or "don't know" were coded as missing data and excluded from analysis. However, in the case of respondents whose response was coded as missing for invasive testing but "yes" or "no" for non-invasive testing (see preceding paragraph), missing data from the invasive testing question were ignored (missing n = 23). Respondents

reporting testing also reported testing timelines in one of the following 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). Invasive testing (sigmoidoscopy, proctoscopy, colonoscopy) was coded as ACS guideline adherent if the testing was reported to be within five, five and ten years, respectively.

To measure reports of true screening versus diagnostic testing behaviors, we created a dichotomous variable combining data on ACS guideline-adherent screening behaviors and the impetus for the reported testing. In order to determine the impetus for the reported testing for 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 'true 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 non-invasive testing was not a "yes"[4] (missing n = 40).

If respondents reported both non-invasive and invasive modes of ACS guideline adherent testing, their answer to the question above with respect to the impetus for the invasive tests determined whether or not it was true screening or diagnostic testing. A total of 287 respondents fit these criteria: 67 were coded as diagnostic and 220 were

coded as true screeners.

Independent variables. The primary independent variable of interest in this study is multiple chronic conditions (MCC), defined as 2 or more chronic conditions. Information on chronic conditions is available from the NHIS in two different ways, and we utilized both methods in creating a count of chronic conditions for each respondent. First, respondents can report the diagnosis of a chronic condition. They are asked, "Have you EVER been told by a doctor or other health professional that you had [condition]?" Second, respondents can report a limitation of daily living and attribute that limitation to the presence of a chronic condition. They are asked, "The next questions ask about difficulties you may have doing certain activities because of a HEALTH PROBLEM. By "health problem" we mean any physical, mental, or emotional problem or illness (not including pregnancy). By yourself, and without using any special equipment, how difficult is it for you to...

...Walk a quarter of a mile - about 3 city blocks?
...Walk up 10 steps without resting?
...Stand or be on your feet for about 2 hours?
...Sit for about 2 hours?
...Stoop, bend, or kneel?
...Reach up over your head?
...Use your fingers to grasp or handle small objects?
...Lift or carry something as heavy as 10 pounds such as a full bag of groceries?
...Push or pull large objects like a living room chair?
...Go out to things like shopping, movies, or sporting events?
...Participate in social activities such as visiting friends, attending clubs and meetings, going to parties?
...Do things to relax at home or for leisure (reading, watching TV, sewing, listening to music)?"

If respondents reported any difficulty with any of these activities, they are asked an open ended question, "What condition or health problem causes you to have difficulty with [the activity]?" Unlike in the first option to reply to a question concerning a

specific condition, they may offer any answer. In addition to the health problem, they report the duration of the condition.

Together, these two sets of questions allowed us to capture data on respondents' reports of the following conditions (listed alphabetically):

Angina
Arthritis
Arthritis (rheumatoid)
Asthma (current)
Benign tumors, cysts
Bronchitis (chronic)
Cancer (excluding colorectal)
Cataracts
Cerebral palsy
Chronic Fatigue
Coronary heart disease
Diabetes
Diabetes (pre-)
Diabetic retinopathy
Emphysema
Epilepsy/seizures
Fibromyalgia
Glaucoma
Gout
Grave's Disease
Hearing/current use of aid
Heart condition (any not mentioned)
Hypertension
Joint condition (other)
Kidney disease (weak or failing)
Liver condition
Lupus
Macular degeneration
Movement disorder
Multiple Sclerosis
Osteoporosis, tendinitis
Stroke
Thyroid (low)
Vision/blindness
Vision/problem seeing w/aids
Weight problem

Chronic disease is characterized by the presence of a disease for 3 months or greater. Therefore, if a respondent reports condition duration of less than three months, the condition was not counted. We utilized coding to prevent duplication of condition counts across the two sections. For example, if a respondent answered that they had been told by a doctor that they have diabetes, and also reported that the diabetes impaired their ability to walk a quarter of a mile, those responses would result in only one chronic condition adding to their count.

We utilized NHIS data directly for variables of 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 doctorate), poverty ratio (tertiles (measured as ratio of family income to poverty threshold): low (under .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 imputations done by the survey administrator. Missing values for all variables except poverty ratio were below 3%, a threshold determined to be acceptable with this sample size.

Statistical analysis.

Data analysis. All data analyses for this chapter were performed with the SAS/STAT statistical software, version 9.3_M1 of the SAS System for Windows, Copyright © 2002-2010 SAS Institute Inc. SAS employs procedures capable of computing appropriate variance estimates for survey estimates generated from analyses of complex sample survey data sets such as the 2008 NHIS (Heeringa, West, & Berglund, 2010; SAS Institute Inc., 2014). Due to the fact that SAS excludes strata with only one PSU including sample from the subpopulation of interest from its variance estimates, we ran selected models in a different software package, Stata 13.0, which enables multiple ad-hoc variance estimation methods for dealing with "singleton" primary sampling unit (PSU) to compare estimated standard errors (StataCorp, 2013). For the logistic regression models including main effect, of 268 estimates only ten had greater than or equal to a 10% difference in standard errors, with the greatest difference at 17%. In most cases, Stata produced higher standard errors; none of these differences resulted in

changes in statistical inferences.

We began our analysis with weighted and unweighted estimates of true screening recommendations and true screening[10] frequencies for each of the independent variables of interest in the full subpopulation as defined above. We then obtained estimates of weighted and unweighted frequencies of both outcomes for each of the independent variables of interest in the smaller subpopulation of ACS guideline-adherent screeners.

We calculated estimates of true screening recommendation and true screening for the full study population by applying the true screening recommendation and true screening rates from the smaller subpopulation to the unadjusted[11] testing rates of the full study population. For example, the unadjusted recommendation rates for the full study population was 61.04%. In the smaller subpopulation, the estimated true screening recommendation rate was 71.81%. We applied this rate to the unadjusted testing rate (61.04% * 71.81%) to calculate an adjusted true screening rate of 43.83%. This same process was repeated for each subgroup of the full study population. Since this calculation reflects the rates of referrals for diagnostic testing of only respondents reporting ACS guideline adherent testing, it is likely a conservative estimate of the true estimate for the full study population. We calculated estimates of the confidence intervals and resulting tests of significance for the adjusted true screening and true screening recommendation rates of the full population. In this process, we were unable

[10] "True screening" refers to pre-symptomatic screening (and not diagnostic testing).

[11] "Unadjusted" rates refers to rates prior to accounting for the impetus for testing (comparable to BRFSS methodology), while "adjusted" rates refers to rates after such accounting (utilizing NHIS methodology).

to account for the fact that we derived the true screening and true screening recommendation rates for the full population from two estimates. This resulted in the possibility of narrower confidence intervals and inferences that were more likely to result in statistically significant values. However, given the precise nature of the subpopulation rate estimates, this issue is likely negligible.

In preparation for fitting multivariate logistic regression models, we examined weighted bivariate associations between independent and dependent variables for each outcome using the Rao-Scott chi-square test to account for the features of the complex sampling design. Our goal with the multivariate logistic regression modeling was not to establish a fitted regression model. Instead, we sought to assess the level of attenuation in the adjusted odds ratios of covariates when MCC was added to the model. Since significant interactions varied across models, making interpretation of odds ratios invalid, we report main effect logistic regression models only.

