Looking to the Future: Incorporating Genomic Information into Disparities Research to Reduce Measurement Error and Selection Bias

Alexandra E. Shields and William H. Crown

Objective. To extend recent conceptual and methodological advances in disparities research to include the incorporation of genomic information in analyses of racial/ethnic disparities in health care and health outcomes.

Data Sources. Published literature on human genetic variation, the role of genetics in disease and response to treatment, and methodological developments in disparities research.

Study Design. We present a conceptual framework for incorporating genomic information into the Institute of Medicine definition of racial/ethnic disparities in health care, identify key concepts used in disparities research that can be informed by genomics research, and illustrate the incorporation of genomic information into current methods using the example of HER-2 mutations guiding care for breast cancer.

Principal Findings. Genomic information has not yet been incorporated into disparities research, though it has direct relevance to concepts of race/ethnicity, health status, appropriate care, and socioeconomic status. The HER-2 example demonstrates how available genetic information can be incorporated into current disparities methods to reduce selection bias and measurement error. Advances in health information infrastructure may soon make standardized genetic information more available to health services researchers.

Conclusion. Genomic information can refine measurement of racial/ethnic disparities in health care and health outcomes and should be included wherever possible in disparities research.

Key Words. Health economics, social determinants of health, racial/ethnic differences in health and health care, personalized medicine, genomics

Despite the national commitment to eliminate health disparities (U.S. Department of Health and Human Services 2000; Institute of Medicine 2001), significant disparities persist in the quality of care received by racial/ethnic groups

and subsequent health outcomes (Institute of Medicine 2003). In 2003, the Institute of Medicine (IOM) defined a health care disparity as "racial or ethnic differences in the quality of health care that are not due to patients' clinical needs and preferences, and appropriateness of intervention" (Institute of Medicine 2003). In measuring health care disparities, most studies focus on the coefficient of the race/ethnicity variable after adjusting for a range of covariates, but this approach does not acknowledge that racial/ethnic differences are mediated by—not independent of—socioeconomic status (McGuire et al. 2006; Agency for Healthcare Research and Quality 2011; Cook et al. current issue). Drawing from methods used in the economics literature on wage discrimination, McGuire, Cook, and others have advocated a methodological approach based on the IOM definition of health care disparity in which disparities are decomposed to estimate the contribution of socioeconomic status to racial/ethnic disparities in health care (McGuire et al. 2006; Cook et al. 2008). Similar methods have begun to appear in the health outcomes literature as well (Crown 2010). At the same time, advances in genomics research, through which patterns in the human genome are examined to identify variants or regions of the genome important to disease etiology or treatment response (National Human Genome Research Institute 2010), are increasing our understanding of the role of genetics in disease risk and treatment response (Manolio et al. 2009), and even the ways in which socioeconomic status can mediate genetic risk through gene–environment interactions (GEI) and epigenetic effects (Olden et al. 2011). A growing number of clinical guidelines, particularly in oncological care, now include practice recommendations based on a patient's genetic status (Amstutz and Carleton 2011). This new knowledge has not yet been widely incorporated into disparities research.

This paper addresses the implications of emerging genomic information for disparities research. We argue that genetic information is an important omitted variable in health services research generally and disparities research in particular, and one that will become even more important as time goes on. Genomic data can help to differentiate geographical ancestry from self-identified race; provide additional information about health status, appropriate care, and expected clinical outcomes; and shed new light on the importance of

Address correspondence to Alexandra E. Shields, Ph.D., 50 Staniford St, Suite 901, Boston, MA 02114; e-mail: ashields@partners.org. Alexandra E. Shields, Ph.D., and William H. Crown, Ph.D., are with the Harvard/MGH Center on Genomics, Vulnerable Populations and Health Disparities, and Mongan Institute for Health Policy, Massachusetts General Hospital, Boston, MA. William H. Crown, Ph.D., is with the OptumInsight Life Sciences, Waltham, MA.

certain socioeconomic factors (e.g., greater exposure to stress) as critical control variables due to their role in amplifying genetic risk of disease. The inclusion of genetic variables in racial/ethnic disparities analyses, where possible, can thus help reduce measurement error and reduce bias in estimates of treatment effects, and allow for a better understanding of the sources of health disparities between self-identified racial/ethnic groups in our society.

We begin with a discussion of the intersection of emerging genomics research and four key concepts in disparities research: race, health status, clinical appropriateness, and socioeconomic status. We then present a conceptual framework showing how genomic information can be incorporated into disparities research, using the case of HER-2 status among women with breast cancer. Finally, recognizing that genomic information has yet to become widely available in health services data, we discuss statistical methods that test and help to control for measurement error due to the omission of variables that capture clinically relevant genetic heterogeneity.

GENOMICS RESEARCH EXPANDING KEY CONCEPTS USED IN DISPARITIES RESEARCH

Genomics research is expanding or refining several key concepts used in disparities research and in health services research more generally. Figure 1 illustrates the IOM definition of racial/ethnic health care disparities, measures commonly used to operationalize this definition (Cook et al. current issue), and how specific measures can be extended or refined using genomic information. Below, we address the intersection of genomics research with four concepts central to disparities research: race, health status, clinical appropriateness, and socioeconomic status.

What Is "Race"?

Foundational to any discussion of disparities research is the meaning of "race." The concept of race has been employed in very different ways in the fields of health services research and genomics research. Disparities research has been dominated by the use of the racial/ethnic categories institutionalized by the Office of Management and Budget (OMB) in 1997 —the same categories used in the U.S. Census—to assess differences in the quality of care received across groups (Office of Management and Budget 1997). The National Institutes of Health (NIH) regulation requiring all

Figure 1:   Incorporating Genomic Information into the IOM Definition of Health Care Disparities[1]
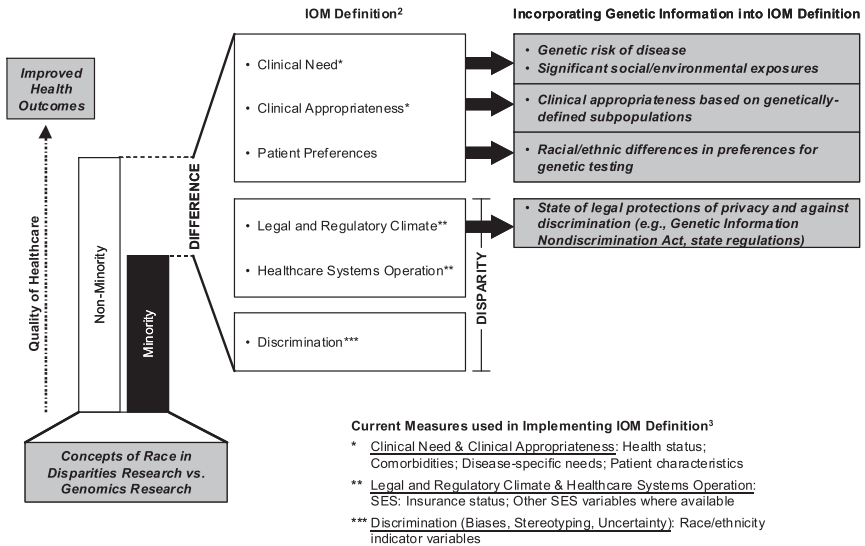
Sources:
[1] Based on Figure 5-1 "Differences, disparities, and discrimination: Populations with equal access to healthcare," Unequal Treatment, 2002
[2] Unequal Treatment, 2002
[3] Cook, McGuire, Zaslavsky, 2012 *(current issue)*

federally funded researchers to report study participants based on the OMB categories was not based on presumed biological differences (National Institutes of Health 1993), but rather by concerns for distributive justice (e.g., equal access to the benefits and burdens of clinical research) (National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research 1978; McCarthy 1994; Shields et al. 2005). The NIH requirement that researchers collect and report their study participants using these categories fostered the use of these same categories in data analysis. Thus, in most early genetics studies, the OMB racial/ethnic categories were used as a proxy for human genetic heterogeneity (i.e., human genetic variation or population structure). Human genetic variation, however, is a "continuous variable" that represents patterns of human mating and migration. These patterns can be mapped spatially (Parra 2007) and are often categorized according to geographical ancestry (Tishkoff and Verrelli 2003). Self-identified race variables have often been used as a proxy for geographical ancestry (e.g., black for "African ancestry"; white

for "European ancestry"), although both geneticists and social scientists have argued that self-identified racial/ethnic categories are poor proxies for capturing human genetic heterogeneity (Royal and Dunston 2004; Shields et al. 2005).

