

Longitudinal assessment of demographic representativeness in the Medical Imaging and Data Resource Center open data commons

Heather M. Whitney¹,^{a,b,*} Natalie Baughan²,^{a,b} Kyle J. Myers³,^{b,c} Karen Drukker⁴,^{a,b}
Judy Gichoya⁵,^{b,d} Brad Bower⁶,^{b,e} Weijie Chen⁷,^{b,f} Nicholas Gruszaszkas⁸,^{a,b}
Jayashree Kalpathy-Cramer⁹,^{b,g} Sanmi Koyejo¹⁰,^{b,h} Rui C. Sá¹¹,^{b,e,i}
Berkman Sahiner¹²,^{b,f} Zi Zhang¹³,^{b,j} and Maryellen L. Giger¹⁴,^{a,b}

^aUniversity of Chicago, Chicago, Illinois, United States

^bThe Medical Imaging and Data Resource Center (midrc.org)

^cPuente Solutions LLC, Phoenix, Arizona, United States

^dEmory University, Atlanta, Georgia, United States

^eNational Institutes of Health, Bethesda, Maryland, United States

^fUnited States Food and Drug Administration, Silver Spring, Maryland, United States

^gUniversity of Colorado Anschutz Medical Campus, Aurora, Colorado, United States

^hStanford University, Stanford, California, United States

ⁱUniversity of California, San Diego, La Jolla, California, United States

^jJefferson Health, Philadelphia, Pennsylvania, United States

ABSTRACT. Purpose: The Medical Imaging and Data Resource Center (MIDRC) open data commons was launched to accelerate the development of artificial intelligence (AI) algorithms to help address the COVID-19 pandemic. The purpose of this study was to quantify longitudinal representativeness of the demographic characteristics of the primary MIDRC dataset compared to the United States general population (US Census) and COVID-19 positive case counts from the Centers for Disease Control and Prevention (CDC).

Approach: The Jensen-Shannon distance (JSD), a measure of similarity of two distributions, was used to longitudinally measure the representativeness of the distribution of (1) all unique patients in the MIDRC data to the 2020 US Census and (2) all unique COVID-19 positive patients in the MIDRC data to the case counts reported by the CDC. The distributions were evaluated in the demographic categories of age at index, sex, race, ethnicity, and the combination of race and ethnicity.

Results: Representativeness of the MIDRC data by ethnicity and the combination of race and ethnicity was impacted by the percentage of CDC case counts for which this was not reported. The distributions by sex and race have retained their level of representativeness over time.

Conclusion: The representativeness of the open medical imaging datasets in the curated public data commons at MIDRC has evolved over time as the number of contributing institutions and overall number of subjects have grown. The use of metrics, such as the JSD support measurement of representativeness, is one step needed for fair and generalizable AI algorithm development.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JMI.10.6.061105](https://doi.org/10.1117/1.JMI.10.6.061105)]

*Address all correspondence to Heather M. Whitney, hwhitney@uchicago.edu

Keywords: COVID-19; representativeness; fairness; artificial intelligence; medical imaging

Paper 23027SSR received Jan. 31, 2023; revised Jun. 21, 2023; accepted Jun. 23, 2023; published Jul. 18, 2023.

1 Introduction

Since the first identification of the SARS-CoV-2 coronavirus (and its associated infectious disease, COVID-19) in late 2019, there have been reports of differences in disease health outcomes by race, ethnicity, sex, and other demographics.¹⁻¹⁴ Additionally, the relative difference in impact of COVID-19 to demographic subgroups has also changed over time.^{15,16} Furthermore, differences in the utilization of medical imaging in healthcare have been observed in various demographic subgroups over the course of the pandemic.¹⁷⁻¹⁹ As a result, the use of medical imaging among different demographic subpopulations can be expected to change over time.

The Medical Imaging and Data Resource Center (MIDRC)²⁰ is a multi-institutional initiative designed to collect, curate, and share medical images and other related resources to support the development of artificial intelligence/machine learning (AI/ML) for diagnosis, treatment, and prognosis of COVID-19 and beyond. MIDRC is hosted at the University of Chicago, funded by the National Institute of Biomedical Imaging and Bioengineering, and co-led by the American College of Radiology[®] (ACR), the Radiological Society of North America (RSNA), and the American Association of Physicists in Medicine (AAPM). Studies are contributed to MIDRC by institutions via a pipeline that includes a collaborative partnership between the ACR[®], the RSNA, the AAPM and Gen 3, a data commons organization. Users can access the data under either a non-commercial research or a commercial use agreement.²¹ MIDRC places a strong emphasis on monitoring and increasing the representativeness of the data, both at specific instances in time and longitudinally, to help support the development of unbiased and generalizable algorithms.