In the second stage of analysis, we performed a four step fitting of weighted multivariate logistic regression models of main effects to the odds of reporting a true screening recommendation (Subpopulation A) and true screening (Subpopulation B), including variables for sociodemographic factors including sex, race, age, education, poverty ratio, health insurance status, and usual source of health care (Breen et al., 2001; O'Malley et al., 2005; Rosen & Schneider, 2004). We estimated the regression parameters using pseudo-maximum likelihood estimation, due to the complex sampling nature of the NHIS (Heeringa et al., 2010), and we used Taylor Series Linearization to compute design-based estimates of standard errors for the estimated regression

parameters. Hypothesis tests were performed using design-adjusted Wald X^2 tests. We report odds ratios and their 95% confidence intervals.

Results

Please refer to **Table 4.1** for weighted estimates of demographic characteristics and MCC distribution for the full population and each subpopulation. Comparing the full study population to subpopulations A and B, the most significant changes in estimates of the distribution of sociodemographic characteristics were in NHB males (+17% (A) and +14% (B)), the youngest age group (+23% (A) and +21% (B)), those with no college education (+27% and +12% (A) and +29% and +13% (B)), those in the low and high poverty ratio tertile (+24% (A and B) and -14% (A) and -13% (B)), those without insurance coverage (+138% (A) and +115% (B)) and those with no usual source of health care (+222% (A) and 170% (B)).

Compared to the full study population, both subpopulations had higher rates of MCC across sociodemographic categories. Across all populations, compared to non-Hispanic Whites (NHW), non-Hispanic Blacks (NHB) reported higher rates of MCC and across all populations. Female respondents reported higher rates of MCC than males. Rates of MCC were greatest in older respondents, those with the least education, and those in the lowest poverty ratio tertile.

Physician recommendation.

Table 4.2 shows weighted estimates of true screening recommendation rates and rates of referrals for diagnostic testing by MCC status for Subpopulation A. **Table**

4.3 shows weighted estimates of the unadjusted recommendation rate and the adjusted true screening recommendation rates by MCC status for the full study population.

Respondents with MCC reported higher rates of referrals for diagnostic testing than their counterparts (31.34% versus 23.55%) but reported much higher rates of recommendations overall (57.39% versus 44.74%). In the full study population, the estimated rate of true screening recommendations was 43.83% once we accounted for the impetus for the testing. Compared to the full population, the rate for individuals without MCC was slightly lower at 41.20%, while the true screening recommendation rate for those with MCC was higher at 46.28%. The difference in adjusted true screening recommendation rates between those with and without MCC was statistically significant at the $p < .0001$ level.

Table 4.1. Weighted Estimates of Demographic Characteristics and MCC Distribution for Full Population, Subpopulation A and Subpopulation B[A]

	Full Study Population				Subpopulation A				Subpopulation B			
	Demographics n = 4355			MCC Yes	Demographics n = 2008			MCC Yes	Demographics n = 2168			MCC Yes
	n	%	CI of %	% (SE of %)	n	%	CI of %	% (SE of %)	n	%	CI of %	% (SE of %)
0-1 Chronic Conditions (MCC = no)	2073	47.36	(45.62, 49.09)		812	40.42	(38.19, 42.64)		900	41.37	(39.20, 43.54)	
2+ Chronic Conditions (MCC = yes)	2282	52.64	(50.91, 54.38)		1196	59.58	(57.36, 61.81)		1268	58.63	(56.46, 60.80)	
Non-Hispanic White	3528	87.14	(86.11, 88.17)	52.03 (0.95)	1675	88.68	(87.46, 89.89)	58.75 (1.19)	1807	88.58	(87.31, 89.84)	58.02 (1.15)
Non-Hispanic Black	827	12.86	(11.83, 13.89)	56.77 (1.90)	333	11.32	(7.52, 9.67)	66.12 (2.89)	361	11.42	(10.16, 12.69)	63.75 (2.75)
Male	1963	45.78	(44.21, 47.36)	47.06 (1.24)	899	45.48	(43.29, 47.67)	54.32 (1.76)	977	45.85	(43.65, 48.06)	53.90 (1.73)
Female	2392	54.22	(52.64, 55.79)	57.36 (1.23)	1109	54.52	(52.33, 56.71)	63.97 (1.60)	1191	54.15	(51.94, 56.35)	62.71 (1.56)
Non-Hispanic White Male	1619	40.44	(38.87, 42.02)	46.39 (1.36)	767	40.90	(38.72, 43.09)	53.14 (1.90)	832	41.20	(39.01, 43.40)	52.98 (1.86)
Non-Hispanic Black Male	344	5.34	(4.76, 5.92)	52.11 (2.83)	132	4.58	(3.77, 5.38)	64.86 (4.59)	145	4.65	(3.83, 5.47)	62.13 (4.50)
Non-Hispanic White Female	1909	46.70	(45.02, 48.38)	56.92 (1.34)	908	47.77	(45.49, 50.05)	63.55 (1.73)	975	47.37	(45.05, 49.70)	62.40 (1.69)
Non-Hispanic Black Female	483	7.52	(6.68, 8.35)	60.09 (2.64)	201	6.75	(5.70, 7.79)	66.98 (3.61)	216	6.77	(5.72, 7.82)	64.85 (3.51)
Ages 50-54	1565	36.03	(34.61, 37.45)	43.06 (1.51)	575	29.39	(27.18, 31.50)	50.32 (2.28)	629	29.66	(27.54, 31.78)	49.00 (2.22)
Ages 55-59	1511	34.06	(32.47, 35.65)	52.95 (1.48)	739	35.89	(33.60, 38.19)	56.24 (1.87)	796	35.79	(33.54, 38.04)	55.91 (1.81)
Ages 60-64	1279	29.91	(28.39, 31.42)	63.84 (1.41)	694	34.77	(23.67, 28.03)	70.85 (1.86)	743	34.55	(32.24, 36.86)	69.81 (1.79)
Less than High School Diploma	455	10.18	(9.04, 11.32)	67.92 (2.38)	161	8.04	(6.66, 9.41)	80.84 (3.34)	172	7.88	(6.59, 9.17)	80.34 (3.21)
High School Degree or GED	1279	28.55	(27.01, 30.09)	54.87 (1.41)	522	25.57	(23.56, 27.57)	65.71 (2.05)	563	25.21	(23.25, 27.18)	63.88 (1.96)
Some College No Degree or Associate Degree	1341	30.87	(29.33, 32.41)	54.80 (1.48)	629	31.54	(29.25, 33.83)	61.10 (2.20)	674	31.41	(29.18, 33.63)	60.68 (2.13)
Bachelor's Degree	739	17.23	(16.00, 18.48)	45.13 (1.94)	369	18.36	(16.59, 20.12)	51.31 (2.65)	410	18.97	(17.23, 20.72)	50.82 (2.47)
Master's, Professional or Doctorate Degree	519	12.66	(11.57, 13.75)	41.66 (2.43)	321	16.50	(14.83, 18.17)	46.59 (2.80)	343	16.26	(14.59, 17.94)	45.85 (2.76)
Poverty Ratio - Low Tertile	1292	29.67	(27.91, 31.49)	65.70 (1.36)	480	23.88	(21.51, 26.43)	76.70 (2.18)	518	23.87	(21.62, 26.27)	74.49 (2.07)
Poverty Ratio - Medium Tertile	1330	30.55	(29.12, 32.02)	51.96 (1.41)	605	30.12	(27.85, 32.49)	61.29 (2.16)	655	30.23	(28.12, 32.43)	59.98 (2.02)
Poverty Ratio - High Tertile	1732	39.78	(38.00, 41.58)	43.43 (1.35)	924	46.00	(43.46, 48.56)	49.58 (1.15)	995	45.90	(43.51, 48.31)	49.60 (1.65)
Insurance - Not Covered	514	11.22	(10.17, 12.27)	47.48 (2.24)	101	4.72	(3.55, 5.89)	64.17 (4.52)	120	5.21	(4.03, 6.38)	62.12 (4.06)
Insurance - Covered	3836	88.78	(87.73, 89.83)	53.30 (0.92)	1905	95.28	(94.11, 96.45)	59.33 (1.15)	2046	94.79	(93.62, 95.97)	58.46 (1.10)
Living with Marital Spouse or Unmarried Partner	2368	55.50	(53.57, 57.44)	49.12 (1.14)	1173	59.24	(56.65, 61.84)	55.15 (1.51)	1267	59.25	(56.68, 61.82)	54.37 (1.46)
Never Married / Unknown Status	479	10.32	(9.34, 11.30)	48.95 (2.51)	182	8.77	(7.49, 10.05)	53.03 (3.81)	201	8.96	(7.67, 10.25)	52.87 (3.59)
Marital Spouse not in HH or Separated	202	4.43	(3.74, 5.12)	58.62 (4.23)	83	3.90	(2.97, 4.84)	62.62 (6.37)	90	3.87	(2.97, 4.78)	61.16 (6.15)
Widowed or Divorced	1306	29.74	(28,02, 31.47)	59.6 (1.47)	570	28.08	(25.87, 30.29)	70.57 (1.99)	610	27.92	(25.70, 30.13)	69.34 (1.87)
Usual Source of Health Care - No	373	8.59	(7.66, 9.53)	31.51 (2.58)	54	2.67	(1.94, 3.40)	48.57 (6.42)	69	3.18	(2.37, 4.00)	46.91 (6.12)
Usual Source of Health Care - Yes	3920	91.41	(90.48, 92.34)	54.66 (0.93)	1954	97.33	(96.59, 98.06)	59.89 (1.17)	2099	96.82	(96.00, 97.63)	59.06 (1.13)
Physician Recommendation - No	1668	38.96	(37.18, 40.73)	44.31 (1.30)			n/a		162	7.21	(5.97, 8.46)	46.96 (3.97)
Physician Recommendation - Yes	2490	61.04	(59.27, 62.82)	58.47 (1.09)			n/a		2006	92.79	(91.54, 94.03)	59.58 (1.13)