In time, the field of human population genetics devised more precise ways to control for underlying population structure that have shown the limitations of the OMB racial/ethnic categories as a useful proxy for genetically meaningful groups. Methods for admixture mapping use data from the entire genome to empirically group individuals according to geographical ancestry based on their genetic data (Patterson et al. 2004). While empirical measures of geographical ancestry may be correlated with self-identified OMB racial/ethnic categories, OMB categories do not begin to capture the biological diversity of even the most basic geographical groupings. For example, a recent study showed that genetically-determined African ancestry among persons in the United States identifying as African American ranged from 1 to 99 percent (Bryc et al. 2010). Similar analyses have shown the extraordinary genetic diversity in European populations previously thought to be fairly homogeneous "whites" (Novembre et al. 2008).

Given the well-documented and unacceptable gaps in the quality of health care received by self-identified white and minority patients in the U.S. health care system, disparities measured according to self-identified race/ethnicity will continue to be powerful policy levers for working toward equity, but these categories should not be considered adequate proxies for human genetic variation. Health services researchers will rarely have data available that include empirical measures of geographical ancestry. When available, these data could provide additional information regarding health status insofar as they capture as yet undiscovered, clinically relevant genetic variants that vary according to geographical ancestry, although the effect size of such variants will likely be modest compared to the contribution of social and environmental factors in explaining disparities (Hirschhorn and Gajdos 2011). In the vast majority of cases, however, data on geographical ancestry will not be available, and the best that can be done is to take great care in specifying precisely what is and is not being captured in the self-identified racial/ethnic variables being used.

Health Status

Patients' clinical need has been measured primarily as health status, using whatever general or disease-specific measures might be available. Recent

advances in genomics research, most notably genome-wide association studies, have identified numerous genetic variants associated with increased disease risk, thereby identifying new subsets of patients likely to have more intensive future clinical needs. For example, seven genes account for approximately 10 percent of variation in diabetes risk (Voight et al. 2010). Thirty genetic variants account for close to 10 percent of inter-individual variation in HDL, LDL, and triglyceride levels indicative of cardiovascular disease (Kathiresan et al. 2009). *BRCA1/2* mutations increase women's lifetime risk of breast cancer from 12 percent in the general population to about 80 percent (King, Marks, and Mandell 2003). Three genetic variants have been shown to account for a 2–4-fold increase in risk for developing age-related macular degeneration (Maller et al. 2006). Multiple variants have also been found to more than double the risk for nondiabetic end-stage renal disease, and these variants account for much of the excess risk among persons of African versus European ancestry (Kao et al. 2008). As data on genetic disease risk become validated and available to health services researchers, they should be included as individual-level measures in disparities research models to control for underlying differences in general or disease-specific health status, as appropriate.

Once individual mutations or constellations of variants have been shown to be reliable indicators of increased disease risk, they should be included in disparities analyses whenever possible as additional measures of health status in order to reduce measurement error. Some of these risk variants will vary in frequency among populations characterized according to geographical ancestry, and thus they will have varying degrees of correlation with the OMB racial/ethnic categories. A subset of these will have substantial clinical relevance. The 8q24 variant, for example, confers significantly increased risk of prostate cancer and is found predominantly in men of African ancestry (Freedman et al. 2006). However, in the proposed framework, we advocate the inclusion of genetic markers of health status as a means of reducing measurement error, not as proxies for race/ethnicity. The magnitude of the reduction in bias and improvements in precision as a result of incorporating genomic markers of health status will vary widely by phenotype and the degree to which the variant (or constellation of variants) is correlated with either the OMB race/ethnicity categories or outcome variables. Research results to date indicate that genetic variation will typically account for a small proportion of variance in disease risk and health outcomes among

OMB-defined racial/ethnic groups, although in many cases genetic variation will be sufficiently informative to meaningfully reduce measurement error.

Clinical Appropriateness

Beyond controlling for health status, or clinical need more generally, genomics research is also beginning to generate information that can be used to match patients to optimal treatments (Beitelshees and Veenstra 2011). Genetically-tailored treatments are most widely established in cancer care, where it has now become standard practice to match patients to appropriate therapy according to their specific tumor biology (i.e., somatic mutations) (Garman, Nevins, and Potti 2007). Guidelines now recommend, for example, that women with breast cancer whose tumors are HER-2 positive receive the medication trastuzumab (Herceptin) to reduce risk of recurrence (Piccart-Gebhart et al. 2005). Approximately 10 percent of breast cancer patients have tumors that respond to trastuzumab; but among patients with HER-2 positive cancers, this figure is as high as 50 percent (Chang 2010). Use of trastuzumab among HER-2 negative women, by contrast, results in little clinical gain while exposing them to potential adverse effects of the drug (e.g., heart damage) (Romond et al. 2005). As another example, certain non-small cell lung cancers over-express the epidermal growth factor receptor (EGFR) (Paez et al. 2004). Among this subset of lung cancer patients, gefitinib (Iressa) inhibits this growth factor receptor and significantly improves survival (Kris et al. 2003). Some academic medical centers now routinely assess EGFR expression among all non-small cell lung cancer patients (Hayden 2009).

A growing number of examples outside oncology exist as well. Genetic variants that code for cytochrome P450 (CYP450) enzymes (i.e., *CYP2C9*, *CYP2C19*, and *CYP2D6*), which determine how fast an individual can metabolize certain medications (e.g., certain beta-blockers, tricyclic antidepressants, and anticoagulants), are being used to classify patients' metabolizing profiles and adjust dosing accordingly (Tomalik-Scharte et al. 2007). This type of genetic information is also being used to guide treatment decisions for thrombosis and HIV/AIDS (Mallal et al. 2008; Schwarz et al. 2008).

As with genetic measures of health status, once a specific genetic status is proven to characterize a clinically relevant subset of patients who would especially benefit from (or be harmed by) a given treatment, this information

should be incorporated into health services research and disparities research whenever possible. Again, these clinically relevant variants will rarely be equally distributed across all human beings, and will more often differ in frequency among persons of differing geographical ancestry. Thus, the degree of correlation between genetic markers important to refining assessment of appropriate care and the OMB racial/ethnic categories used in disparities research will vary, depending on the variant(s) pertinent to the study at hand. To the extent that genomic data are important for guiding treatment choices, their omission in a given analysis will introduce a correlation between the treatment variable and the error term in the equation and will result in biased estimates of treatment effects.