The purpose of this study was to (1) introduce the use of a metric to measure representativeness of the imaging datasets compared to relevant groups and (2) report on the evolution of the representativeness since the ingestion of datasets from contributors began in August 2021.

2 Materials and Methods

2.1 Dataset

Data used in this study were composed of metadata for the imaging studies available at the MIDRC open data commons²² in the open-A1 and open-R1 datasets (i.e., those ingested by the ACR[®] and RSNA and curated and harmonized by AAPM and Gen3). In this study, we refer to this specific collection as “MIDRC data.” The metadata had been submitted by data contributors in accordance with the MIDRC data dictionary.²³ Assignment of unique patients into the open data commons occurs at the ingestion of data within MIDRC according to a multidimensional stratified sampling algorithm,²⁴ with ~80% of unique patients assigned to the open data commons and 20% of unique patients assigned to a sequestered data commons. The sequestration algorithm is designed and tested for balance among groupings including but not limited to demographic categories.

For the purposes of this study, demographic categories were analyzed as follows: age at index event (i.e., the first occurrence in MIDRC, usually the first COVID-19 test), sex, race, ethnicity, and the combination of race and ethnicity. The latter was used in accordance with guidance from the Office of Management and Budget, which identifies patients as Hispanic or, if non-Hispanic, by their race.²⁵ Patients may have multiple imaging studies in MIDRC but, for each unique patient, the characteristics at the index event were used in this study.

2.2 Comparison Groups

The demographic distributions of unique patients in the MIDRC data were compared against two relevant population distributions. Because at the time of this study all open-A1 and open-R1 data contributing to the MIDRC data described here had been collected in the United States, we compared the demographic distributions (1) between all cases (unique patients) within the MIDRC

data and population in the United States 2020 Census²⁶ and (2) between COVID-19 positive cases within the MIDRC data and population in the COVID-19 case surveillance public use data from the Centers for Disease Control and Prevention (CDC).²⁷

2.3 Statistical Analysis

The Jensen-Shannon distance^{28–30} (JSD) was used as a metric to measure the difference between any two population categorical distributions, with the two comparison groups being termed S and T in this study. It is based upon the Jensen-Shannon divergence³¹ (called D_{JS} in this study) and the Kullback-Leiber divergence D_{KL} . The D_{KL} is defined for two distributions as

$$D_{KL}(S||T) = \sum_x S(x) \log_2 \frac{S(x)}{T(x)}, \quad (1)$$

where $S(x)$ and $T(x)$ are the distribution functions of any two populations S and T , and x is the variable of interest which in this study is any of the demographic variables under investigation. Because all the demographic variables x in this study are discrete, the distribution functions are probability mass functions (i.e., represented by the fraction of patients at each bin of x). Subsequently, the JSD is defined as

$$JSD = \sqrt{D_{JS}(S||T)}, \quad (2)$$

where

$$D_{JS}(S||T) = \frac{1}{2} D_{KL}(S||M) + \frac{1}{2} D_{KL}(T||M), \quad (3)$$

and

$$M = \frac{1}{2}(S + T). \quad (4)$$

The logarithm within D_{KL} can be determined through other bases (such as the natural logarithm). When \log_2 is used, the D_{JS} and JSD are bounded between 0 and 1, which is advantageous for our purpose. The sum in equation (1) was taken over each bin x for which both comparison groups were non-zero. A JSD of zero indicates that there is no difference between compared distributions, while a JSD of 1 indicates that there is no similarity between them. In this study, more representative distributions (compared to the reference distribution) will have a lower JSD.

In this study, the JSD was used to compare the following distributions at each MIDRC batch ingestion date:

1. cumulative counts of all unique patients in the MIDRC data to the US Census counts ($JSD_{MIDRC (all) to census}$),
2. the cumulative counts of all unique COVID-19 positive cases in the MIDRC data to the cumulative COVID-19 positive counts (derived from case counts) reported by the CDC ($JSD_{MIDRC (C19+) to CDC (C19+)}$), and
3. the cumulative COVID-19 positive counts reported by the CDC to the US Census counts ($JSD_{CDC (C19+) to census}$).

