

Demographic Bias in Human Cell Studies

Jessica Snyder¹, Iva Bojic^{1,2}, Aarom Gerow³, Carlo Ratti^{1,2}

¹*Massachusetts Institute of Technology, SENSEable City Lab*

²*Singapore-MIT Alliance for Research and Technology*

³*University of Chicago, Computation Institute*

December 21, 2016

More than 50,000 woman will be diagnosed with breast cancer between 2010-2020 **in the US alone**, costing an estimated \$48 billion in medical treatment [1, 2]. The National Institutes of Health (NIH) and the National Cancer Institute (NCI) invested a combined \$4.1 billion in breast cancer research from 2012 through 2014, with the goal of restoring the nation's breast cancer patient population to good health. With a stable and improving health care system, breast cancer survival rates are at a record high — more than 90% — increasing each year Center for Disease Control (CDC) statistics are available. **However, are all women, no matter their ethnicity, equally likely to survive?**

The modern paradigm in cancer research has established that many sub-types of breast cancer are unique pathologies. This motivates using specific traits of the tumor as well as the patient's genetics to make treatment decisions. ~~As the medical community pivots towards increasingly personalized care based on genetic markers, we reflect on the inertia of breast cancer research. After acknowledging that new genomic techniques change how we group patients, lets consider past use of ethnic groups.~~ Before gene sequencing was widely available, the CDC grouped the US breast cancer patient population by ethnicity, to identify who might be left out of improving health care options. The purpose was to expand healthy lifestyle choices, public awareness, diagnostics, and treatment options to include considerations of non-majority patients.

Here, we use the CDC ethnic groups to assess how well each group has been represented in the available bank of human cell lines and publication that use those cell lines. The breast cancer patient population grouped by CDC-defined ethnicities includes Asian, Black, Hispanic, American Indian and White. The more than 230k US women diagnosed with breast cancer are grouped by ethnicity, and show that breast cancer patients are predominantly White women, 190k or 83% of the entire incidence population. Black patients comprise 12% of the incidence population and the remaining 5% include Asian, Hispanic and Native American women. **The results show** the incidence rate of breast cancer among White woman has been decreasing since 1999, while **conversely, for the same period** the rate among Black and Asian women has increased (Figure 1A).

One ongoing success is the steadily declining morbidity for all patients since 1999 (Figure 1B). The CDC statistics, however, shows that Black patients face the worst prognosis for survival after diagnosis, ~~Black patients comprise more of the death population than the incidence: the incidence population fraction is less than the death population fraction (Figure 1C). Black patients are less likely to survive than others,~~ highlighting an acknowledged disparity in the efficacy of breast cancer treatment for Black women [3, 4]. **NamelySurprisingly**, when the fractions of each ethnicity in the incidence population are compared to the

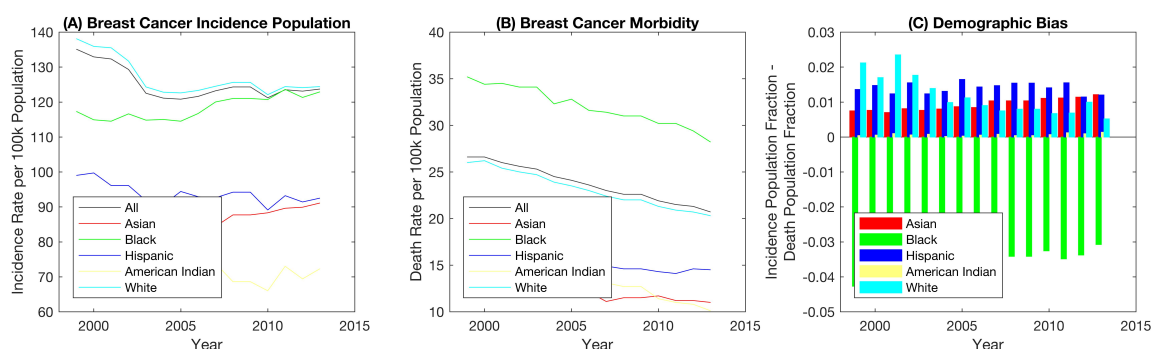


Figure 1: Incidence, morbidity and bias in US breast cancer population by ethnicity from 1998-2013.