Socioeconomic Status

In disparities research, socioeconomic status is conceptualized as mediating racial/ethnic disparities through health systems. However, a growing body of research shows how socioeconomic status also mediates health status through GEI and epigenetic effects, where environmental exposures that track with socioeconomic status are found to amplify genetic risk of disease (Hunter 2005; Jirtle and Skinner 2007). GEI research, for example, investigates how complex relationships between genes and environmental stressors produce differential disease risk. A specific genetic variant may be common across all human populations but only deleteriously affect a biological process in the presence of a particular environmental exposure. For example, carriers of *CD14* mutations are more susceptible to experiencing decreased lung function when exposed to airborne endotoxins (Smit et al. 2011). Poor diet (Low and Tai 2007) and air pollution (Moysich et al. 1999; Park et al. 2006) interact with specific genes to increase risk for breast cancer and cardiovascular disease. As genomics research identifies specific environmental exposures that are mediated by socioeconomic status and amplify genetic risk of disease, the inclusion of such exposure information, if available, would further reduce measurement error in disparities analyses. As a growing number of studies link air pollution to compounded genetic risk for decreased lung function (Ege et al. 2011), for instance, one can imagine geographic information systems (GIS)-linked air pollution exposure data becoming an expected covariate in future studies investigating asthma or chronic obstructive pulmonary disorder disparities.

A Methodological Framework for Incorporating Genomic Information into Health Disparities Research

Among the many intersections between current conceptual and methodological frameworks for examining racial/ethnic health disparities and genomic information, genomic information can be immediately applied to three areas: (a) accounting for previously unmeasured genetic variation that affects health status or treatment response; (b) refining measurement of "appropriate care"; and (c) incorporating genetic information indicating likely benefit from a given treatment into the assessment of health outcomes. We illustrate these areas using the example of HER-2 status guiding breast cancer care.

Accounting for Previously Unmeasured Genetic Variation

To date, the growing knowledge about genetic variation and the role of genetics in predicting an individual's risk of disease or response to a given treatment has rarely been incorporated into disparities research methods. Determining the best methods for addressing the issue of previously unobserved genetic influences that may be correlated with both treatment and outcomes is a current challenge for the health services research field. As increasing amounts of evidence-based, clinically relevant genetic information becomes available, the incorporation of such data into analyses of health disparities can be expected to reduce measurement error and selection bias, and improve the value of racial/ethnic disparities research.

We describe below an analytic framework that incorporates genetic variables in disparities research in anticipation of the growing availability of genomic data, recognizing that many of the data needed to implement the proposed framework are not yet widely available to health services researchers. Central to this framework is the acknowledgment that there is clinically relevant genetic heterogeneity within and between the OMB-defined racial/ethnic groups. Ideally, one would want to control for genetic variability within and between self-identified OMB racial/ethnic groups with robust markers of geographical ancestry, as well as known markers relevant to health status and disease risk. This would address the selection bias problems (i.e., bias introduced by the omission of variables that are correlated with both treatment and patient outcomes) associated with unobserved heterogeneity in treatment selection and response. In addition to including genetic variables in our

framework, we utilize, within each racial/ethnic group, the decomposition methods that have been used to assess racial/ethnic disparities in health care in the recent literature (Cook et al. current issue). Finally, these methods are extended to incorporate racial/ethnic disparities in nonrandom selection (matching patients to treatments based on their genetic profile) into treatment and heterogeneity of treatment response (due to one's genetic make-up) within treatment categories. The resulting analytic framework combines statistical methods from both disparities research and outcomes research, and incorporates genetic information to enable more robust analyses of racial/ethnic disparities in health care utilization and outcomes.

Any number of factors can undermine the reliability of statistical inferences drawn from observational data. Common problems in observational data analysis include omitted variables, measurement error, joint causation, and unobserved factors (e.g., positive HER-2 status) that are correlated with both treatment selection and patient outcomes. The use of race/ethnicity categories as proxies for underlying genetic variation is a measurement error problem because of the imprecision of the OMB race categories and the wide genetic variation found within these racial categories (Novembre et al. 2008; Bryc et al. 2010). As a result, disease risk among genetic subgroups within a given OMB racial/ethnic category would be expected to vary. The potential clinical benefit to be derived from a given treatment can also vary dramatically according to genotype. Thus, genetic variation found within the OMB racial/ethnic categories has implications both for errors in treatment selection (i.e., appropriate care) as well as observed variation in treatment response and health outcomes. Selection bias in treatment choice and variation in treatment response can be reduced by identifying clinically relevant genetic subgroups. As genetic data become increasingly available, the inclusion of these variables will enable researchers to explain a greater proportion of disease burden and outcomes.

Some variables that are important predictors of health care utilization (e.g., health status, socioeconomic status) can be readily observed by researchers, but others (e.g., family history, genetic traits) may not be observed. The inability to include such variables can lead to a special type of measurement error known as omitted variable bias. The most direct solution is to include the missing data when possible. However, when this is not possible, researchers can use statistical methods that test and correct for selection bias in treatment and outcomes. To illustrate these points more concretely, consider the example of breast cancer treatment. A basic framework is provided in Equations (1–4).

Refining Measurement of "Appropriate Care"

$$Y_{\text{B}} = \hat{C_0} + \hat{C_1}X_2 + \hat{\epsilon_1} \tag{1}$$

$$Y_{\text{W}} = \hat{D_0} + \hat{D_1}X_2 + \hat{\epsilon_2} \tag{2}$$

Equations (1) and (2) represent the clinician's diagnostic and treatment process. Y_{B} and Y_{W} represent the prescribed treatments for blacks and whites, respectively. As discussed above, the OMB categories are a blunt instrument for classifying disease risk and controlling for heterogeneity of treatment selection. White women are known to be at higher risk for developing breast cancer than black women, but black women are diagnosed at later stages of disease and have higher mortality (Jemal et al. 2009). However, once diagnosed, the differential risk of disease may have limited utility in guiding treatment; matching women to the best treatment to optimize outcomes is the priority. As detailed above, within each racial category, information about HER-2 status has become a critical predictor of treatment selection. Thus, genetic testing for HER-2 receptor status should help to improve access to appropriate care for women with breast cancer, regardless of race. To the extent that black women experience poorer access to appropriate treatments for breast cancer, testing for HER-2 status may reduce racial disparities in breast cancer treatment. Using the counterfactual approach (online Supporting Information Appendix SA2), it would be possible to evaluate racial disparities in treatment with trastuzumab under the assumption that black women had the same mean HER-2 status and other characteristics as white women.

While the example of HER-2 is useful in illustrating the value of matching patients to treatment based on their mutation status, it simultaneously highlights a limitation of currently available genomic data of particular relevance to disparities researchers. Though roughly the same proportion of African American and white women will have a negative HER-2 status (Lund et al. 2010; Swede et al. 2011), black women with HER-2 negative status will be far more likely to have triple negative cancer (which is a much more aggressive form of breast cancer) (Carey et al. 2006; Cleator, Heller, and Coombes 2007). The vast majority of genomic studies in breast cancer, however—as with all conditions—have been conducted in European ancestry populations (Haga 2010), and so the genetics research needed to fully understand the etiology and treatment of triple negative cancer has not yet been performed. Large scale genomic studies of women of African ancestry with breast cancer are only now being undertaken (Hutter et al. 2011).

The racial disparities literature has largely focused on differential access to appropriate care by race. In the following sections, the racial disparities framework is extended to include an assessment of disparities in patient outcomes that may arise from disparities in care.