[A] Full population defined as adults ages 50-64 with no history of colorectal cancer. Subpopulation A defined as adults ages 50-64 with no history of colorectal cancer and who reported a physician recommendation to screen and any ACS guideline adherent testing. Subpopulation B defined as adults ages 50-64 with no history of colorectal cancer and who reported any ACS guideline adherent testing.

Note: all values are unweighted n, weighted %, weighted CI of %. Poverty ratio data imputed with Stata.

Table 4.2. Weighted Estimates of Screening Recommendation Rates and Diagnostic Referral Rates for Subpopulation A

	n	Screening Recommendation %	CI of %	Diagnostic Referral %	CI of %	Rao-Scott Chi-Square	Pr > ChiSq
Subpopulation A	2008	71.81	(69.61, 74.01)	28.19	(25.99, 30.39)		
0-1 Multiple Chronic Conditions (MCC = no)	812	76.45	(73.48, 79.42)	23.55	(20.58, 26.51)	16.29	<.0001
2+ Chronic Conditions (MCC = yes)	1196	68.66	(65.90, 71.42)	31.34	(28.58, 34.10)		

Note: unweighted n, weighted % and CI of %

Table 4.3. Weighted Estimates of Unadjusted Recommendation Rates and Adjusted True Recommendation Rates for Full Population

	n	Unadjusted Recommendation Rate %	CI of %	Tests of Significance[1]		Adjusted True Recommendation Rate %	CI of %	Tests of Significance[2]	
Full Study Population (n missing = 197)	4158	61.04	(59.27, 62.82)			43.83			
0-1 Multiple Chronic Conditions (MCC = no)	1965	53.89	(51.32, 56.45)	79.88	<.0001	41.20	(39.22, 43.17)	4.50	<.0001
2+ Chronic Conditions (MCC = yes)	2193	67.41	(65.36, 69.45)			46.28	(44.87, 47.69)		

[1] Rao-Scott Chi-Square and Pr > ChiSq calculated with SAS.

[2] Confidence intervals, t-test and P>|t| calculated for overall effects with Stata

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

* applying true screening recommendation and true screening rates from subpopulation A to unadjusted rates from full study population.

Table 4.4 shows estimates of unadjusted and adjusted CRC testing and screening rates by MCC status for Subpopulation B. **Table 4.5** shows estimates of unadjusted and adjusted CRC testing and screening rates by MCC status for the full study population.

Table 4.4. Weighted Estimates of True Screening Rates and Diagnostic Testing Rates for Subpopulation B

	n	True Screening %	CI of %	Diagnostic Testing %	CI of %	Rao-Scott Chi-Square	Pr > ChiSq
Subpopulation B	2168	79.94	(77.96, 81.92)	20.06	(18.08, 22.04)		
0-1 Multiple Chronic Conditions (MCC =	900	84.08	(81.47, 86.68)	15.92	(13.32, 18.53)	16.37	<.0001
2+ Chronic Conditions (MCC = yes)	1268	77.03	(74.49, 79.57)	22.97	(20.43, 25.51)		

Note: unweighted n, weighted % and CI of %

Table 4.5. Weighted Estimates of Unadjusted Testing Rates and Adjusted True Screening Rates for Full Population

	n	Unadjusted Testing Rate %	CI of %	Tests of Significance[1]		Adjusted True Screening Rate %	CI of %	Tests of Significance[2]	
Full Study Population (n missing = 40)	4315	51.39	(49.63, 53.15)			41.08			
0-1 Multiple Chronic Conditions (MCC = no)	2058	44.74	(42.36, 47.12)	68.41	<.0001	37.62	(35.62, 39.62)	5.31	<.0001
2+ Chronic Conditions (MCC = yes)	2257	57.39	(55.17, 59.62)			44.21	(42.50, 45.93)		

[1] Rao-Scott Chi-Square and Pr > ChiSq calculated with SAS.
[2] Confidence intervals, t-test and P>|t| calculated for overall effects with Stata
Note: unweighted n, weighted % and CI of %
* applying true screening recommendation and true screening rates from subpopulation to unadjusted population rates

In the subpopulation of respondents ages 50-64 reporting ACS guideline adherent screening, true screening rates were higher in the general population than in those with MCC (79.94% versus 77.03%). The highest rates of true screening were in those with no MCC at 84.08%. The difference in rates between those with and without MCC was statistically significant ($p < .0001$).