The CDC to US Census comparison was used as a reference against which the comparison of MIDRC distributions can be considered.

Additionally, the temporal difference in the JSD was determined between the JSD when comparing all cases in MIDRC to the US Census and the JSD when comparing all COVID-19 positive cases to the COVID-19 positive case counts from the CDC. If this difference is positive, the distribution of COVID-19 positive unique patients in MIDRC to the CDC cumulative COVID-19 case counts is more representative than the distribution of all unique patients in MIDRC to the US general population. If this difference is negative, the distribution of all unique patients in MIDRC to the US general population is more representative than the distribution of COVID-19 positive unique patients in MIDRC to the CDC cumulative case counts.

Note that in this study, no measures of statistical difference were assessed, since the goal of the measure of representativeness here is to measure degree of similarity according to counts.

Additionally, no sampling of distributions was conducted, due to the nature of the data (counts of individuals), none of which are inherently considered samples in this study.

3 Results

3.1 Dataset

As of April 3, 2023 (the most recent batch ingestion date at time of manuscript preparation), there were 9 unique contributing sites and over 55,000 unique patients represented in the MIDRC data (Fig. 1).

The proportions of the MIDRC data by demographic category and COVID-19 status as of April 3, 2023 are given in Fig. 2.

The most recent distributions of unique patients within each demographic category, both within the MIDRC data and the comparison groups, are given in Fig. 3 along with JSD results. Longitudinal measurements of the demographic data for the MIDRC data are available in the [Supplementary Material](#).

The JSD measured changes in the similarity of the MIDRC data (both when considering all imaging studies and those from COVID-19 positive patients only) to their comparison groups (the United States general population and case counts from the CDC, respectively) (Figs. 4–8). The comparison of age at index to the United States general population and the cumulative case counts as recorded by the CDC has remained relatively stable in these sets of patients over time, with little difference in their level of representativeness (Fig. 4).

The comparison of MIDRC unique patients by sex to the United States general population and the cumulative case counts as recorded by the CDC has demonstrated more representativeness (i.e., more similarity) in the distribution of all unique patients to the United States general population (lower JSD) than the comparison of MIDRC positive patients to the cumulative case count from the CDC (higher JSD) (Fig. 5). Over time, there has been a slight increase in the difference of the comparisons as the proportion of male unique patients has increased.

The representativeness of MIDRC unique patients by race to the United States general population and the cumulative case counts as recorded by the CDC reached almost equal similarity within the MIDRC data in August 2022 (Fig. 6). However, it is important to note that the measurement of representativeness for all three comparisons is impacted by the proportion of subjects for which race is not reported, which is over 30% in the most recent CDC cumulative case counts and around 10% in the most recent distributions within the MIDRC data. The MIDRC data also have substantially higher proportions of unique patients with reported race as Black than the US general population and the CDC cumulative case counts.

The comparison of MIDRC unique patients by ethnicity to the United States general population has been substantially more similar than the comparison of MIDRC COVID-19 positive patients to the cumulative CDC case counts (Fig. 7). This is likely a result of the substantial percentage of cases within the CDC data for which ethnicity is not available (over 40%).

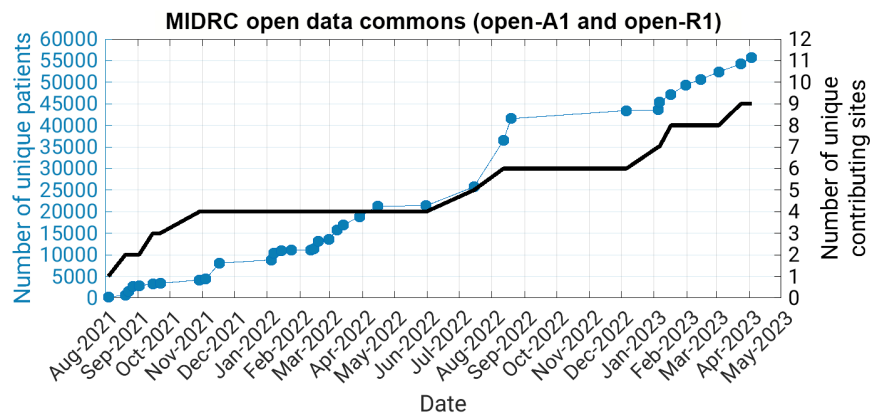


Fig. 1 Cumulative number of unique patients and number of unique contributing sites in the open MIDRC data commons since the launch of the data commons in August 2021 through time of manuscript preparation.