Incorporating Genomic Information into the Assessment of Health Outcomes

$$O_B = \hat{A}_0 + \hat{A}_1 X_B + \hat{A}_2 Y_B + \hat{\epsilon_3} \qquad (3)$$

$$O_W = \hat{B}_0 + \hat{B}_1 X_W + \hat{B}_2 Y_W + \hat{\epsilon_4} \qquad (4)$$

Equations (3) and (4) illustrate predicted outcomes O_B and O_W contingent upon observed patient characteristics X_B and X_W and treatment choice Y_B and Y_W for blacks and whites, respectively. One would expect that the use of HER-2 positive gene expression results to guide treatment selection should also lead to improvements in treatment outcomes within each racial category. By reducing within-group variation, the ability to measure true between group differences will be improved.

Once again, Equations (3) and (4) could be used to implement the traditional counterfactual approach to assess racial/ethnic differences in treatment outcomes. The counterfactual approach would enable researchers to examine whether differential outcomes by race/ethnicity persist after controlling for the same rate of treatment with trastuzumab, as well as other patient characteristics such as health status and socioeconomic status.

DISCUSSION

Advances in genomics research are generating new insight into the role of genetics in disease risk and treatment response. The pace of discovery is impressive, promising an increasing number of clinically relevant genetic markers that will become validated as evidence-based indicators of health status or clinical need and used to guide medical care. Due to patterns of human migration, these genetic markers often vary in frequency according to geographical ancestry, which in turn is often correlated with, but not to be confused with, groups based on self-identified race/ethnicity. The complex interplay of human genetic diversity and the OMB categories of race/ethnicity commonly deployed in disparities research present distinctive challenges.

To incorporate emerging genomic data into cutting-edge disparities research, several key concepts and measures, including race, health status, clinical appropriateness, and socioeconomic status, must be refined or expanded. Furthermore, appropriate genomic variables that account for population structure, health status, and appropriate care should be included wherever possible to reduce measurement error in disparities research models. Realizing the potential value of genomic information to disparities research will depend on addressing key normative and practical issues.

First and foremost, standards for evaluating the scientific reliability of emerging genomic information are needed. When is the scientific evidence strong enough to warrant including genetic information in disparities studies? In general, the scientific process for identifying control variables in any health services research is a function of evidence from the literature. In this sense, the threshold for inclusion of genomic variables in disparities studies is no different. The emerging field of personalized medicine, however, has been slow to embrace evidence-based medicine (Khoury et al. 2008), and the pace of research has made it difficult to reach consensus regarding the clinical utility of many genetic markers. Currently, the Centers for Disease Control's Evaluation of Genomic Applications in Practice and Prevention Initiative (Teutsch et al. 2009) is the premier source of independent, evidence-based assessment of the clinical utility and reliability of genomic applications. The U.S. Preventive Services Task Force (Agency for Healthcare Research and Quality 2012), Cochrane Collaboration (Sivell et al. 2007), and Secretary's Advisory Committee on Genetics, Health and Society (National Institutes of Health 2012) have also published reports assessing the validity and reliability of specific genetic applications, as have professional organizations recommending the use of specific mutations to guide treatment (Carlson et al. 2006; Levin et al. 2006; Lynch 2006; Burtness et al. 2009; Greenberg et al. 2009). When deciding whether mutations associated with disease risk or treatment response should be included in disparities research, their use in clinical practice and inclusion in evidence-based guidelines are reasonable evidentiary thresholds for health services researchers to consider.

The practical challenge for health services researchers trying to operationalize clinical guidelines based on genomic information lies in the lack of research infrastructure. The genomic information necessary to define denominator populations is not readily available in claims data or even electronic medical records. Germline (e.g., *BRCA1/2*) or somatic (e.g., HER-2) mutation status, which is used to determine appropriate treatment,

is often only available in medical charts or other data sources (e.g., cancer registries). Few datasets outside cancer include genomic information. Even within fully automated electronic health records (EHRs), mutation status is rarely entered electronically (Wilke et al. 2011). This data gap limits the feasibility of mounting large-scale studies assessing disparities in receipt of guideline-recommended care where genomic information is used to define the applicable patient populations.

Some health plans and health systems have begun collecting genetic information on patients' genetic status for select variants key to directing care, particularly in the areas of cancer and cardiovascular disease. The VA Hypertension Primary Care Longitudinal cohort, for example, has been incorporating genetic markers into EHRs since 2003 (Salem et al. 2010; Puppala et al. 2011). Recently, Kaiser Permanente was funded to link genetic and clinical data for up to 500,000 enrollees to assess the efficacy and toxicity of drugs, as well as the effect of different environmental exposures across different genetic profiles (Thomson et al. 2010; Wilke et al. 2011). The Kaiser Permanente Research Program on Genes, Environment and Health has now genotyped the DNA and the length of chromosome tips of more than 100,000 individuals. This program is merging these data with patients' EHRs, providing a valuable, integrated database. Such data resources will become more common in the coming years (Kaiser Permanente 2011). Other initiatives to incorporate genomic data into EHR systems include the Harvard University/Partners Healthcare system i2b2, the Vanderbilt BioVu44 database, and the multi-center, National Human Genome Research Institute-funded eMERGE (electronic Medical Records and Genomics Network) (Kohane 2011).

The infrastructure to facilitate the routine incorporation of genomic data into EHRs in the future is key to making use of the added power of genetic information to understand variation in disease risk and treatment outcomes among different populations, and to craft interventions most likely to ensure optimal outcomes for all. Federal initiatives determined to accelerate adoption of EHRs will likely be a major force in developing the infrastructure that will begin capturing clinically relevant genetic information and making it available to health services researchers. Adoption rates of EHRs remains low nationally, at approximately 13 percent among physicians and 8 percent among hospitals (DesRoches et al. 2008; Jha et al. 2009a), with safety net providers facing particular barriers to adoption (Shields et al. 2007; Jha et al. 2009b). However, large providers and health systems are already beginning to transform their information systems by merging data (including genomic data)

previously collected only in research into EHR formats (Massachusetts General Hospital 2009; Wilke et al. 2011).

As the first generation of genetic testing data becomes available to health services researchers, these data should be interpreted with care. Concerns about privacy and genetic discrimination may result in biased samples of patients for whom germline genetic test results are available, thus limiting generalizability. Patient preferences for testing play a significant role in determining whether a patient receives genetic testing (Bruno et al. 2010). Importantly, patient preferences for genetic testing have been shown to differ according to patients' race and socioeconomic status, with minority patients less likely to be aware of and less likely to believe in the benefits of testing (Peters, Rose, and Armstrong 2004; Halbert, Kessler, and Mitchell 2005). In reality, patients are rarely equipped to make informed decisions about their health care preferences (O'Connor et al. 2007; Peters et al. 2007). The complex nature of genetic information may make it more difficult for patients to understand available information and may lead to heightened anxieties about how genetic information will be used. The passage of the Genetic Information Nondiscrimination Act of 2008 may allay these fears over time.

The field of genomics is growing at an exponential rate, and new technological capacity has yielded results that have eclipsed findings from only a few years ago. As a result, an increasing wealth of clinically relevant, validated genomic data will become available to health services researchers for use in disparities research models in the coming years. Although we believe social, environmental, and political factors will continue to account for the lion's share of racial/ethnic disparities in the United States, we also believe that genetic data will be valuable in accounting for unobserved differences in health status and treatment heterogeneity due to genetic variation that may contribute to underlying prevalence of illness, treatment selection, and treatment response. As genomics research continues to identify clinically relevant patient subpopulations and knowledge of the complex factors that predict disease and treatment outcomes increases, we anticipate a gradual shift from statistical models with extremely large variability in selecting appropriate treatments and in predicting health outcomes toward models that explain an ever growing proportion of inter-individual variation. This added analytical power, in turn, has the potential to increase our understanding of the various sources of racial/ethnic disparities in health care and health outcomes, and generate new interventions aimed at reducing disparities.

