Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

The Astounding Breadth of Health Disparity: Phenome-Wide Effects of Race on Disease Risk

Jill M. Pulley, M.B.A., Rebecca N. Jerome, M.L.I.S., M.P.H., Gordon R. Bernard, M.D., Jana K. Shirey-Rice, Ph.D., Yaomin Xu, Ph.D., Consuelo H. Wilkins, M.D., M.S.C.I.

Acknowledgements: The authors express their sincere gratitude to Siwei Zhang for data analysis and graphics development, and to Ingrid Mayer, Tuya Pal, and Xiao-ou Shu for sharing clinical insights.

Abstract: Objective: We conducted a phenotype-wide association study (PheWAS) to compare diagnoses among Blacks with those of Whites in one health center in Tennessee using data from 1,883,369 patients.

Methods: We used our deidentified EHR, the Synthetic Derivative, to assess risk of diagnoses associated with Black as compared with White race using Firth logistic regression with covariates including age, sex, and density of clinical encounters.

Results: There were anchoring associations in both directions, including the highest increased risk for Blacks of having sickle cell anemia, and strongest *decreased* risk of basal cell carcinoma. Results included established areas of disparity and many novel associations.

Conclusions: PheWAS is a viable tool for calculating risk associated with any biomarker. The current analysis provide a new approach to generating hypotheses and understanding the breadth of health disparities. Future analyses will further explore causality, risk factors, and potential confounders not accounted for here.

Keywords: Health disparities■Racial disparities■Phenome-wide association study

Author affiliations: Jill M. Pulley, Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA; Rebecca N. Jerome, Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA; Gordon R. Bernard, Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA; Jana K. Shirey-Rice, Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, TN, USA; Yaomin Xu, Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN, USA; Consuelo H. Wilkins, Office of Health Equity, Division of Geriatrics, Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA; Department of Internal Medicine, Meharry Medical College, Nashville, TN, USA

Correspondence: Consuelo H. Wilkins, M.D., M.S.C.I., 1005 Dr. D.B. Todd, Jr., Blvd., Biomedical Sciences Building, Nashville, TN 37208, USA. Fax: 615-320-9457., email: consuelo.h.wilkins@Meharry-Vanderbilt.org

https://doi.org/10.1016/j.jnma.2020.08.009

INTRODUCTION

Health and healthcare disparities, and their myriad influences on the wellbeing of individuals in affected groups, are a major focus of initiatives in the United States. The National Institute on Minority Health and Health Disparities designated health disparity populations include "racial/ethnic minorities, socioeconomically disadvantaged populations, underserved rural populations, and sexual and gender minorities."[1] Health disparities are complex, and we know in particular that racial health disparities are multifactorial and that the 'variable' of race is often correlated with other factors (e.g., socioeconomic level, experience of discrimination, habits, health system interactions), each of which can independently and interdependently influence health.

Thoughtful design of strategies to mitigate the untoward effects of disparities requires a sound understanding of both the scope and magnitude of health disparities affecting a group. The phenome-wide association study (PheWAS) provides a powerful and validated methodology for visualizing the effect of an exposure on the relative risk of all diagnoses documented in the electronic health record (EHR).[2,3] While most commonly used to explore effects of genetic variation, PheWAS is readily adaptable to explore effects of other exposures such as race.

To aid in estimating the breadth of racial health disparities, we conducted a PheWAS to compare diagnoses among Blacks with those of Whites at one health center in Tennessee. The analysis was not undertaken to ignore the importance of other factors. Rather, it was intended to assess variance in disease risk holistically, across many diseases, to: 1) visualize and obtain insight into the overall phenome-wide burden, 2) evaluate concordance between individual disease risk in the PheWAS analysis compared to established disparity to demonstrate utility; and 3) identify disparities among rarer diseases that might be overlooked in public health literature.