Applying the subpopulation true screening rates to the full study population, true screening rates in the full study population were estimated to be 41.08% after accounting for the impetus for testing. Similar to recommendation results, the adjusted population rate of true screening was lower in those without MCC (37.62%) and higher

in those with MCC (44.21%). After adjustment, only those with zero chronic conditions had true screening rates lower than the full study population, due in large part by the higher rate of any ACS guideline adherent testing in respondents with MCC. Again, the difference between those with and without MCC was statistically significant ($p < .0001$).

Multivariate Logistic Regression Modeling

Table 4.6 provides estimated odds ratios from multivariate logistic regression models predicting true screening recommendations in Subpopulation A. **Table 4.7** provides estimated odds ratios from multivariate logistic regression models predicting true screening in Subpopulation B.

Our primary interest in these analyses was to determine whether MCC accounts for any of the effects demonstrated in the CRC screening recommendation and CRC screening literature to date, namely effects of race/ethnicity, age, sex, education, poverty ratio, insurance coverage, and a usual source of health care. For these analyses, we used multivariate logistic regression to compare estimates of these effects in modeling both outcomes without MCC and with MCC. In contrast to other regression approaches, we were not seeking to identify a 'best fit' model, therefore tests of interactions were not performed and all main effects remained in both models irrespective of statistical significance. Despite this not being a 'best fit' model, it is of interest to note that controlling for other covariates, MCC was statistically significant at the $p < .02$ level and negatively associated with the odds of reporting a true screening recommendation, with an adjusted odds ratio of 0.754 (CI [0.605, 0.941], $p = 0.0123$) and statistically significant at the $p < .01$ level and negatively associated with the odds of

reporting true screening, with an adjusted odds ratio of 0.642 (CI [0.504, 0.818], p = 0.0003). All odds ratio estimates reported reflect statistical adjustment for other covariates in the multivariate model.

True Screening Recommendations.

In the models predicting true screening recommendations, comparing models without and with MCC we find that MCC has little effect on the adjusted odds ratios of other covariates. Race/ethnicity and age effects increased slightly with MCC in the model but both remained statistically significant across models. NHB were more likely to report a true screening recommendation than their NHW counterparts. Older age was only slightly associated with the outcome. The effect of sex increased slightly in models with MCC and remained statistically significant. Males were more likely than females to report a true screening recommendation.

The effects of characteristics of socioeconomic status varied across models. Global tests for education showed that the effect was not statistically significant in either model, while global tests for poverty ratio showed stable statistical significance across models, with a stronger effect on the outcome in the model with MCC. As the ratio of family income to poverty threshold increased, so, too, did odds of reporting a true physician recommendation. Neither insurance coverage nor source of usual care was statistically significant in any model and both varied only slightly across models without and with MCC.

True Screening.

Comparing models without and with MCC, we found again that MCC had little effect on the adjusted odds ratios of other covariates in the model. Race/ethnicity was significant at the $p < .001$ level across all models, and the effect was slightly higher in models with MCC versus without MCC. NHB were more likely than NHW to report true screening across both models. Age remained statistically significant across models, with mild positive effects on the likelihood of true screening as age increases. Sex effects were slightly attenuated with MCC in the model but remained statistically significant across all models. Males were more likely to report true screening than females.

Characteristics of socioeconomic status had varying levels of association with true screening. Global tests for education show that the effect was highly significant in the model without MCC, but no longer statistically significant in the model with MCC. Global tests for poverty ratio show stable statistical significance across models, with an attenuated effect on the outcome in models with MCC. Finally, insurance coverage was not statistically significant in any model.

Features of the health care system had positive associations with the odds of true screening. Although the presence of a usual source of health care was statistically significant at the $p < .02$ level in each of the models, however, a physician recommendation was not statistically significant in predicting true screening in any of the regression models.

Table 4.6. Multivariate Main Effects Models Predicting True Recommendation Without and With MCC in Model

		Without MCC in Model					With MCC in Model					
Predictor*	**Category**	**OR**	**95% Confidence Limits (OR)**		**t for H0**	**Pr > |t|**	**OR**	**95% Confidence Limits (OR)**		**t for H0**	**Pr > |t|**	**% Change (OR)**
intercept		0.469	0.232	0.949	-2.1	0.0354	0.568	0.277	1.168	-1.54	0.124	21%
MCCCat	*Yes*						0.754	0.605	0.941	-2.5	0.0123	
Race/Ethnicity	Non-Hispanic Black	1.749	1.279	2.392	3.5	0.0005	1.766	1.291	2.414	3.56	0.0004	1%
Age	Centered Age (57)	1.048	1.022	1.074	3.66	0.0003	1.053	1.026	1.080	3.93	<.0001	1%
Sex	Male	1.605	1.302	1.979	4.43	<.0001	1.573	1.278	1.937	4.28	<.0001	-2%
Education	High School Graduate or GED	1.756	1.121	2.749	2.46	0.0139	1.726	1.102	2.703	2.38	0.0172	-2%
	Some College (no degree) or Associates Degree	1.608	1.065	2.426	2.26	0.0238	1.573	1.042	2.374	2.16	0.0309	-2%
	Bachelor's Degree	1.534	0.920	2.558	1.64	0.1008	1.480	0.888	2.468	1.51	0.1323	-3%
	Master's, Professional or Doctorate	1.517	0.917	2.509	1.62	0.1044	1.438	0.868	2.383	1.41	0.1586	-5%
Poverty Ratio	Medium Tertile	1.539	1.108	2.137	2.57	0.0101	1.485	1.064	2.073	2.33	0.02	-3%
	High Tertile	2.449	1.782	3.366	5.53	<.0001	2.317	1.672	3.210	5.05	<.0001	-5%
Insurance	Covered	1.546	0.941	2.541	1.72	0.0858	1.578	0.953	2.612	1.77	0.076	2%
Usual Source of Health Care	Yes	1.064	0.573	1.974	0.2	0.8444	1.096	0.583	2.059	0.29	0.7755	3%

Note: Without MCC in model, global test for Education: F = 1.65 Pr > F = 0.1585; for Poverty Ratio: F = 15.60 Pr > F <.0001. With MCC in model, Global test for Education: F = 1.58 Pr > F = 0.1772; for Poverty Ratio: F = 13.20 Pr > F < .0001. No changes in statistical significance of main effects between models at p < .05 level.

* Referent group for MCC is No, for Race/Ethnicity is Non-Hispanic White, for Sex is Female; for Education, < High School; for Poverty Ratio, Low Tertile; for Insurance, Not Covered; for Usual Source of Health Care, No.