ACKNOWLEDGMENTS

REFERENCES

Agency for Healthcare Research and Quality. 2011. *2010 National Healthcare Disparities Report*. Rockville, MD: Agency for Healthcare Research and Quality.

Agency for Healthcare Research and Quality. 2012. "U.S. Preventive Services Task Force" [accessed on February 1, 2012]. Available at http://www.ahrq.gov/clinic/uspstfix.htm.

Amstutz, U., and B. C. Carleton. 2011. "Pharmacogenetic Testing: Time for Clinical Practice Guidelines." *Clinical Pharmacology and Therapeutics* 89 (6): 924–7.

Beitelshees, A. L., and D. L. Veenstra. 2011. "Evolving Research and Stakeholder Perspectives on Pharmacogenomics." *Journal of the American Medical Association* 306 (11): 1252–3.

Bruno, M., M. Digennaro, S. Tommasi, B. Stea, T. Danese, F. Schittulli, and A. Paradiso. 2010. "Attitude Towards Genetic Testing for Breast Cancer Susceptibility: A Comparison of Affected and Unaffected Women." *European Journal of Cancer Care* 19 (3): 360–8.

Bryc, K., A. Auton, M. R. Nelson, J. R. Oksenberg, S. L. Hauser, S. Williams, A. Froment, J. M. Bodo, C. Wambebe, S. A. Tishkoff, and C. D. Bustamante. 2010. "Genome-Wide Patterns of Population Structure and Admixture in West Africans and African Americans." *Proceedings of the National Academy of Sciences* 107 (2): 786–91.

Burtness, B., M. Anadkat, S. Basti, M. Hughes, M. E. Lacouture, J. S. McClure, P. L. Myskowski, J. Paul, C. S. Perlis, L. Saltz, and S. Spencer. 2009. "NCCN Task Force Report: Management of Dermatologic and Other Toxicities Associated with EGFR Inhibition in Patients with Cancer." *Journal of the National Comprehensive Cancer Network* 7 (suppl 1): S5–S21; quiz S22-24-S5-21; quiz S22-24.

Carey, L. A., C. M. Perou, C. A. Livasy, L. G. Dressler, D. Cowan, K. Conway, G. Karaca, M. A. Troester, C. K. Tse, S. Edmiston, S. L. Deming, J. Geradts, M. C. U. Cheang, T. O. Nielsen, P. G. Moorman, H. S. Earp, and R. C. Millikan. 2006. "Race, Breast Cancer Subtypes, and Survival in the Carolina Breast Cancer Study." *Journal of the American Medical Association* 295 (21): 2492–2502.

Carlson, R. W., S. J. Moench, M. E. Hammond, E. A. Perez, H. J. Burstein, D. C. Allred, C. L. Vogel, L. J. Goldstein, G. Somlo, W. J. Gradishar, C. A. Hudis, M. Jahanzeb, A. Stark, A. C. Wolff, M. F. Press, E. P. Winer, S. Paik, B. M. Ljung, and the NCCN HER2 Testing in Breast Cancer Task Force. 2006. "HER2 Testing in Breast Cancer: NCCN Task Force Report and Recommendations." *Journal of the National Comprehensive Cancer Network* 4: S1–S22.

Chang, H. R. 2010. "Trastuzumab-Based Neoadjuvant Therapy in Patients with HER2-Positive Breast Cancer." *Cancer* 116 (12): 2856–2867.

Cleator, S., W. Heller, and R. C. Coombes. 2007. "Triple-Negative Breast Cancer: Therapeutic Options." *Lancet Oncology* 8 (3): 235–244.

Cook, B. L., T. G. McGuire, E. Meara, and A. M. Zaslavsky. 2008. "Adjusting for Health Status in Non-Linear Models of Health Care Disparities." *Health Services and Outcomes Research Methodology* 9 (1): 1–21.

Cook, B. L., T. G. McGuire, and A. M. Zaslavsky. 2012. "Measuring Racial/Ethnic Disparities in Health Care: Methods and Practical Issues." *Health Services Research* doi: 10.1111/j.1475-6773.2012.01413.x.

Crown, W. H. 2010. "There's a Reason They Call Them Dummy Variables: A Note on the Use of Structural Equation Techniques in Comparative Effectiveness Research." *PharmacoEconomics* 28 (10): 947–955.

DesRoches, C. M., E. G. Campbell, S. R. Rao, K. Donelan, T. G. Ferris, A. Jha, R. Kaushal, D. E. Levy, S. Rosenbaum, A. E. Shields, and D. Blumenthal. 2008. "Electronic Health Records in Ambulatory Care–A National Survey of Physicians." *New England Journal of Medicine* 359 (1): 50–60.

Ege, M. J., D. P. Strachan, W. O. C. M. Cookson, M. F. Moffatt, I. Gut, M. Lathrop, M. Kabesch, J. Genuneit, G. Büchele, B. Sozanska, A. Boznanski, P. Cullinan, E. Horak, C. Bieli, C. Braun-Fahrländer, D. Heederik, and E. von Mutius. 2011. "Gene-Environment Interaction for Childhood Asthma and Exposure to Farming in Central Europe." *Journal of Allergy and Clinical Immunology* 127 (1): 144.e1–4.

Freedman, M. L., C. A. Haiman, N. Patterson, G. J. McDonald, A. Tandon, A. Waliszewska, K. Penney, R. G. Steen, K. Ardlie, E. M. John, I., Oakley-Girvan, A. S. Whittemore, K. A. Cooney, S. A. Ingles, D. Altshuler, B. E. Henderson, and D. Reich 2006. "Admixture Mapping Identifies 8q24 as a Prostate Cancer Risk Locus in African-American men." *Proceedings of the National Academy of Sciences* 103 (38): 14068–68.

Garman, K. S., J. R. Nevins, and A. Potti. 2007. "Genomic Strategies for Personalized Cancer Therapy." *Human Molecular Genetics* 16 (R2): R226–26.

Greenberg, P. L., C. K. Rigsby, R. M. Stone, H. J. Deeg, S. D. Gore, M. M. Millenson, S. D. Nimer, M. R. O'Donnell, P. J. Shami, and R. Kumar. 2009. "NCCN Task Force: Transfusion and Iron Overload in Patients with Myelodysplastic Syndromes." *Journal of the National Comprehensive Cancer Network* 7 (suppl 9): S1–S16.

Haga, S. B. 2010. "Impact of Limited Population Diversity of Genome-Wide Association Studies." *Genetics in Medicine* 12 (2): 81–4.

Halbert, C. H., L. J. Kessler, and E. Mitchell. 2005. "Genetic Testing for Inherited Breast Cancer Risk in African Americans." *Cancer Investigation* 23 (4): 285–295.

Hayden, E. C. 2009. "Personalized Cancer Therapy Gets Closer." *Nature* 458 (7235): 131–2.

Hirschhorn, J. N., and Z. K. Z. Gajdos. 2011. "Genome-Wide Association Studies: Results from the First Few Years and Potential Implications for Clinical Medicine." *Annual Review of Medicine* 62 (1): 11–24.

H.R. 493–110th Congress. 2008. "Genetic Information Nondiscrimination Act of 2008, H.R. 493, 110th Congress, 2nd Session."

Hunter, D. J. 2005. "Gene-Environment Interactions in Human Diseases." *Nature Reviews Genetics* 6: 287–98.