MATERIALS AND METHODS

We extracted diagnoses for all individuals with documented White or Black race from our Derivative (SD), a deidentified version of the entire EHR at our medical center.[4] Our EHR currently includes more than 3 million patients of all ages; the current study extracted data from 1990 to 2019. We employed PheWAS to analyze variability in risk of all documented diagnoses associated with Black as compared with White race using demographic data from the SD (99% accuracy compared to genetic ancestry[5]). International Classification of Diseases, Ninth Revision (ICD-9) and ICD-10-CM codes were converted to phenotype codes (phecodes).[6,7] For each phecode, a

case was defined as having a minimum of two phecodes on different dates; controls were those having no related phecodes as is standard. Firth logistic regression was performed using R with covariates including age, sex, and the number of ages with a clinical encounter recorded. We report associations using the Bonferroni corrected p value of 2.7×10^{-5} (minimum detectable bound $p = 5 \times 10^{-324}$). For comparisons with the published literature, we extracted published odds ratios (OR) or calculated relative risk comparing Blacks and Whites.

RESULTS AND DISCUSSION

Figure 1 illustrates the phenome wide results, including phenotypes with increased (top section) or decreased risk (bottom section) among Blacks as compared with Whites, representing a diverse range of disease types and affected organ systems. A dynamic version of the PheWAS results, with hover-over labels for all phenotypes, is available online at https://prod.tbilab.org/phewas_race.

Dataset characteristics and anchoring

Our analysis included 1,883,369 patients, including 269,872 Blacks and 1,613,497 Whites. Mean age at last encounter was 37.8 years (range 0-90 years). Approximately 52.8% (n = 994,930) were females and 47.2% (n = 888,439) were males. We found anchoring associations as the absolute strongest OR in both directions in the data, including the highest *increased* risk for Blacks of sickle cell anemia[8] (OR 94.7; 95% CI 79.14, 114.51; $p < 5 \times 10^{-324}$), and strongest *decreased* risk of basal cell carcinoma[9] (OR 0.009; 95% CI 0.005, 0.01; $p < 5 \times 10^{-324}$). Agreement with previous research estimating risk magnitude was also apparent (Table 1). Notably, almost all pregnancy complications were higher risk in Black women whereas many congenital anomalies carried higher risk in Whites. There were some apparent areas of discordance. For example, the risk of low birth weight among Blacks was lower in our data (OR 1.30; 95% CI 1.25, 1.35; $p < 5 \times 10^{-324}$) as compared with the literature, while the odds of end stage renal disease was greater (OR 5.2; 95% CI 4.9-5.4; $p < 5 \times 10^{-324}$). Further, though the odds of diabetes or cerebrovascular disease were similar between our results and the literature, the odds of downstream sequelae including diabetic retinopathy and end stage renal disease were larger among Blacks in our data than estimates reported elsewhere.

Figure 1. Manhattan plot of PheWAS results comparing phenotype risk between Blacks and Whites. The plot illustrates represents log odds ratio for phenotypes with increased or decreased odds among Blacks as compared with Whites; significance is set at the Bonferroni-corrected threshold of $p < 2.7 \times 10^{-5}$.