death population, Black patients' fraction increases, meaning more Black woman die after a breast cancer diagnosis than other ethnicities (Figure 1C). ~~Black patients comprise more of the death population than the incidence, the incidence population fraction is less than the death population fraction, figure 1C.~~ White woman comprised the same 83% of the death population as they did the incidence population, while Asian, Hispanic and Native American fractions decrease slightly. ~~This is indicating that breast cancer treatment is less effective in treating Bblack~~ blacks than all other races.

Raising the survival rate of Black breast cancer patients to the national average requires a critical review of societal, behavioral and biological factors, broadly categorized as societal or extrinsic factors (largely the responsibility of policy research) and the biological, intrinsic factors such as genetics and cell structure within tumor sub-types (largely the responsibility of medical researchers and clinicians) [5, 6]. Socioeconomic-dependence of high-quality health care access and inclusive diagnostic techniques are important without question [7]. Failing to diagnosis at early stages, as well as faster growing, more aggressive tumor sub-types, disproportionately effect Black patients [8]. This disparity is attributal to a mixture of extrinsic and intrinsic problems affecting Black women. Could it be due to access to medical care? Yes. Might it also explained by bias in medical research? Possibly. Here, we review breast cancer research tools and their representation in research publications.

Data-driven analyses of disease burden and treatment effectiveness have been used in the past to guide medical research funding and health policy decisions towards the most productive population-level outcomes [9]. Vetting the influence of intrinsic factors of breast cancer is critical to guide the action of fundamental medical research, and in particular, the demographics of cell lines used in foundational research. Are there intrinsic factors of the tumor sub-types causing morbidity, which are disproportionately present in ~~in Bblack~~ black breast cancer patients? Environmental exposure and genetic background determine the a cancer's subtypes, aggressiveness, and ultimately, patient prognosis. Highly aggressive tumor subtypes are correlated to risk factors such as obesity, which disproportionately effects Bblack women. Self-reported race correlated to breast cancer properties, such as estrogen receptor status and stage of diagnosis, more than <XX???XXX> percent African ancestry, suggesting social determinants shape the disease physiology more than genetic predispositions [5]. Currently, alternative treatment options for patients with a genetic predisposition for hormone receptor-negative breast cancers, which are also likely Bblack, is limited [10].

The use of cell models may hinder generalization of results. Select biopsies from patients are cataloged and inventoried, creating a bank of cell line standards which allows experimental results to be compared across time and between labs. Because cell lines are derived from patients, the they carry unique genetic traits specific to the individual, and which diminish the applicability of some results to other members of

the patient population. One way to help avoid this limitation is to diversify the catalog of available cell lines. How the research community uses available cell lines shapes how representative finding will be and the efficacy of developed treatments among the patient population.

Comparing the fraction of an ethnicity in the incidence population, death population, cell line bank and publication record, we expect the most prevalent ethnicity in the incidence and death population to be the same. We can also expect the ethnicities to be the proportionally represented in the donor to the cell lines and for those cell lines to be the most used in research. White patients represent 83% of the incidence population and the same percentage of the death population. Whites comprise slightly less — though still a majority — of the donated cell lines, and more than 95% of cell lines cited in the publication record were from people of **W**hite ethnicity. So far, the picture of **W**hites' representation appears proportional. Research effort should reflect the greatest need, representing the majority of patients at greatest risk.