Table 4.7. Multivariate Main Effects Models Predicting True Screening Without and With MCC in Model

		Without MCC in Model					With MCC in Model					
Predictor*	**Category**	**OR**	**95% Confidence Limits (OR)**		**t for H0**	**Pr > |t|**	**OR**	**95% Confidence Limits (OR)**		**t for H0**	**Pr > |t|**	**% Change (OR)**
intercept		0.446	0.208	0.956	-2.08	0.0380	0.576	0.268	1.236	-1.42	0.1568	29%
MCCCat	*Yes*						0.642	0.504	0.818	-3.59	0.0003	
Race/Ethnicity	Non-Hispanic Black	1.956	1.348	2.838	3.53	0.0004	1.974	1.355	2.874	3.55	0.0004	1%
Age	Centered Age (57)	1.052	1.023	1.082	3.59	0.0003	1.061	1.031	1.091	4.09	<.0001	1%
Sex	Male	1.581	1.240	2.016	3.7	0.0002	1.539	1.207	1.960	3.48	0.0005	-3%
Education	High School Graduate or GED	1.570	1.005	2.454	1.98	0.0476	1.508	0.963	2.362	1.79	0.0727	-4%
	Some College (no degree) or Associates Degree	1.556	1.024	2.367	2.07	0.0386	1.494	0.985	2.266	1.89	0.0588	-4%
	Bachelor's Degree	1.663	1.011	2.736	2	0.0453	1.559	0.946	2.571	1.74	0.0816	-6%
	Master's, Professional or Doctorate	1.356	0.810	2.271	1.16	0.2469	1.233	0.743	2.046	0.81	0.418	-9%
Poverty Ratio	Medium Tertile	1.374	0.991	1.905	1.91	0.0563	1.309	0.941	1.820	1.6	0.1098	-5%
	High Tertile	1.919	1.377	2.675	3.86	0.0001	1.785	1.273	2.503	3.37	0.0008	-7%
Insurance	Covered	1.302	0.797	2.129	1.05	0.2919	1.327	0.799	2.202	1.09	0.274	2%
Usual Source of Health Care	Yes	1.965	1.115	3.463	2.34	0.0194	2.089	1.167	3.738	2.48	0.0132	6%
Physician Recommendation	Yes	1.326	0.873	2.016	1.32	0.1860	1.398	0.920	2.124	1.57	0.1163	5%

Note: Without MCC in model, global test for Education: F = 8.81 Pr > F <.0001; for Poverty Ratio: F = 11.12 Pr > F <.0001. With MCC in model, global test for Education: F = 1.28 Pr > F = 0.2753; for Poverty Ratio: F = 5.86 Pr > F = 0.0031. Other than the change in statistical significance for Education, no changes in statistical significance of main effects between models at p < .05 level.

* Referent group for MCC is No, for Race/Ethnicity is Non-Hispanic White, for Sex is Female; for Education, < High School; for Poverty Ratio, Low Tertile; for Insurance, Not Covered; for Usual Source of Health Care, No; for Physician Recommendation, No.

Discussion

Our primary goal in this study was to examine the role that MCC has on two behavioral outcomes in CRC screening: a physician recommendation to screen and engagement with screening itself. Although we found that rates of referrals for diagnostic testing and diagnostic testing itself were higher in those with MCC, overall rates of both outcomes are unexpectedly higher in those with MCC when compared to those without. Our secondary goal was to explore how the presence of MCC in multivariate models changes the association of key predictors with the odds of both outcomes. Multivariate logistic regression models revealed that although MCC does not account for much of the effect of sociodemographic variables, it is negatively associated with both outcomes, adjusting for other covariates in the model.

Comparing our results to those described earlier, we hypothesize that the contradictory outcomes in rates are due to the difference in outcome variable specification. We have estimated rates of true screening recommendations and true screening for the full population, while other studies typically conflate diagnostic testing (and referrals) with true screening (and recommendations). And while our multivariate models were not fitted to a 'best fit' model, the adjusted odds ratios associated with MCC in models of both outcomes merit the attention of future research.

This study is subject to some unavoidable limitations. First, we analyzed data collected in 2008, as analyses began as the 2010 survey was being released. 2008 data remained the best option, however, due to its specificity with its questions pertaining to physician recommendations. A general limitation of using NHIS data is the inability to

conduct analyses by geographic region with much certainty; other data collected by state-level entities such as the Department of Public Health or data that are representative to each state's population (including BRFSS data) would be more appropriate. CRC screening rates vary considerably by state, and within state by county; this variation should be considered in future studies of CRC screening (Rim et al., 2011; Weir et al., 2003).

Common to survey data, NHIS data are subject to recall bias on the part of the respondents for both dependent and independent variables. Prior studies have found that self-report of CRC screening behaviors and MCC are demonstrated to be moderate to high in both sensitivity and specificity with no clear patterns in differences in the accuracy of self-report by age, sex, race or family history (Baier et al., 2000; Martin, Leff, Calonge, Garrett, & Nelson, 2000). 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 (Centers for Disease Control and Prevention, 2012). In addition, there may be differential bias in comprehension of survey questions. Differential interpretation of questions by race, gender, and socioeconomic status may explain some of our findings. However, the pivotal questions of interest were follow-up questions to screening responses. These questions received very little attention in cognitive testing, as the majority of the cognitive testing subjects (which numbered only 9 over age 40) indicated they did not undergo CRC testing, resulting in little to no cognitive testing on the follow-up questions (B. Taylor, personal communication, June 17, 2014). Our

practice of deriving full population rates using estimates from unadjusted testing rates and adjusted true screening and true screening recommendation rates of the subpopulation resulted in estimates that may have confidence intervals narrower than what they should be. This is due to the fact that we are unable to account for the fact that we are treating the rates used in estimation as fixed when it is, in fact, an estimate. The confidence intervals for the true screening and true screening recommendation rates of the full population exclude the variance estimates of the true screening and true screening recommendation rates of the study's subpopulation. Therefore, the resulting tests of statistical significance will be more likely to show statistical significance than they would have had the results not been derived from two estimates. In the future, researchers should explore the use of jackknifing as a tool to incorporate the variance of both estimates when calculating confidence intervals and resulting statistical tests for the full population.

The artificiality of the measurement and cutpoints of MCC may be seen as a limitation to this study, as a chronic disease count does not reveal information to support or refute the hypothesized mechanisms of influence, including competing demands in the clinical encounter and diminished individual resources to complete screening. However, in an effort to be consistent with existing literature in other cancer screening modalities as well as federal initiatives, it will be important to maintain the measurement and cutpoints as proposed, and allow future qualitative and quantitative research to determine alternative ways of measuring and operationalizing chronic disease burden in this population.

Finally, no measure of availability of services is being incorporated into these analyses, as variables of interest are limited to those at the individual level. Clearly, structural factors such as health care policy (including mandated insurance coverage of CRC screening), physician supply and market conditions (e.g. availability of technology) are relevant to both screening recommendations and services; however, these structural factors are beyond the scope of the present study.

To our knowledge, this is the first study to explore the impact of MCC on CRC screening outcomes using a nationally representative survey sample. The results of this study are highly relevant to both researchers and practitioners. As we face a growing population with MCC in the future, understanding how MCC creates individual and institutional context in cancer screening behaviors will become critical. For researchers, it will be important to develop and test measures that can accurately assess chronic disease burden. Studies on MCC burden show linear associations between MCC and individual economic burdens, frequency of health care interaction and other related outcomes (Anderson, 2010; US Department of Health & Human Services, 2011). However, there is little consistency across studies with respect to the approach to measuring chronic disease burden. A measure that can capture the burden of chronic disease on individuals' social, financial, and physical resources and on the ways in which they engage with the health care system will be useful in future studies in cancer screening and other preventative health behaviors.