Hutter, C. M., A. M. Young, H. M. Ochs-Balcom, C. L. Carty, T. Wang, C. T. L. Chen, T. E. Rohan, C. Kooperberg, and U. Peters. 2011. "Replication of Breast Cancer GWAS Susceptibility Loci in the Women's Health Initiative African American SHARe Study." *Cancer Epidemiology Biomarkers & Prevention* 20 (9): 1950–59.

Institute of Medicine. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academy Press.

Institute of Medicine. 2003. *Unequal Treatment: Confronting Racial and Ethnic Disparities in Health Care*. Washington, DC: The National Academies Press.

Jemal, A., R. Siegel, E. Ward, Y. Hao, J. Xu, and M. J. Thun. 2009. "Cancer Statistics, 2009." *CA: A Cancer Journal for Clinicians* 59 (4): 225–49.

Jha, A. K., C. M. DesRoches, E. G. Campbell, K. Donelan, S. R. Rao, T. G. Ferris, A. Shields, S. Rosenbaum, and D. Blumenthal. 2009a. "Use of Electronic Health Records in US Hospitals." *New England Journal of Medicine* 360 (16): 1628–38.

Jha, A. K., C. M. DesRoches, A. E. Shields, P. D. Miralles, J. Zheng, S. Rosenbaum, and E. G. Campbell. 2009b. "Evidence of an Emerging Digital Divide among Hospitals That Care for the Poor." *Health Affairs* 28 (6): w1160–w70.

Jirtle, R. L., and M. K. Skinner. 2007. "Environmental Epigenomics and Disease Susceptibility." *Nature Reviews Genetics* 8 (4): 253–62.

Kaiser Permanente. 2011. "DNA of 100,000 Kaiser Permanente Members Genotyped in 15 Months, Creating Novel Resource for Research." *Division of Research* [accessed on November 2, 2011]. Available at: http://researchnews.kaiser.org/?p=1811.

Kao, W. H. L., M. J. Klag, L. A. Meoni, D. Reich, Y. Berthier-Schaad, M. Li, J. Coresh, N. Patterson, A. Tandon, N. R. Powe, N. E. Fink, J. H. Sadler, M. R. Weir, H. E. Abboud, S. G. Adler, J. Divers, S. K. Iyengar, B. I. Freedman, P. L. Kimmel, W. C. Knowler, O. F. Kohn, K. Kramp, D. J. Leehey, S. B. Nicholas, M. V. Pahl, J.

R. Schelling, J. R. Sedor, D. Thornley-Brown, C. A. Winkler, M. W. Smith, and R. S. Parekh. 2008. "MYH9 Is Associated with Nondiabetic End-Stage Renal Disease in African Americans." *Nature Genetics* 40 (10): 1185–92.

Kathiresan, S., C. J. Willer, G. M. Peloso, S. Demissie, K. Musunuru, E. E. Schadt, L. Kaplan, D. Bennett, Y. Li, T. Tanaka, B. F. Voight, L. L. Bonnycastle, A. U. Jackson, G. Crawford, A. Surti, C. Guiducci, N. P. Burtt, S. Parish, R. Clarke, D. Zelenika, K. A. Kubalanza, M. A. Morken, L. J. Scott, H. M. Stringham, P. Galan, A. J. Swift, J. Kuusisto, R. N. Bergman, J. Sundvall, M. Laakso, L. Ferrucci, P. Scheet, S. Sanna, M. Uda, Q. Yang, K. L. Lunetta, J. Dupuis, P. I. W. de Bakker, C. J. O'Donnell, J. C. Chambers, J. S. Kooner, S. Hercberg, P. Meneton, E. G. Lakatta, A. Scuteri, D. Schlessinger, J. Tuomilehto, F. S. Collins, L. Groop, D. Altshuler, R. Collins, G. M. Lathrop, O. Melander, V. Salomaa, L. Peltonen, M. Orho-Melander, J. M. Ordovas, M. Boehnke, G. R. Abecasis, K. L. Mohlke, and L. A. Cupples. 2009. "Common Variants at 30 Loci Contribute to Polygenic Dyslipidemia." *Nature Genetics* 41 (1): 56–65.

Khoury, M. J., A. Berg, R. Coates, J. Evans, S. M. Teutsch, and L. A. Bradley. 2008. "The Evidence Dilemma in Genomic Medicine." *Health Affairs* 27 (6): 1600–11.

King, M.-C., J. H. Marks, and J. B. Mandell. 2003. "Breast and Ovarian Cancer Risks due to Inherited Mutations in BRCA1 and BRCA2." *Science (New York, N.Y.)* 302 (5645): 643–46.

Kohane, I. S. 2011. "Using Electronic Health Records to Drive Discovery in Disease Genomics." *Nature Reviews Genetics* 12 (6): 417–28.

Kris, M. G., R. B. Natale, R. S. Herbst, T. J. Lynch, D. Prager, C. P. Belani, J. H. Schiller, K. Kelly, H. Spiridonidis, A. Sandler, K. S. Albain, D. Cella, M. K. Wolf, S. D. Averbuch, J. J. Ochs, and A. C. Kay. 2003. "Efficacy of Gefitinib, an Inhibitor of the Epidermal Growth Factor Receptor Tyrosine Kinase, in Symptomatic Patients with Non–Small Cell Lung Cancer." *Journal of the American Medical Association* 290 (16): 2149–58.

Levin, B., J. S. Barthel, R. W. Burt, D. S. David, J. M. Ford, F. M. Giardiello, S. B. Gruber, A. L. Halverson, S. Hamilton, W. Kohlmann, K. A. Ludwig, P. M. Lynch, C. Marino, E. W. Martin Jr, R. J. Mayer, B. Pasche, S. J. Pirruccello, A. Rajput, M. S. Rao, M. Shike, G. Steinbach, J. P. Terdiman, D. Weinberg, and S. J. Winawer. 2006. "Colorectal Cancer Screening Clinical Practice Guidelines." *Journal of the National Comprehensive Cancer Network* 4 (4): 384–420.

Low, Y. L., and E. S. Tai. 2007. "Understanding Diet-Gene Interactions: Lessons from Studying Nutrigenomics and Cardiovascular Disease." *Mutation Research* 622 (1-2): 7–13.

Lund, M. J., E. N. Butler, B. Y. Hair, K. C. Ward, J. H. Andrews, G. Oprea-Ilies, A. R. Bayakly, R. M. O'Regan, P. M. Vertino, and J. W. Eley. 2010. "Age/Race Differences in HER2 Testing and in Incidence Rates for Breast Cancer Triple Subtypes: A Population-Based Study and First Report." *Cancer* 116 (11): 2549–59.

Lynch, P. M. 2006. "Current Approaches in Familial Colorectal Cancer: A Clinical Perspective." *Journal of the National Comprehensive Cancer Network* 4 (4): 421–30.

Mallal, S., E. Phillips, G. Carosi, J. M. Molina, C. Workman, J. Tomaži, E. Jägel-Guedes, S. Rugina, O. Kozyrev, J. F. Cid, P. Hay, D. Nolan, S. Hughes, S. Hughes. A.

Hughes, S. Ryan, N. Fitch, D. Thorborn, and A. Benbow 2008. "HLA-B* 5701 Screening for Hypersensitivity to Abacavir." *New England Journal of Medicine* 358 (6): 568–79.