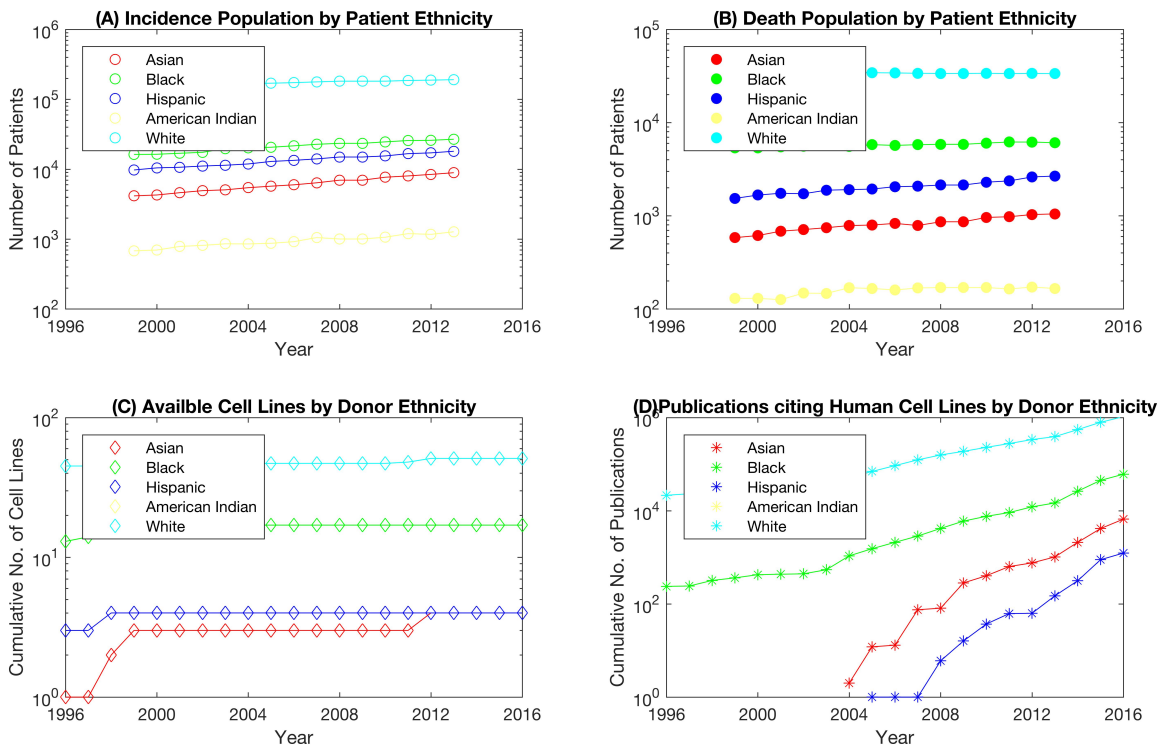


Figure 2: From top left: (A) Incidence of breast cancer from 1999 to 2013, grouped by ethnicity. (B) Death counts from breast cancer by ethnicity. (C) Proportion of available cell lines by ethnicity. (D) Proportion of citations of cell lines in published research, grouped by ethnicity.

The breast cancer patient population includes all ethnic groups. Sorting the fraction each group comprises by the incidence number shows the incidence rate is in line with the population sizes: White, Black, Hispanic, Asian and Native American (Figure 2A). Matching the trend, the death population shows the same order from most populous to least (Figure 2B). Given that research is limited by resources and time, it is not surprising that the breast cancer cell line bank composition reflects the incidence and death population proportions (Figure 2C). Of the 154 cell lines available, 50%, 78 cell lines were formed from donors of

unknown ethnicity. Of breast cancer cell lines with known donor ethnicity, 51 were from White donors, 17 from Black donors, 4 Hispanic, 4 Asian, and none from Native American donors.

Understanding the importance of standards, reproducibility and the legacy of scientific research, we expect to observe a bias in the publication record toward available cell lines. This would mean cell lines donated by White patients should be the most cited lines in research. Of the more than 1.2 million publications citing a cell line, 1.0 million cited a cell line donated by White patients, 60k cited a line from a Black patient, 6k cited a line an Asian patient, 1k cited Hispanic cell lines and none cited a cell lines from a Native Americans (as none are available; Figure 2D).