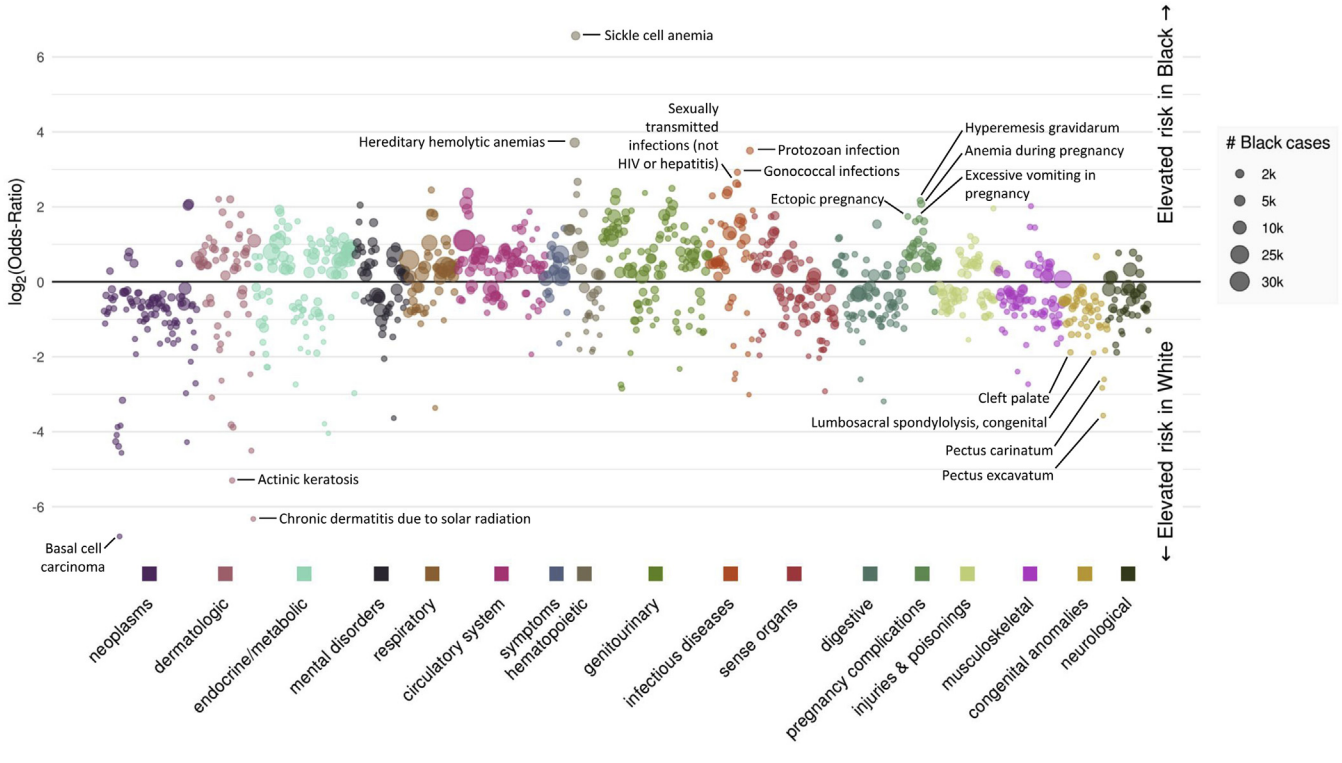


Table 1. Odd ratios (ORs) of selected diagnoses among blacks compared with whites.

Phecode description	PheWAS OR	Relative risk estimate from the literature	Source, relative risk data
Morbid obesity	1.9	2.0	Sturm and Hattori, 2013[15]
Preeclampsia and eclampsia	2.2	1.6	Fingar et al., 2017[16]
Hypertension	2.2	1.4	Office of Minority Health[17]
Diabetes	1.7	1.6	Office of Minority Health[18]
Systemic lupus erythematosus	1.8	3[a]	CDC[19]
Senile dementia	3.0	2.6	Chen and Zissimopoulos, 2018[20]
Cerebrovascular disease	1.5	1.5	Office of Minority Health[21]
Low birth weight[a]	1.3	2.4	Ratnasiri et al., 2018[22]
End stage renal disease	5.2	3.5	Office of Minority Health[18]
Diabetic retinopathy	2.7	1.6	Zhang et al., 2010[23]

[a]*This OR estimate is based on comparison of risk among women, consistent with the predominance of this disease among females.*

Greatest risk among Blacks across the phenome

The disease categories with the greatest racial disparity (as indicated by the highest ORs) include HIV, end stage renal disease, hypertension, uterine fibroids, diabetes, sarcoidosis, asthma, atherosclerosis, and glaucoma (Tables 1 and 2). PheWAS recapitulated widely established areas of disparity, but also identified several diseases for which significant disparities do not appear to be as well studied or understood.

The neoplasm category of phecodes showed the lowest relative volume of phenotypes among Blacks, with seven phenotypes with higher risk among Blacks and 89 neoplasm-related phenotypes with higher risk among Whites. High risk disease categories with the largest number of patients (shown by the size of the circle in Figure 1) generally conform to known prevalence among Blacks (although PheWAS represents health system data, not the general public) include those in Table 2. In addition, the risk of readily remediable high health disparity conditions, such as vitamin D deficiency (OR 1.54; 95% CI 1.49, 1.58; $p < 5 \times 10^{-324}$, remains apparent.