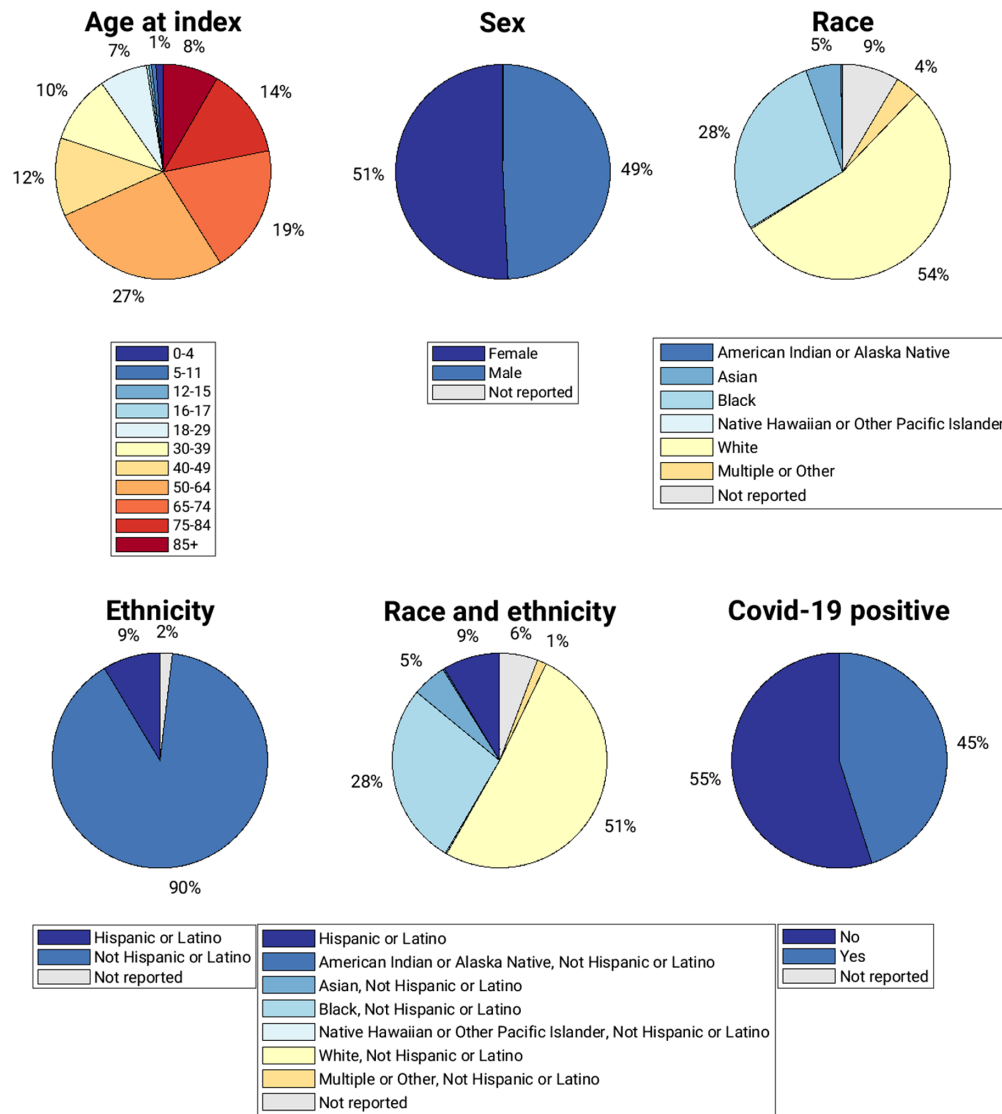


Fig. 2 Pie charts of the percentages of unique patients in the MIDRC data as of April 3, 2023 by demographic category and COVID-19 status. The presentation of demographic data in pie chart form here is the same as the bar graph representation for “MIDRC (all)” (blue bars) in each sub-figure of Fig. 3.