At the institutional level, our evidence supports consideration for shifting the responsibility of physician recommendation delivery to all health care providers. In

most cases, this responsibility of cancer screening recommendations falls solely on primary care physicians, which is useful for patients using primary care for preventative health care. However, we know that patients with MCC use the health care system differently and more frequently than their healthier counterparts including a shift in use from primary care to more frequent specialty or emergency care (Kane, Priester, & Totten, 2005; Vogeli et al., 2007). Furthermore, when patients with MCC visit their primary care provider, the short time afforded to their visit may be spent in discussions of issues related to their multiple chronic conditions such as visits with specialists, new medications, mental health concerns and other disease-related care, rather than preventative care (Anderson, 2010; Gonzalez et al., 2001; Messina, 2011). The potential for improvement in recommendation and screening rates in this growing population starts with increasing the opportunities for cancer screening counseling by promoting recommendations outside of primary care including emergency care, specialty care, and non-traditional sources of health care.

References

AARP. (2010, February). Preventive Health Screenings Among Midlife and Older Adults. *AARP*. Retrieved May 5, 2011, from http://www.aarp.org/health/fitness/info-02-2010/prevmed.html

American Cancer Society. (2012). Colorectal Cancer. Retrieved January 2, 2003, from http://www.cancer.org/acs/groups/cid/documents/webcontent/003096-pdf.pdf

Amos, C., Bosken, C., Soliman, A., & Frazier, M. (2005). Risk Factors for Colorectal Cancer Susceptibility. In J. Ajani, P. Lynch, N. Janjan, & S. Curley (Eds.), *Gastrointestinal Cancer* (pp. 67–84). Springer New York. Retrieved from http://dx.doi.org/10.1007/0-387-27285-2_6

Anderson, G. F. (2010). *Chronic care: making the case for ongoing care*. Robert Wood Johnson Foundation.

Baier, M., Calonge, N., Cutter, G., McClatchey, M., Schoentgen, S., Hines, S., … Ahnen, D. (2000). Validity of self-reported colorectal cancer screening behavior. *Cancer Epidemiology Biomarkers & Prevention*, *9*(2), 229–232.

Becker, E., Elliott, H., Griffith, D., Alexander, G., & Morris, A. (2010). Qualitative Research and Racial Disparities in Screening Behaviors: Understanding Barriers and Facilitators of Colorectal Cancer Screening. Presented at the Academy Health Annual Research Meeting, Boston, MA.

Brawarsky, P., Brooks, D., Mucci, L., & Wood, P. (2004). Effect of physician recommendation and patient adherence on rates of colorectal cancer testing. *Cancer Detection and Prevention*, *28*, 260–268.

Breen, N., Wagener, D., Brown, M., Davis, W., & Ballard-Barbash, R. (2001). Progress in cancer screening over a decade: results of cancer screening from the 1987, 1992, and 1998 National Health Interview Surveys. *JNCI Journal of the National Cancer Institute*, *93*, 1704.

Brody, J. E. (2011). Tackling care as chronic ailments pile up. *New York Times, February*, *21*.

Centers for Disease Control and Prevention. (2012, June 13). NHIS - About the National Health Interview Survey. Retrieved March 14, 2014, from http://www.cdc.gov/nchs/nhis/about_nhis.htm

Cummings, J., & Jackson, B. (2008). Race, Gender, and SES Disparities in Self-Assessed Health, 1974-2004. *Research on Aging*, *30*, 137.

Fleming, S. T., Pursley, H. G., Newman, B., Pavlov, D., & Chen, K. (2005). Comorbidity as a predictor of stage of illness for patients with breast cancer. *Medical Care*, *43*(2), 132–140.

Gonzalez, E. C., Ferrante, J. M., Van Durme, D. J., Pal, N., & Roetzheim, R. G. (2001). Comorbid illness and the early detection of cancer. *Southern Medical Journal*, *94*, 913&hyhen; 920.

Gross, C., Andersen, M., Krumholz, H., McAvay, G., Proctor, D., & Tinetti, M. (2006). Relation between Medicare screening reimbursement and stage at diagnosis for older patients with colon cancer. *JAMA*, *296*, 2815.

Heeringa, S. G., West, B. T., & Berglund, P. A. (2010). Applied survey data analysis.

Holden, D. J., Jonas, D. E., Porterfield, D. S., Reuland, D., & Harris, R. (2010). Systematic review: enhancing the use and quality of colorectal cancer screening. *Annals of Internal Medicine*, *152*, 668.

House, J., Kessler, R., & Herzog, A. (1990). Age, socioeconomic status, and health. *The Milbank Quarterly*, 383–411.

Inter-University Consortium for Political and Social Research. (2012). National Health Interview Survey, 2008 (ICPSR 27341). Retrieved from http://www.icpsr.umich.edu/icpsrweb/ICPSR/studies/27341?q=27341&searchSource=icpsr-landing

Kane, R. L., Priester, R., & Totten, A. M. (2005). *Meeting the challenge of chronic illness*. JHU Press.

Kiefe, C. I., Funkhouser, E., Ph, D., Fouad, M. N., & May, D. S. (1998). Chronic disease as a barrier to breast and cervical cancer screening. *Journal of General Internal Medicine*, *13*, 357.

Kington, R. S., & Smith, J. P. (1997). Socioeconomic status and racial and ethnic differences in functional status associated with chronic diseases. *American Journal of Public Health*, *87*(5), 805–810.

Klabunde, C., Breen, N., & Meissner, H. (2005). Use of Colonoscopy for Colorectal Cancer Screening. *Cancer Epidemiology Biomarkers & Prevention*, *14*, 2279.

Ko, C. W., Kreuter, W., & Baldwin, L. M. (2005). Persistent demographic differences in colorectal cancer screening utilization despite Medicare reimbursement. *BMC Gastroenterol*, *5*, 10. doi:10.1186/1471-230X-5-10

Martin, L. M., Leff, M., Calonge, N., Garrett, C., & Nelson, D. E. (2000). Validation of self-reported chronic conditions and health services in a managed care population. *American Journal of Preventive Medicine*, *18*(3), 215–218.

Messina, C. R. (2011). Barriers to Colorectal Cancer Screening: Patient, Physician, and System Factors. In *Colorectal Cancer Screening* (pp. 57–66). Springer.

O'Malley, A., Forrest, C., Feng, S., & Mandelblatt, J. (2005). Disparities despite coverage: gaps in colorectal cancer screening among Medicare beneficiaries. *Archives of Internal Medicine*, *165*, 2129.

Potter, J. D., & Hunter, D. (2002). Colorectal Cancer. In H.-O. Adami, D. Hunter, & D. Trichopoulos (Eds.), *A Textbook of Cancer Epidemiology* (pp. 188–211). New York: Oxford University Press.

Rim, S. H., Joseph, D. A., Steele, C. B., Thompson, T. D., Seeff, L. C., & (CDC), C. for D. C. and P. (2011). Colorectal cancer screening—United States, 2002, 2004, 2006, and 2008. *MMWR Surveill Summ*, *60*(suppl 15), 42–46.

Rosen, A., & Schneider, E. (2004). Colorectal cancer screening disparities related to obesity and gender. *Journal of General Internal Medicine*, *19*, 332–338.

SAS Institute Inc. (2014). SAS (Version 9.3). Cary, NC: SAS Institute Inc.

Schneider, E. (2009). Disparities and Colorectal Cancer. In H. K. Koh (Ed.), *Toward the Elimination of Cancer Disparities* (pp. 161–178). New York: Springer.