Other less well reported phenotypes with sizable populations included dermatophytosis (OR 2.00; 95% CI 1.89, 2.12; $p < 5 \times 10^{-324}$, iron deficiency anemia (OR 2.61; 95% CI 2.53, 2.71; $p < 5 \times 10^{-324}$), and fever of unknown origin (OR 1.67; 95% CI 1.64, 1.70; $p < 5 \times 10^{-324}$). While these are likely related to underlying disease such as diabetes, other immune dysfunction, or sickle cell, their appearance in the data might be reflecting the known ripple effect of disparities; that is, that individual health disparities are compounded, producing new, incremental increases in comorbidities over time in the Black population.

Immune-related, rare, and mental health diseases among blacks

Phenotypes carrying risks of immunocompromise, which may be particularly relevant in times of community outbreaks of communicable disease, are also notable. In addition to HIV and type 2 diabetes, we also see increased risk of various autoimmune diseases conferring risk of immune compromise due to the disease process and/or need for immunosuppressing treatment regimens (Table 2). Less reported in the public health literature than common diseases, Blacks have an increased risk of many rare diseases (Table 2 and Figure 2). Blacks in this analysis also have an increased risk of many psychiatric diagnoses (Table 2).

Utility of PheWAS in assessing relative risk

Using a large disease-agnostic, real world database of diagnoses, we applied PheWAS which can calculate risk associated with any biomarker (here, we used race as the social construct). The data are credible, recapitulating known relative risk. Appropriate disease complexity is reflected (such as, a cluster of pregnancy-related complications). Long-term consequences of risk factors (e.g., cerebral atherosclerosis) are also present in the data (e.g., dementia).

Table 2. Risk of additional anchoring, immune-related, mental health, or rare disease diagnoses among blacks compared with whites.

Category	Phecode	Phenotype	Odds ratio	Confidence interval	P-value
Additional anchoring diagnoses	071	HIV	5.08	(4.84, 5.32)	$<5 \times 10^{-324}$
	218.1	Uterine leiomyoma	4.21	(4.03, 4.39)	$<5 \times 10^{-324}$
	697	Sarcoidosis	3.62	(3.32, 3.95)	$<5 \times 10^{-324}$
	495	Asthma	2.06	(2.02, 2.10)	$<5 \times 10^{-324}$
	440	Atherosclerosis	1.40	(1.32, 1.48)	$<5 \times 10^{-324}$
	365	Glaucoma	2.41	(2.33, 2.50)	$<5 \times 10^{-324}$
Immune-related diseases	695.4	Lupus	1.87	(1.75, 2.00)	$<5 \times 10^{-324}$
	704.11	Alopecia areata	1.66	(1.43, 1,92)	$<1.56 \times 10^{-10}$
	709.4	Polymyositis	2.55	(2.04, 3.15)	$<5.66 \times 10^{-15}$
	250.1	Type 1 diabetes	1.30	(1.25, 2.35)	$<5 \times 10^{-324}$
	242.1	Graves' disease	1.30	(1.20, 1.41)	$<7.04 \times 10^{-10}$
Rare diseases	364.41	Keratoconus	3.34	(2.87, 3.88)	$<5 \times 10^{-324}$
	731.1	Paget's disease	4.05	(2.96, 5.48)	$<1.89 \times 10^{-15}$
	709.4	Polymyositis	2.55	(2.04, 3.15)	$<5.66 \times 10^{-15}$
	433.32	Moyamoya disease	2.77	(2.11, 3.61)	$<2.81 \times 10^{-12}$
	204.4	Multiple myeloma	1.40	(1.28, 1.54)	$<2.02 \times 10^{-12}$
	281.1	Megaloblastic anemia	1.44	(1.30, 1.59)	$<3.03 \times 10^{-11}$
	270.33	Amyloidosis	1.70	(1.43, 2.00)	$<3.61 \times 10^{-9}$
	261.41	Rickets	1.78	(1.39, 2.26)	$<8.53 \times 10^{-6}$
	270.34	Alpha-1-antitrypsin deficiency	0.06	(0.02, 0.14)	$<5 \times 10^{-324}$
	499	Cystic fibrosis	0.10	(0.07, 0.13)	$<5 \times 10^{-324}$
	270.12	Phenylketonuria	0.07	(0.02, 0.16)	$<5 \times 10^{-324}$
	253.5	Pituitary dwarfism	0.15	(0.12, 0.19)	$<5 \times 10^{-324}$
	386.1	Meniere's disease	0.32	(0.26, 0.39)	$<5 \times 10^{-324}$
	530.15	Eosinophilic esophagitis	0.33	(0.28, 0.39)	$<5 \times 10^{-324}$
	752.11	Spina bifida	0.37	(0.31, 0.43)	$<5 \times 10^{-324}$
	359.1	Muscular dystrophies	0.41	(0.32, 0.51)	$<5 \times 10^{-324}$
	270.1	Disturbances of amino-acid transport	0.42	(0.34, 0.53)	$<5 \times 10^{-324}$