The comparison of MIDRC unique patients by the combination of race and ethnicity to the United States general population has been more similar than the comparison of MIDRC COVID-19 positive patients to the cumulative CDC case counts (Fig. 8). This may also be impacted by the substantial percentage of cases within the CDC data for which race and ethnicity is not available.

4 Discussion

The representativeness of the MIDRC data continues to change over time as the number of contributing institutions and the overall number of unique patients grow. The evolution of the impact of the COVID-19 pandemic to various demographic groups also continues to change over time, as shown by the changes in $JSD_{CDC(C19+)}_{to\ census}$ (not discussed in detail here). Using the JSD contributes to quantifying the comprehensive representativeness of the data and supports several initiatives related to health-related research and development at the federal level, including the strategic plan of the National Institute on Minority Health and Health Disparities³² (specifically, Goal 7: “ensure appropriate representation of minority and other health disparity

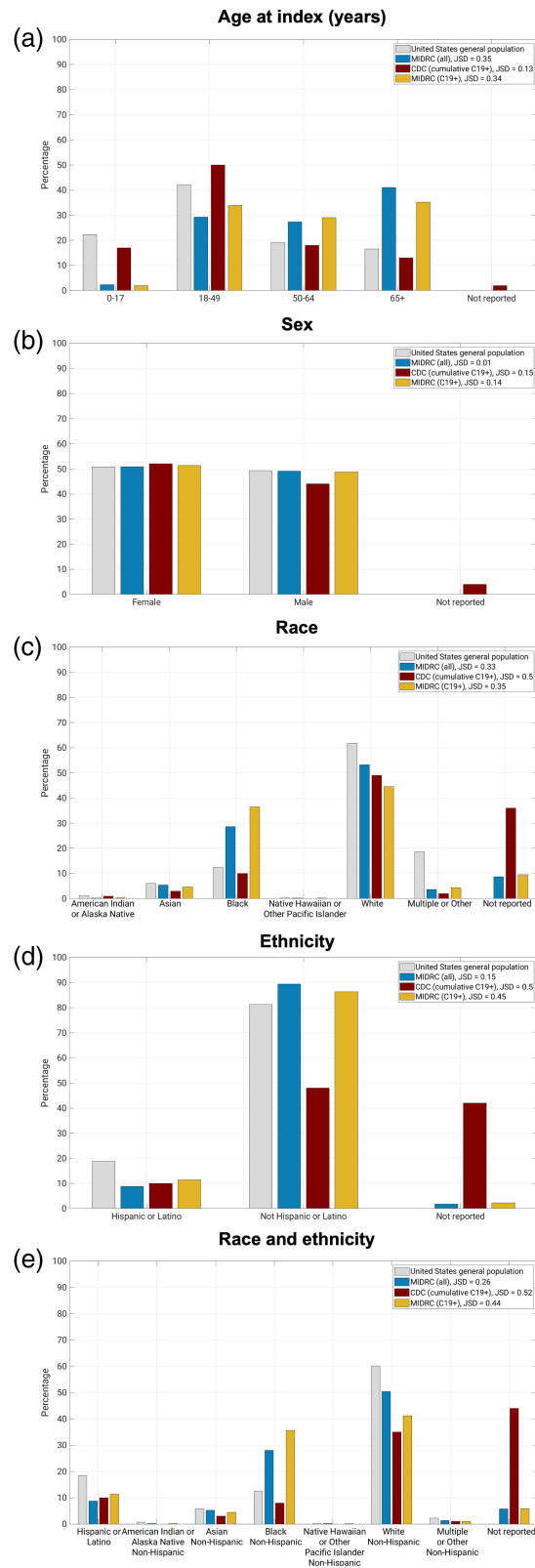


Fig. 3 Distributions of cases within the MIDRC data as of March 24, 2023 (the latest ingestion date that can be compared to CDC data) and comparison groups (US general population from the 2020 census and cumulative case counts from the CDC). In these figures, the JSD is shown for (1) all cases in the MIDRC data compared to the US general population ($JSD_{MIDRC (all) \text{ to census}}$), (2) cumulative CDC COVID-19 positive case counts compared to the US general population ($JSD_{CDC (C19+) \text{ to census}}$), and (3) MIDRC COVID-19 positive case counts compared to the CDC