Shokar, N. K., Carlson, C. A., & Weller, S. C. (2008). Factors associated with racial/ethnic differences in colorectal cancer screening. *The Journal of the American Board of Family Medicine*, *21*, 414–426.

Singh-Manoux, A., Ferrie, J. E., Chandola, T., & Marmot, M. (2004). Socioeconomic trajectories across the life course and health outcomes in midlife: evidence for the accumulation hypothesis? *International Journal of Epidemiology*, *33*(5), 1072–1079.

Smedley, B., Stith, A., & Nelson, A. (2002). *Unequal treatment: confronting racial and ethnic disparities in health care*. Natl Academy Pr.

StataCorp. (2013). Stata Version 13 (Version Stata Statistical Software: Release 13). College Station, TX: StateCorp LP.

Subramanian, S., Klosterman, M., Amonkar, M., & Hunt, T. (2004). Adherence with colorectal cancer screening guidelines: a review. *Preventive Medicine*, *38*, 536–550.

US Department of Health & Human Services. (2010, February). NIH Colorectal Cancer Screening Conference - Panel Statement. Retrieved April 16, 2014, from http://consensus.nih.gov/2010/colorectalstatement.htm

US Department of Health & Human Services. (2011). *Healthy People 2020*.

Vogeli, C., Shields, A. E., Lee, T. A., Gibson, T. B., Marder, W. D., Weiss, K. B., & Blumenthal, D. (2007). Multiple chronic conditions: prevalence, health consequences, and implications for quality, care management, and costs. *Journal of General Internal Medicine*, *22*(3), 391–395.

Wee, C., McCarthy, E., & Phillips, R. (2005). Factors associated with colon cancer screening: the role of patient factors and physician counseling. *Preventive Medicine*, *41*, 23–29.

Weir, H. K., Thun, M. J., Hankey, B. F., Ries, L. A. G., Howe, H. L., Wingo, P. A., … Edwards, B. K. (2003). Annual Report to the Nation on the Status of Cancer, 1975-2000, Featuring the Uses of Surveillance Data for Cancer Prevention and Control. *J. Natl. Cancer Inst.*, *95*, 1276–1299. doi:10.1093/jnci/djg040

Williams, D., & Jackson, P. (2005). Social sources of racial disparities in health. *Health Affairs*, *24*, 325.

CHAPTER FIVE

Implications for Researchers, Practitioners, and Policymakers

Colorectal cancer is largely preventable through the early detection of pre-cancerous polyps, yet estimates of population screening rates are far below a level that will prevent an increasing burden of morbidity and mortality from the disease. Accurate assessment of screening is critical as we design interventions to promote screening and improve population health.

An often-cited limitation of CRC screening research is the conflation of diagnostic testing and true screening. To my knowledge, this is the first study to measure the impact of this limitation on population screening rates, rates of physician recommendations and across sociodemographic groups. I hypothesized that conflation of diagnostic testing and true screening would result in overestimated screening rates and that the overestimation would be higher in those with less socioeconomic and social privilege. The results support these hypotheses and demonstrate not only significant overestimation of rates, but also substantial variation in rates of overestimation across levels of privilege. For both screening and recommendations, it is clear from these studies that the omission of two questions, 'why did you engage in screening?' and 'why did your physician recommend the screening test?', results in

gross overestimation of screening and recommendation rates and false impressions about predictors of screening and screening recommendations.

To improve public health, we must increase rates of true screening. Conflating diagnostic testing and true screening when researchers specify behavioral outcomes is misguided, as these are distinct behaviors. Diagnostic testing is important to determine the causes of symptoms; however, true screening at the pre-symptomatic stage is what will improve population health. Measuring both and calling them both screening is wrong, and leads to highly problematic overestimation of true screening rates. By failing to distinguish the two behaviors, accepting the BRFSS limitation, and disseminating rates, trends and disparities data based on it, the research community is misinforming researchers, practitioners, policymakers and the American public about the nature of the underutilization of screening.

When researchers conflated diagnostic testing and true screening in existing research, their results masked several unexpected findings discovered in these studies. Contrary to the results noted in published literature that shows advantages for NHW and women, my study shows that NHW women engaged in the highest rate of diagnostic testing. Using BRFSS methodology and ignoring the distinction between diagnostic testing and true screening, the advantage over NHB women significantly less than when the outcome is correctly specified (9.98% versus 16.78%). This type of error is easily corrected at the survey level for future research, but the mistake is not easily adjusted as we look back at past research, since the rate of overestimation due to conflation varies greatly across subgroups.

I also found that the misspecification of the outcome significantly changed the degree to which established predictors of screening were predictive of true screening. I hypothesized that predictors of the outcome would change in magnitude when the model is correctly specified, but I did not expect that the model would change so drastically. Specifically, both aspects of health care delivery – the physician recommendation and a usual source of health care – were predictive of any ACS guideline adherent testing but were insignificant in the model of true screening behaviors. Overall, the results challenge the status quo of accepting limitations that are preventable by more effective survey design. Some limitations are unavoidable – but the conflation of diagnostic testing with true screening is solved by the addition of two question – 'why did you engage in screening?' and 'why did your physician recommend the screening test?'.

The results of both studies reveal that NHW females are the most likely to engage in, and receive referrals for, diagnostic testing. These results are perplexing, as they are inconsistent with research on health care utilization and patterning of CRC disease. With respect to health care utilization and gender, research has shown that women are more likely to use preventative and curative health care services than their male counterparts (Green & Pope, 1999; Bertakis et al., 2000). I expected that females' engagement with true CRC screening, not diagnostic testing, might be higher. Instead, I found that females reported higher rates of diagnostic testing than their male counterparts. Higher diagnostic testing rates in women across both race/ethnicity groups may suggest differences in diagnostic testing patterns due to physiological

differences (including higher risk of abdominal pain and chronic GI disease). In health care utilization patterning, we typically find a higher level of engagement with the health care system in NHW when compared to NHB, particularly in older adults. I expected that this would result in higher rate estimates of true screening; instead, I found higher rates of diagnostic testing and referrals in NHW despite these high diagnostic testing rates, they had higher true screening rates than NHB females. Although these results may reflect gender-specific differences in study measure conceptualizations, specifically differences between true screening and diagnostic testing, these differences would not likely account for the combination of gender and race/ethnicity contributions to the pre- and post-adjustment rate differences. Instead, we hypothesize that differential perceptions of risk across gender and race/ethnicity groups may be a significant contributor to these findings. Prior research has often focused on the higher burden of CRC in males and in NHB populations. This may be contributing to a much lower perception of risk, both on the part of the patient and her health care provider, for NHW females. Therefore, rather than engaging in true screening, NHW females are more likely than their male and NHB counterparts to engage in diagnostic testing after symptoms are present. It will be important to explore perceptions of risk in these populations, and in their health care providers, as we seek to determine best practices for multilevel interventions to increase true screening rates.

Although this study sample is comprised of individuals without a history of CRC, I would expect the patterning of diagnostic testing versus true screening behaviors to correspond with patterns by gender and race/ethnicity in CRC incidence and stage at

diagnosis. For example, CRC incidence and mortality are significantly higher for men than women across race/ethnicity groups (American Cancer Society, 2011). Based on morbidity and mortality data, I would therefore expect that males engage in more diagnostic testing than their female counterparts. Instead, the estimates show that females are more likely to engage in, and receive referrals for, diagnostic testing. With respect to gender and stage at diagnosis, there is little evidence of differences between race/ethnicity groups; I only find differences between groups by race/ethnicity (American Cancer Society, 2011). This suggests that the higher rate of diagnostic testing in NHW females may be the result of other factors and merits further study.