How does the fraction of women in the incidence population compare to publications that cite a breast cancer cell line donated by a White woman? Since 1999, the fraction of White women has been less than the fraction of cell lines donated by White women. That is, cell lines over-represent White women by more than 15% overall and nearly 19% in 2013 (Figure 3A). Other ethnicities have remained under-represented by the same metric. The under-representation of Hispanics increased from 3% to nearly 6% in 2013. Perhaps worse, the trend against Hispanic representation in cited cell lines is greater when comparing the death population fraction.

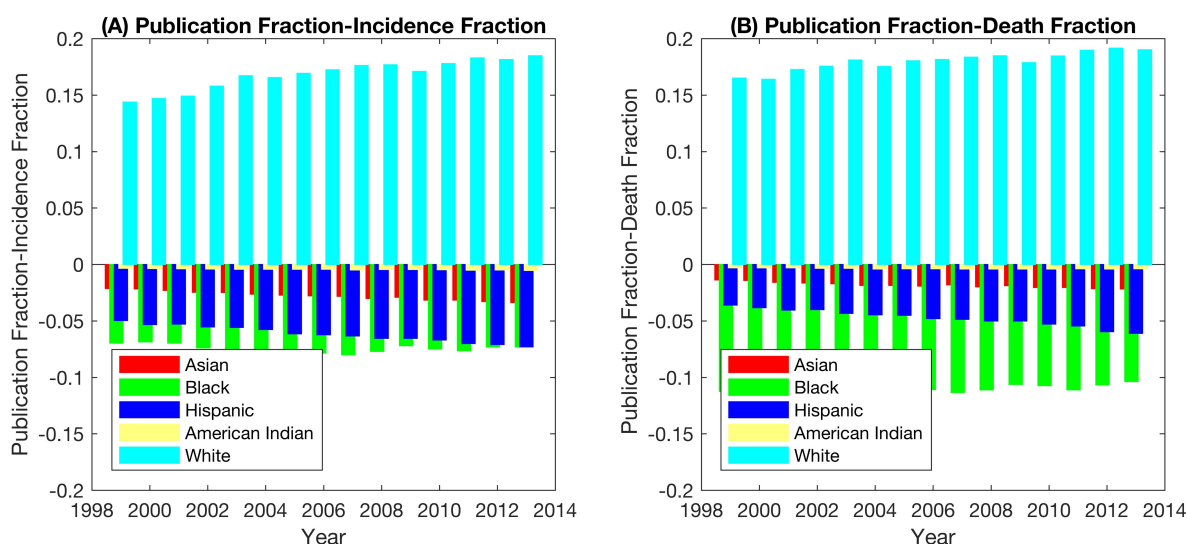


Figure 3: Representative bias in breast cancer medical research determined by comparing the fraction of each demographic group in the breast cancer death count to the proportional of cell lines cited in research publications and US patents from 1999 to 2013. White cell lines are proportionally over-represented, and getting worse, while Asian and Hispanic people are increasingly under-represented and Black patients remain steadily under-represented. Patents are similarly biased against non-Whites, more biased against Blacks, but with similar trends as research publications.

To this end, as Black patients comprise more of the death population than the incidence, CDC statistics suggests that breast cancer treatment is less effective in treating Black women than women of all other races. And even though previous studies showed Black breast cancer patients presented elevated risk for aggressive, intrinsic factors of breast cancer <NEED A CITATION>, instead of putting more resources into research using cell lines from Black donors, we have just shown totally the opposite effect – Whites are over-represent by more than 15%.

Today the only regulative we have is the one of the US Food and Drug Administration (FDA) who recognized genetic factors sensitize pharmaceutical efficacy and responded by implementing the Demographic Rule in 1998. The rule standardizes race and ethnicity representative during clinical trials based on the target therapeutic population as identified by the drug's sponsor. However, clinical trials are just the last step of the research process which typically progresses from cell models, to animal models, and finally human models. Should we extend this regulation also to the preclinical stage?