Maller, J., S. George, S. Purcell, J. Fagerness, D. Altshuler, M. J. Daly, and J. M. Seddon. 2006. "Common Variation in Three Genes, Including a Noncoding Variant in CFH, Strongly Influences Risk of age-Related Macular Degeneration." *Nature Genetics* 38 (9): 1055–59.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. Mackay, S. A. McCarroll, and P. M. Visscher 2009. "Finding the Missing Heritability of Complex Diseases." *Nature* 461 (7265): 747–53.

Massachusetts General Hospital. 2009. *Massachusetts General Hospital Cancer Center Opens Molecular Pathology Lab to Genetically Profile all Patient Tumors.* Boston: News, Massachusetts General Hospital.

McCarthy, C. R. 1994. "Historical Background of Clinical Trials Involving Women and Minorities." *Academic Medicine* 69 (9): 695–98.

McGuire, T. G., M. Alegria, B. L. Cook, K. B. Wells, and A. M. Zaslavsky. 2006. "Implementing the Institute of Medicine Definition of Disparities: An Application to Mental Health Care." *Health Services Research* 41 (5): 1979–2005.

Moysich, K. B., P. G. Shields, J. L. Freudenheim, E. F. Schisterman, J. E. Vena, P. Kostyniak, H. Greizerstein, J. R. Marshall, S. Graham, and C. B. Ambrosone. 1999. "Polychlorinated Biphenyls, Cytochrome P4501A1 Polymorphism, and Postmenopausal Breast Cancer Risk." *Cancer Epidemiology, Biomarkers & Prevention*, A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology, 8 (1): 41–44.

National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. 1978. *The Belmont Report: Ethical Principles and Guidelines for the Protection of Human Subjects of Research.* Washington, DC: Government Printing Office.

National Human Genome Research Institute. 2010. "Genetic and Genomic Science" [accessed on March 12, 2012]. Available at http://www.genome.gov/19016904.

National Institutes of Health. 1993. "National Institutes of Health Revitalization Act: Subtitle B, Clinical Research Equity Regarding Women and Minorities." 42 U.S. C.: National Institutes of Health Revitalization Act.

National Institutes of Health. 2012. "Secretary's Advisory Committee on Genetics, Health, and Society" [accessed on March 12, 2012]. Available at http://oba.od.nih.gov/SACGHS/sacghs_home.html.

Novembre, J., T. Johnson, K. Bryc, Z. Kutalik, A. R. Boyko, A. Auton, A. Indap, K. S. King, S. Bergmann, M. R. Nelson, M. Stephens, and C. D. Bustamante 2008. "Genes Mirror Geography within Europe." *Nature* 456 (7218): 98–101.

O'Connor, A. M., J. E. Wennberg, F. Legare, H. A. Llewellyn-Thomas, B. W. Moulton, K. R. Sepucha, A. G. Sodano, and J. S. King. 2007. "Toward the 'Tipping Point': Decision Aids and Informed Patient Choice." *Health Affairs* 26 (3): 716–25.

Office of Management and Budget. 1997. *Recommendations from the Interagency Committee for the Review of the Racial and Ethnic Standards to the Office of Management and Budget Concerning Changes to the Standards for the Classification of Federal Data on Race and Ethnicity*, pp 36873–946. Washington, DC: Federal Register.

Olden, K., N. Freudenberg, J. Dowd, and A. E. Shields. 2011. "Discovering how Environmental Exposures Alter Genes Could Lead to New Treatments for Chronic Illnesses." *Health Affairs* 30 (5): 833–41.

Paez, J. G., P. A. Jänne, J. C. Lee, S. Tracy, H. Greulich, S. Gabriel, P. Herman, F. J. Kaye, N. Lindeman, T. J. Boggon, K. Naoki, H. Sasaki, Y. Fujii, M. J. Eck, W. R. Sellers, B. E. Johnson, and M. Meyerson 2004. "EGFR Mutations in Lung Cancer: Correlation with Clinical Response to Gefitinib Therapy." *Science* 304 (5676): 1497–1500.

Park, S. K., M. S. O'Neill, R. O. Wright, H. Hu, P. S. Vokonas, D. Sparrow, H. Suh, and J. Schwartz. 2006. "HFE Genotype, Particulate Air Pollution, and Heart Rate Variability: A Gene-Environment Interaction." *Circulation* 114 (25): 2798–805.

Parra, E. J. 2007. "Human Pigmentation Variation: Evolution, Genetic Basis, and Implications for Public Health." *American Journal of Physical Anthropology* 134 (S45): 85–105.

Patterson, N., N. Hattangadi, B. Lane, K. E. Lohmueller, D. A. Hafler, J. R. Oksenberg, S. L. Hauser, M. W. Smith, S. J. O'Brien, D. Altshuler, M. J. Daly, and D. Reich 2004. "Methods for High-Density Admixture Mapping of Disease Genes." *American Journal of Human Genetics* 74 (5): 979–1000.

Peters, N., A. Rose, and K. Armstrong. 2004. "The Association between Race and Attitudes about Predictive Genetic Testing." *Cancer Epidemiology Biomarkers & Prevention* 13 (3): 361–5.

Peters, E., J. Hibbard, P. Slovic, and N. Dieckmann. 2007. "Numeracy Skill and the Communication, Comprehension, and Use of Risk-Benefit Information." *Health Affairs* 26 (3): 741–48.

Piccart-Gebhart, M. J., M. Procter, B. Leyland-Jones, A. Goldhirsch, M. Untch, I. Smith, L. Gianni, J. Baselga, R. Bell, C. Jackisch, D. Cameron, M. Dowsett, C. H. Barrios, G. Steger, C. S. Huang, M. Andersson, M. Inbar, M. Lichinitser, I. Láng, U. Nitz, H. Iwata, C. Thomssen, C. Lohrisch, T. M. Suter, J. Rüschoff, T. Suto, V. Greatorex, C. Ward, C. C. Straehle, E. McFadden, M. S. Dolci, R. D. Gelber, and Herceptin Adjuvant (HERA) Trial Study Team. 2005. "Trastuzumab after Adjuvant Chemotherapy in HER2-Positive Breast Cancer." *New England Journal of Medicine* 353 (16): 1659–72.

Puppala, S., D. K. Coletta, J. Schneider, S. L. Hu, V. S. Farook, T. D. Dyer, R. Arya, J. Blangero, R. Duggirala, R. A. DeFronzo, and C. P. Jenkinson 2011. "Genome-Wide Linkage Screen for Systolic Blood Pressure in the Veterans Administration Genetic Epidemiology Study (VAGES) of Mexican-Americans and Confirma-

tion of a Major Susceptibility Locus on Chromosome 6q14. 1." *Human Heredity* 71 (1): 1–10.

Romond, E. H., E. A. Perez, J. Bryant, V. J. Suman, C. E. Jr Geyer, N. E. Davidson, E. Tan-Chiu, S. Martino, S. Paik, P. A. Kaufman, S. M. Swain, T. M. Pisansky, L. Fehrenbacher, L. A. Kutteh, G. Vogel, D. W. Visscher, G. Yothers, R. B. Jenkins, A. M. Brown, S. R. Dakhil, E. P. Mamounas, W. L. Lingle, P. M. Klein, J. N. Ingle, and N. Wolmark 2005. "Trastuzumab Plus Adjuvant Chemotherapy for Operable HER2-Positive Breast Cancer." *New England Journal of Medicine* 353 (16): 1673–84.

Royal, C. D. M., and G. M. Dunston. 2004. "Changing the Paradigm from 'Race' to Human Genome Variation." *Nature Genetics* 36 (suppl 11): S5–7.