Fig. 3 (Continued) COVID-19 positive case counts ($JSD_{MIDRC(C19+)} \text{ to } CDC(C19+)$). The JSD is bounded between 0 and 1, where 0 indicates that two distributions are the same as measured by the JSD and 1 indicates that they are completely different. MIDRC, The Medical Imaging and Data Resource Center; US, United States; CDC, Centers for Disease Control and Prevention; and C19+, cases positive for COVID-19. (a) Age at index; (b) sex; (c) race; (d) ethnicity; (e) race and ethnicity.

populations in NIH-funded research”) and action plans and guidance from the Food and Drug Administration.^{33–35}

The goal of the development of fair and generalizable AI/ML algorithms in medical imaging has rightfully been the topic of much attention.^{36–43} Goals of algorithmic fairness and generalizability involve those of equal outcomes (including equalized outcomes) and of equal performance.⁴⁴ It should be noted that defining the representativeness of data is a crucial part of developing and deploying algorithms with fairness and generalizability in mind. Indeed, defining representativeness is one of many careful procedures needed in AI/ML, with others including but not limited to definition of the purpose of data collection, aim of the model, and careful identification of the task. As has been noted by others,⁴⁵ representativeness can involve representativity of data in the sense of coverage of the “input space” (i.e., the training and/or the test data) and/or representativeness to population distributions. The use of the JSD generally can support the assessment of representativeness of distributions for either aim; in this study, we used the JSD to measure representativeness of the MIDRC data to demographic distributions.

Assessment of representativeness of data by demographic categories is but one part of ensuring fairness and generalizability at various stages of AI/ML pipelines. These include data collection (by identifying protected groups and their representation and addressing unequal representation in data through intentional collection efforts), model tuning and evaluation (such as comparing deployment data with training data across subgroups), and performance monitoring (including monitoring for data shifts, such as changes of impact in disease to subgroups over time^{44,46}). Population characterization (and potentially matching synthetically^{47,48}), cross-population modeling,^{49,50} and class balancing⁵¹ can be useful in AI/ML algorithm development to identify and avoid model bias.

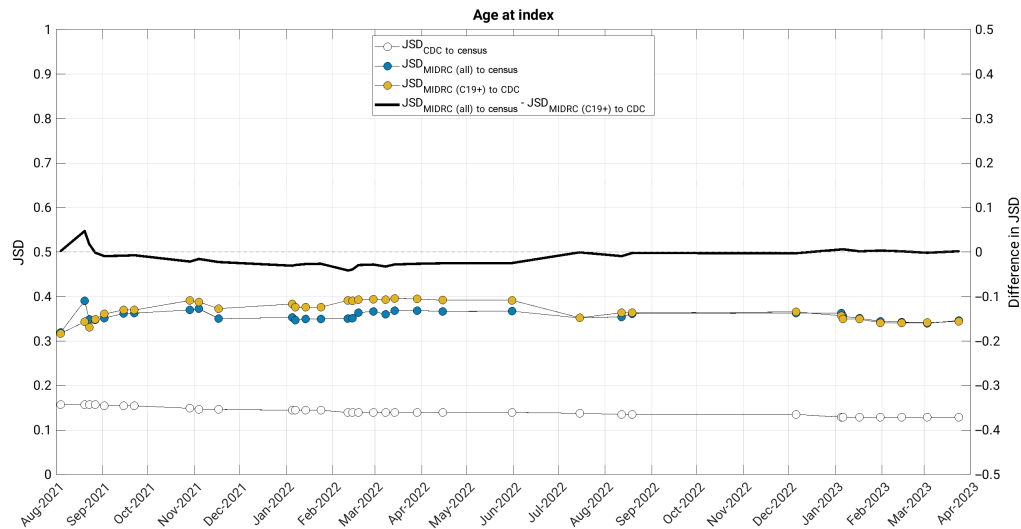


Fig. 4 The Jensen-Shannon distance (JSD) over time for age at index for (blue data markers) all unique patients, (gold data markers) all unique COVID-19 positive patients in the MIDRC data, and (white data markers) the JSD for comparing the CDC data to the US general population (for reference). The difference in JSD over time between all unique patients and all unique COVID-19 positive patients in the MIDRC data is also shown (black line). The similarity of both all unique patients and all unique COVID-19 positive patients in the MIDRC data has remained fairly constant to their respective comparison groups over time in terms of JSD. MIDRC, The Medical Imaging and Data Resource Center; US, United States; CDC, Centers for Disease Control and Prevention; and C19+, cases positive for COVID-19.