Recent efforts from NIH to balance sex in cell and animal studies would definitely suggest so [11]. Namely, similar to the Demographic Rule, in 1993 Congress passed the NIH Revitalization Act to increase the representation of women and minorities in clinical trials. However, despite of multiple calls for actions, publications often continued to neglect sex-based considerations and analysis in preclinical studies..

References

- [1] Angela B Mariotto, K Robin Yabroff, Yongwu Shao, Eric J Feuer, and Martin L Brown. Projections of the cost of cancer care in the United States: 2010–2020. *Journal of the National Cancer Institute*, 2011.
- [2] Hannah K Weir, Trevor D Thompson, Ashwini Soman, Bjørn Møller, and Steven Leadbetter. The past, present, and future of cancer incidence in the United States: 1975 through 2020. *Cancer*, 121(11):1827–1837, 2015.
- [3] Robert T Greenlee, Taylor Murray, Sherry Bolden, and Phyllis A Wingo. Cancer statistics, 2000. *CA: A Cancer Journal for Clinicians*, 50(1):7–33, 2000.
- [4] Amy Berkman, Bernard F Cole, Philips A Ades, Samantha Dickey, Stephen T Higgins, Amy Trentham-Dietz, Brian L Sprague, and Susan G Lakoski. Racial differences in breast cancer, cardiovascular disease, and all-cause mortality among women with ductal carcinoma in situ of the breast. *Breast Cancer Research and Treatment*, 148(2):407–413, 2014.
- [5] Kerryn W Reding, Christopher S Carlson, Orsalem Kahsai, Christina C Chen, Andrew McDavid, David R Doody, Chu Chen, Kimberly Lowe, Leslie Bernstein, Linda Weiss, et al. Examination of ancestral informative markers and self-reported race with tumor characteristics of breast cancer among black and white women. *Breast Cancer Research and Treatment*, 134(2):801–809, 2012.
- [6] Kevin Brennan and James M Flanagan. Is there a link between genome-wide hypomethylation in blood and cancer risk? *Cancer Prevention Research*, 5(12):1345–1357, 2012.
- [7] Vickie L Shavers, Linda C Harlan, and Jennifer L Stevens. Racial/ethnic variation in clinical presentation, treatment, and survival among breast cancer patients under age 35. *Cancer*, 97(1):134–147, 2003.
- [8] Nataliya G Batina, Amy Trentham-Dietz, Ronald E Gangnon, Brian L Sprague, Marjorie A Rosenberg, Natasha K Stout, Dennis G Fryback, and Oguzhan Alagoz. Variation in tumor natural history contributes to racial disparities in breast cancer stage at diagnosis. *Breast Cancer Research and Treatment*, 138(2):519–528, 2013.

- [9] Jane J Kim, Anna NA Tosteson, Ann G Zauber, Brian L Sprague, Natasha K Stout, Oguzhan Alagoz, Amy Trentham-Dietz, Katrina Armstrong, Sandi L Pruitt, Carolyn M Rutter, et al. Cancer models and real-world data: Better together. *Journal of the National Cancer Institute*, 108(2):djv316, 2016.
- [10] Dezheng Huo, Francis Ikpatt, Andrey Khramtsov, Jean-Marie Dangou, Rita Nanda, James Dignam, Bifeng Zhang, Tatyana Grushko, Chunling Zhang, Olayiwola Oluwasola, et al. Population differences in breast cancer: Survey in indigenous African women reveals over-representation of triple-negative breast cancer. *Journal of Clinical Oncology*, 27(27):4515–4521, 2009.
- [11] Janine A Clayton, Francis S Collins, et al. NIH to balance sex in cell and animal studies. *Nature*, 509(7500):282–283, 2014.