Salem, R. M., B. Pandey, E. Richard, M. M. Fung, E. P. Garcia, V. H. Brophy, N. J. Schork, D. T. O'Connor, and V. Bhatnagar. 2010. "The VA Hypertension Primary Care Longitudinal Cohort: Electronic Medical Records in the Post-Genomic era." *Health Informatics Journal* 16 (4): 274–74.

Schwarz, U. I., M. D. Ritchie, Y. Bradford, C. Li, S. M. Dudek, A. Frye-Anderson, R. B. Kim, D. M. Roden, and C. M. Stein. 2008. "Genetic Determinants of Response to Warfarin during Initial Anticoagulation." *New England Journal of Medicine* 358 (10): 999–1008.

Shields, A. E., M. Fortun, E. M. Hammonds, P. A. King, C. Lerman, R. Rapp, and P. F. Sullivan. 2005. "The use of Race Variables in Genetic Studies of Complex Traits and the Goal of Reducing Health Disparities: A Transdisciplinary Perspective." *American Psychologist* 60 (1): 77–103.

Shields, A. E., P. Shin, M. G. Leu, D. E. Levy, R. M. Betancourt, D. Hawkins, and M. Proser. 2007. "Adoption of Health Information Technology in Community Health Centers: Results of a National Survey." *Health Affairs* 26 (5): 1373–83.

Sivell, S., R. Iredale, J. Gray, and B. Coles. 2007. "Cancer Genetic Risk Assessment for Individuals at Risk of Familial Breast Cancer." *Cochrane Database of Systematic Reviews (Online)* 2: CD003721.

Smit, L. A. M., D. Heederik, G. Doekes, G. H. Koppelman, R. W. B. Bottema, D. S. Postma, and I. M. Wouters. 2011. "Endotoxin Exposure, CD14 and Wheeze among Farmers: A Gene-Environment Interaction." *Occupational and Environmental Medicine* 68 (11): 826–31.

Swede, H., D. I. Gregorio, S. H. Tannenbaum, J. A. Brockmeyer, C. Ambrosone, L. L. Wilson, M. A. Pensa, L. Gonsalves, R. G. Stevens, and C. D. Runowicz. 2011. "Prevalence and Prognostic Role of Triple-Negative Breast Cancer by Race: A Surveillance Study." *Clinical Breast Cancer* 11 (5): 332–41.

Teutsch, S. M., L. A. Bradley, G. E. Palomaki, J. E. Haddow, M. Piper, N. Calonge, W. D. Dotson, M. P. Douglas, and A. O. Berg. 2009. "The Evaluation of Genomic Applications in Practice and Prevention (EGAPP) Initiative: Methods of the EGAPP Working Group." *Genetics in Medicine* 11 (1): 3–3.

Thomson, J., D. Garry, M. Poncz, S. Center, B. Torok-Storb, A. Friedman, J. Cooke, J. D. G. Institutes, D. S. San Francisco, and A. Hatzopoulos. 2010. "Kaiser Permanente & UCSF Receive NIH Award for Genetic Epidemiology Research." *Journal of Investigative Medicine* 58 (1): 6–9.

Tishkoff, S. A., and B. C. Verrelli. 2003. "Patterns of Human Genetic Diversity: Implications for Human Evolutionary History and Disease." *Annual Review of Genomics and Human Genetics* 4: 293–340.

Tomalik-Scharte, D., A. Lazar, U. Fuhr, and J. Kirchheiner. 2007. "The Clinical Role of Genetic Polymorphisms in Drug-Metabolizing Enzymes." *Pharmacogenomics Journal* 8 (1): 4–15.

U.S. Department of Health and Human Services. 2000. *Healthy People 2010: Understanding and Improving Health.* Washington, DC: U.S. Government Printing Office.

Voight, B. F., L. J. Scott, V. Steinthorsdottir, A. P. Morris, C. Dina, R. P. Welch, E. Zeggini, C. Huth, Y. S. Aulchenko, G. Thorleifsson, L. J. McCulloch, T. Ferreira, H. Grallert, N. Amin, G. Wu, C. J. Willer, S. Raychaudhuri, S. A. McCarroll, C. Langenberg, O. M. Hofmann, J. Dupuis, L. Qi, A. V. Segre, M. van Hoek, P. Navarro, K. Ardlie, B. Balkau, R. Benediktsson, A. J. Bennett, R. Blagieva, E. Boerwinkle, L. L. Bonnycastle, K. B. Bostrom, B. Bravenboer, S. Bumpstead, N. P. Burtt, G. Charpentier, P. S., ChinesM. Cornelis, D. J. Couper, G. Crawford, A. S. F. Doney, K. S. Elliott, A. L. Elliott, M. R. Erdos, C. S. Fox, C. S. Franklin, M. Ganser, C. Gieger, N. Grarup, T. Green, S. Griffin, C. J. Groves, C. Guiducci, S. Hadjadj, N. Hassanali, C. Herder, B. Isomaa, A. U. Jackson, P. R. V. Johnson, T. Jorgensen, W. H. L. Kao, N. Klopp, A. Kong, P. Kraft, J. Kuusisto, T. Lauritzen, M. Li, A. Lieverse, C. M. Lindgren, V. Lyssenko, M. Marre, T. Meitinger, K. Midthjell, M. A. Morken, N. Narisu, P. Nilsson, K. R. Owen, F. Payne, J. R. B. Perry, A.-K. Petersen, C. Platou, C. ProencaI. Prokopenko, W. Rathmann, N. W. Rayner, N. R. Robertson, G. Rocheleau, M. Roden, M. J. Sampson, Saxena, R., B. M. Shields, P. Shrader, G. Sigurdsson, T. Sparso, K. Strassburger, H. M. Stringham, Q. Sun, A. J. Swift, B. Thorand, J. Tichet, T. Tuomi, R. M. van Dam, T. W. van Haeften, T. van Herpt, J. V. van Vliet-Ostaptchouk, G. B. Walters, M. N. Weedon, C. Wijmenga, J. Witteman, R. N. Bergman, S. Cauchi, F. S. Collins, A. L. Gloyn, U. Gyllensten, T. Hansen, W. A. Hide, G. A. Hitman, A. Hofman, D. J. Hunter, K. Hveem, M. Laakso, K. L. Mohlke, A. D. Morris, C. N. A., Palmer, P. P., PramstallerI. Rudan, E. Sijbrands, L. D. Stein, J. Tuomilehto, A. Uitterlinden, M. Walker, N. J. Wareham, R. M. Watanabe, G. R. Abecasis, B. O. Boehm, H. Campbell, M. J. Daly, A. T. Hattersley, F. B. Hu, J. B. Meigs, J. S. Pankow, O. Pedersen, H. E. Wichmann, I. Barroso, J. C. Florez, T. M. Frayling, L. Groop, R. Sladek, U. Thorsteinsdottir, J. F. Wilson, T. Illig, P. Froguel, C. M. van Duijn, K. Stefansson, and D. Altshuler, M., Boehnke, and M. I., McCarthy. 2010. "Twelve Type 2 Diabetes Susceptibility Loci Identified through Large-Scale Association Analysis." *Nature Genetics* 42 (7): 579–89.

Wilke, R. A., H. Xu, J. C. Denny, D. M. Roden, R. M. Krauss, C. A. McCarty, R. L. Davis, T. Skaar, J. Lamba, and G. Savova. 2011. "The Emerging Role of Electronic Medical Records in Pharmacogenomics." *Clinical Pharmacology and Therapeutics* 89 (3): 379–86.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of this article:

Appendix SA1: Author Matrix.
Appendix SA2: Decomposition Methods, the Counterfactual, and Sample Selection Bias Models.

Please note: Wiley-Blackwell is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.