NHB suffer a significantly higher burden of CRC incidence and mortality when compared to their NHW counterparts; compared to Whites, NHB are less likely to be diagnosed at the earliest, most treatable, stage of CRC disease (American Cancer Society, 2011). I would therefore expect to see higher diagnostic testing estimates in NHB, as screening is most often associated with earlier stage at diagnosis. Instead, the estimates reveal higher rates of diagnostic testing in NHW. Despite the higher rates of diagnostic testing, true screening rates of NHW remained higher than NHB, but the disparity was much less after adjustment for the impetus for testing. These results suggest that the diagnostic testing in NHW may not contribute to increased disease risk and instead may represent a benign difference in diagnostic test utilization. However, this cannot be fully explored until we collect and analyze screening/diagnostic testing data on individuals who have been diagnosed with CRC.

A full explanation for both of these unexpected results can not be ascertained through this type inquiry. Instead, the results provide strong motivation for future work utilizing both qualitative and quantitative analytic approaches. Qualitatively, understanding psychosocial and instrumental barriers and facilitators across demographic groups is critical to the design and implementation of effective interventions. Interventions based on prior 'screening' data need to be reconsidered, as these data provided false estimates across sociodemographic groups and under- and overestimated differences between and within groups. Engaging a wide range of subjects in qualitative research spanning sociodemographic groups and intersections (e.g. race/ethnicity and poverty ratio) will help as we attempt to fully elucidate the mechanisms contributing to our findings, particular those findings that are contradictory to the help-seeking and CRC risk literature. In addition to suggestions already articulated herein, quantitative approaches also require a more critical approach to subject selection, ensuring adequate representation from respondents across sociodemographic groups and intersections of race/ethnicity and socioeconomic status. Furthermore, surveys of health care providers (including providers outside of primary care) will be useful as we determine if perceived risk and perceived adherence to screening recommendations contribute to the patterns in our results.

My third study is in response to the narrow lens through which we view predictors of screening and recommendations. A focus on sociodemographic characteristics of both patient and physician has ignored the growing reality of older adults who face multiple chronic conditions that can interfere with preventative care at

the individual and institutional level. I hypothesized that MCC would be negatively associated with the odds of screening recommendations and screening itself. My results show, however, that although population rates were higher in those with MCC than the population rate and their healthier counterparts, MCC was negatively associated with both outcomes in multivariate models. Considering the results of earlier chapters showing that the physician's role is not as important as once thought, the impact of MCC on reduced odds of screening in multivariate models may show that individual level, not institutional, factors are likely to influence screening rates in individuals with MCC.

The measures utilized in these studies designating testing as true screening or diagnostic in nature are subject to limitations that merit further exploration. Patterning of testing type across demographic categories may reflect differing interpretations of the survey questions. However, the patterning of diagnostic testing across the age continuum is still highly problematic and less likely to be the result of such a limitation. Recall issues related to the timing of the testing may also affect these estimates. For example, a respondent's odds of remembering the impetus for testing nine years earlier may be less reliable than another's recall of a more recent test. Unfortunately, cognitive testing on screening questions (and particularly their follow-up questions) is limited. Data provided in personal correspondence from the National Center for Health Statistics (NCHS) shows that questions related to testing were tested on few respondents, ranging from one to nine depending on modality. The follow-up question inquiring about the impetus for testing was only tested on two individuals, neither of

who struggled to answer the question (B. Taylor, personal communication, June 17, 2014). Future research on variation in interpreting this question is needed to understand if any of the findings reported herein are due to measurement issues.

Despite these limitations, the results of these three studies challenge the research community in several ways. The results speak loudly to the need for changes in survey methodology in cancer screening. We must avoid the limitation of conflating diagnostic testing and true screening by asking 'why?' in cancer screening research. Immediate changes in BRFSS methodologies are required so that we may begin to accurately assess national- and state-level estimates of screening patterns. We must critically examine assumptions that have been made based on BRFSS data including data on trends, disparities, and health care quality benchmarks. We must also reassess the importance and relevance of the physician recommendation, as it is not predictive of screening when the outcome is correctly specified. Further research is necessary to examine the nature of the relationship between MCC and CRC screening to determine what aspects of MCC promote true screening and what aspects of MCC interfere with appropriate and timely screening. Finally, cancer registries should consider the addition of information on the method by which patients have been diagnosed with CRC. The results of my studies are helpful to assess relationships among variables in individuals without cancer, but understanding rates of diagnostic testing versus true screening and comorbidities and their associations with stage at diagnosis will prove to be valuable as we seek to improve public health.

My results should also help inform the activities and focus of practitioners. The health care system is faced with an enormous challenge of the generation entering older adulthood. Unlike the generation that preceded them, this generation faces a burden of chronic disease that is unprecedented. The resulting changes in health care utilization including higher rates of specialty should not impact preventative care such as CRC screening. Although the results of my second study show that a physician recommendation is not predictive of true screening, the fact remains that CRC screening remains an event that requires coordination and action of both practitioner and patient. Extending responsibility of this coordinated effort outside of primary care will be critical as these older adults enter the at-risk age group for CRC.

Finally, the results of all of these studies are informative for policymakers. The need for accountability for adequate data collection and accurate data analysis by federally funded agencies is critical to the proper design and success of screening interventions. The results support the need for continued funding to improve nationally representative surveys to capture all data necessary to correctly specify behavioral outcomes. In addition, the use of past BRFSS data should be avoided in future studies, especially when assessing quality benchmarks (e.g. the Agency for Healthcare Research and Quality Report). Finally, we need to critically examine the utility of Healthy People-style goals and benchmarks for screening. Federal goals and progress reports that are not specific to subgroups miss the problematic variation in rates and the unique multilevel contexts that drive that variation.

Overall, I found that there are many answers to the question of "What are we missing?" Challenging the status quo of methodological approaches to survey design and analysis as well as valid suspects for predictors of screening led to significant findings in the estimation of screening rates, rates of screening recommendations, and the role that MCC plays in the odds of both. The results have implications for these areas as well as trends in screening rates over time (which may reflect increasing diagnostic testing, not screening), intervention effectiveness (counting colonoscopies pre- and post-intervention is inadequate) and disparities (changes in rates resulted in reductions in some disparities and increases in others). I hope to extend this research to other outcomes including mammography, where the same methodological limitations are present in the literature. In the end, I hope that these studies lead to improvements in our understanding of CRC screening and improvements in population health.

References

American Cancer Society. (2011). Colorectal Cancer Facts & Figures 2011-2013. Retrieved from http://www.cancer.org/acs/groups/content/@epidemiologysurveilance/documents/document/acspc-028323.pdf

Bertakis, K. D., Azari, R., Helms, L. J., Callahan, E. J., & Robbins, J. A. (2000). Gender differences in the utilization of health care services. Journal of Family Practice, 49(2), 147-152.

Green, C. A., & Pope, C. R. (1999). Gender, psychosocial factors and the use of medical services: a longitudinal analysis. *Social science & medicine*, *48*(10), 1363-1372.