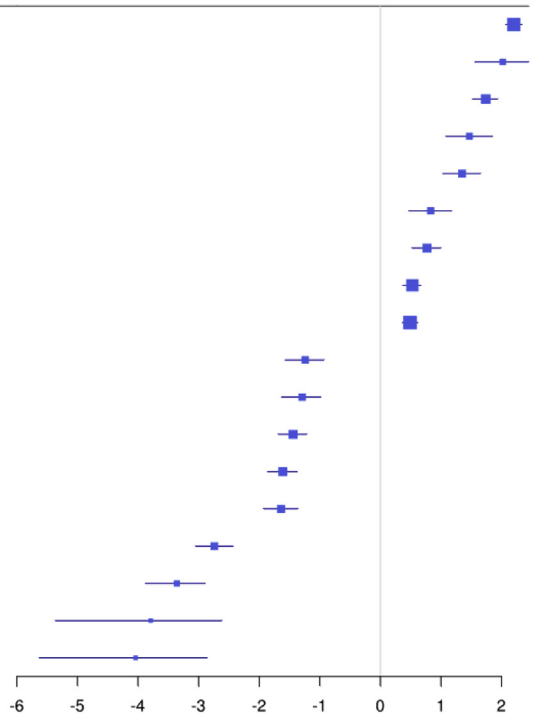
continued…

continued…

Category	Phecode	Phenotype	Odds ratio	Confidence interval	P-value
Mental health diseases	295.1	Schizophrenia	2.99	(2.82, 3.17)	$<5 \times 10^{-324}$
	290.1	Dementias	2.06	(1.95, 2.18)	$<5 \times 10^{-324}$
	312	Conduct disorders	1.89	(1.81, 1.99)	$<5 \times 10^{-324}$
	295.3	Psychosis	1.84	(1.72, 1.96)	$<5 \times 10^{-324}$
	292.6	Hallucinations	1.76	(1.58, 1.96)	$<5 \times 10^{-324}$
	305.21	Anorexia nervosa	0.08	(0.05, 0.12)	$<5 \times 10^{-324}$
	300.3	Obsessive-compulsive disorders	0.24	(0.20, 0.28)	$<5 \times 10^{-324}$
	300	Anxiety disorders	0.59	(0.58, 0.61)	$<5 \times 10^{-324}$
	313.2	Tics and stuttering	0.49	(0.42, 0.58)	$<5 \times 10^{-324}$
	301	Personality disorders	0.58	(0.52, 0.65)	$<5 \times 10^{-324}$
	296.1	Bipolar	0.73	(0.70, 0.77)	$<5 \times 10^{-324}$
	313.3	Autism	0.74	(0.70, 0.78)	$<5 \times 10^{-324}$
	296	Mood disorders	0.78	(0.77, 0.80)	$<5 \times 10^{-324}$
	313	Pervasive developmental disorders	0.86	(0.84, 0.89)	$<5 \times 10^{-324}$

Figure 2. Forest plot of PheWAS results comparing phenotype risk between Blacks and Whites among rare diseases. The plot illustrates represents log odds ratio for rare disease phenotypes with increased or decreased odds among Blacks as compared with Whites; significance is set at the Bonferroni-corrected threshold of $p < 2.7 \times 10^{-5}$.

Phenotype	Number of Black Cases	log2(Odds-Ratio)
Acquired acanthosis nigricans	1026	2.2
Osteitis deformans	59	2.02
Keratoconus	263	1.74
Moyamoya disease	84	1.47
Polymyositis	118	1.35
Rickets or osteomalacia	102	0.83
Amyloidosis	178	0.77
Megaloblastic anemia	505	0.53
Multiple myeloma	635	0.49
Disturbances of amino-acid transport	85	-1.24
Muscular dystrophies	84	-1.29
Spina bifida	154	-1.44
Eosinophilic esophagitis	154	-1.61
Meniere's disease	107	-1.64
Pituitary dwarfism	87	-2.74
Cystic fibrosis	40	-3.36
Phenylketonuria [PKU]	4	-3.79
Alpha-1-antitrypsin deficiency	4	-4.04



The spectrum of risks noted above are concordant with those inducing increased risk of COVID-19 infection: hypertension, diabetes, heart disease, asthma, obesity, and immune compromising conditions are likely playing a significant role in the increased COVID-19 disease severity and mortality experienced by Blacks in communities across the United States. The implications of these issues are potentially further worsened by delayed or cancelled health visits among those who cannot access telehealth formats.

Limitations

All of the limitations of the PheWAS method apply to this work, and have been described.[2] We note several of particular relevance to the current report. First, these codes do not separate biologic risk from risks associated with systematic differences in health system factors such as utilization or diagnostic biases; for example, the differences in mental health conditions are also concordant with previous literature on systemic biases in diagnoses among Blacks as compared with Whites.[10-13]

Other important factors also affect health and healthcare disparities and may lead to selection bias, including access to care, trust in the health system, and insurance status. For example, we observed many fewer diagnostic codes indicating neoplasms among Blacks, in contradiction with the published literature. This discrepancy is perhaps explained at least in part by insurance characteristics; many of our cancer clinics do not accept Medicaid; further, cancer incidence, as estimated in the current study, and mortality are different issues. Incorporation of data representing additional key exposures and outcomes such as these into future modeling will further inform our discussion of the breadth and implications of health and healthcare disparities. Despite these limitations, PheWAS represents a useful complement to existing approaches to visualizing health disparities, can highlight diseases of particular relevance to various audiences, and aid in decision making regarding high priority health disparities research and other programs.

IMPLICATIONS

As stated, any given disease can have many individual (but not independent) risk factors such as genetics, socioeconomics, lifestyle, healthcare access, stressors, environmental exposures, and many others. But these factors converge in the Black population to produce drastically

poorer health. All of the multifactorial risks that correlate with race and contribute to poorer health are implicitly included within the aggregate results described above, experienced in the real world in their composite by the individuals whose diagnoses comprise these data. As health systems charged with maintaining the health of the public, we need to better understand and recognize the overwhelming disparity that exists among Black patients, both a single disease at a time, and in their totality. Indeed, the preponderance of health risk in Blacks culminates in variable longevity; Whites live on average 4 years longer than Blacks.[14] Poorer health is an important driver of that loss of life, with socioeconomic and other factors being principal underlying components.

FUNDING

The project described was supported by CTSA award No. UL1 TR002243 from the National Center for Advancing Translational Sciences. Its contents are solely the responsibility of the authors and do not necessarily represent official views of the National Center for Advancing Translational Sciences or the National Institutes of Health.

CONFLICT OF INTEREST

None.

REFERENCES

1. Alvidrez, J., Castille, D., Laude-Sharp, M., Rosario, A., & Tabor, D. (2019). The national Institute on minority health and health disparities research framework. *Am J Public Health, 109*(Suppl 1), S16–S20.
2. Denny, J. C., Ritchie, M. D., Basford, M. A., et al. (2010). PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. *Bioinformatics, 26*(9), 1205–1210.
3. Denny, J. C., Bastarache, L., Ritchie, M. D., et al. (2013). Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol, 31*(12), 1102–1110.
4. Danciu, I., Cowan, J. D., Basford, M., et al. (2014). Secondary use of clinical data: the Vanderbilt approach. *J Biomed Inform, 52*, 28–35.
5. Hall, J. B., Dumitrescu, L., Dilks, H. H., Crawford, D. C., & Bush, W. S. (2014). Accuracy of administratively-assigned ancestry for diverse populations in an electronic medical record-linked biobank. *PLoS One, 9*(6), e99161.
6. Wei, W.-Q., Bastarache, L. A., Carroll, R. J., et al. (2017). Evaluating phecodes, clinical classification software, and ICD-9-CM codes for phenome-wide association studies in the electronic health record. *PLoS One, 12*(7), e0175508.
7. Wu, P., Gifford, A., Meng, X., et al. (2019). Mapping ICD-10 and ICD-10-CM codes to phecodes: workflow development and initial evaluation. *JMIR Med Inform, 7*(4), e14325.
8. CDC. (2016). *Data & Statistics on Sickle Cell Disease | CDC*. Centers for Disease Control and Prevention. https://www.cdc.gov/ncbddd/sicklecell/data.html. Accessed February 3, 2020.
9. Guy, G. P., Thomas, C. C., Thompson, T., et al. (2015). Vital signs: melanoma incidence and mortality trends and projections - United States, 1982-2030. *MMWR Morb Mortal Wkly Rep, 64*(21), 591–596.
10. Akinhanmi, M. O., Biernacka, J. M., Strakowski, S. M., et al. (2018). Racial disparities in bipolar disorder treatment and research: a call to action. *Bipolar Disord, 20*(6), 506–514.
11. Medlock, M., Weissman, A., Wong, S. S., et al. (2017). Racism as a unique social determinant of mental health: development of a didactic curriculum for psychiatry residents. *MedEdPORTAL, 13*.
12. DeCoux Hampton, M. (2007). The role of treatment setting and high acuity in the overdiagnosis of schizophrenia in African Americans. *Arch Psychiatr Nurs, 21*(6), 327–335.
13. Schwartz, R. C., & Blankenship, D. M. (2014). Racial disparities in psychotic disorder diagnosis: a review of empirical literature. *World J Psychiatry, 4*(4), 133–140.
14. Bond, M. J., & Herman, A. A. (2016). Lagging life expectancy for black men: a public health imperative. *Am J Public Health, 106*(7), 1167–1169.
15. Sturm, R., & Hattori, A. (2013). Morbid obesity rates continue to rise rapidly in the United States. *Int J Obes, 37*(6), 889–891.
16. Fingar, K., Mabry-Hernandez, I., Ngo-Metzger, Q., Wolff, T., Steiner, C., & Elixhauser, A. (2017). *Delivery Hospitalizations Involving Preeclampsia and Eclampsia, 2005-2014 #222*. Agency for Healthcare Research and Quality. https://hcup-us.ahrq.gov/reports/statbriefs/sb222-Preeclampsia-Eclampsia-Delivery-Trends.jsp?utm_source=ahrq&utm_medium=en-1&utm_term=&utm_content=1&utm_campaign=ahrq_en4_25_2017. Accessed February 12, 2020.
17. Heart disease and African Americans - the office of minority health. https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=4&lvlid=19. Accessed February 12, 2020.
18. Diabetes and African Americans - the office of minority health. https://minorityhealth.hhs.gov/omh/browse.aspx?lvl=4&lvlid=18. Accessed May 27, 2020.
19. Lupus in women | CDC. https://www.cdc.gov/lupus/basics/women.htm. Accessed February 12, 2020.

20. Chen, C., & Zissimopoulos, J. M. (2018). Racial and ethnic differences in trends in dementia prevalence and risk factors in the United States. *Alzheimers Dement (N Y), 4*, 510–520.

21. Stroke and African Americans - the office of minority health. https://www.minorityhealth.hhs.gov/omh/browse.aspx?lvl=4&lvlid=28. Accessed May 27, 2020.

22. Ratnasiri, A. W. G., Parry, S. S., Arief, V. N., et al. (2018). Recent trends, risk factors, and disparities in low birth weight in California, 2005–2014: a retrospective study. *Maternal Health Neonatol Perinatol, 4*(1), 15.

23. Zhang, X., Saaddine, J. B., Chou, C.-F., et al. (2010). Prevalence of diabetic retinopathy in the United States, 2005-2008. *J Am Med Assoc, 304*(6), 649–656.