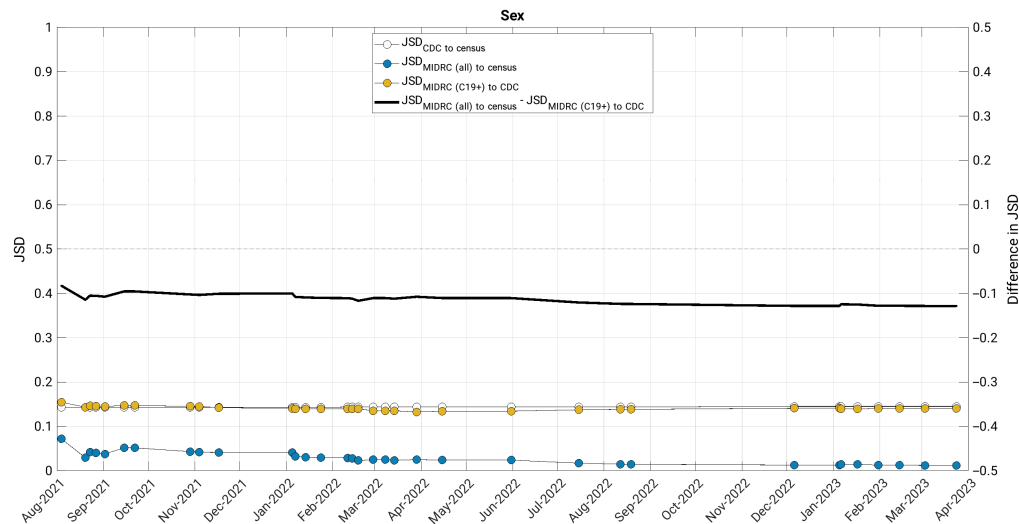


Fig. 5 The Jensen-Shannon distance (JSD) over time for sex for (blue data markers) all unique patients, (gold data markers) all unique COVID-19 positive patients in the MIDRC data, and (white data markers) the JSD for comparing the CDC data to the US general population (for reference). The difference in JSD over time between all unique patients and all unique COVID-19 positive patients in the MIDRC data is also shown (black line). The distribution of all unique cases in the MIDRC data has been more representative of the US population than the distribution of all unique COVID-19 cases to the CDC cumulative case counts, and this higher representativeness has slightly increased as the number of unique cases has increased. MIDRC, The Medical Imaging and Data Resource Center; US, United States; CDC, Centers for Disease Control and Prevention; and C19+, cases positive for COVID-19.

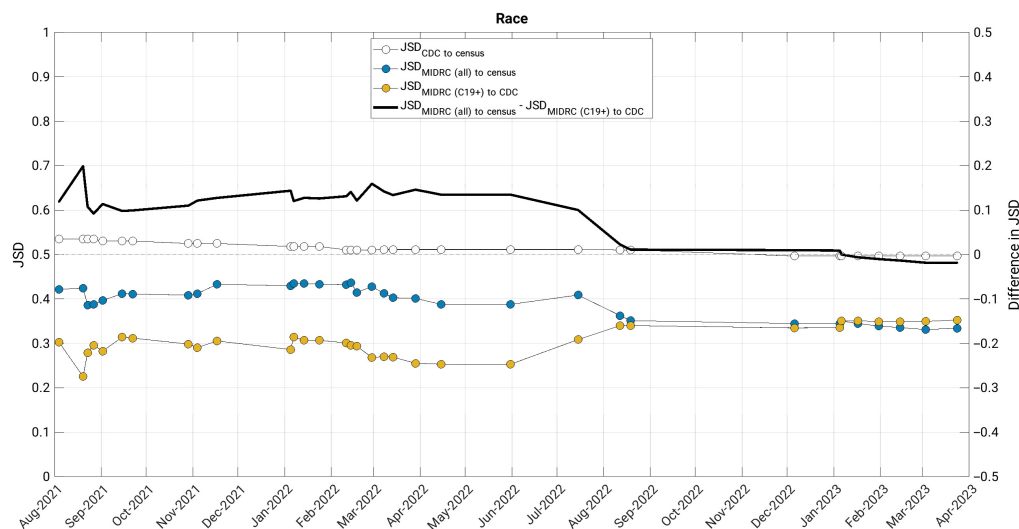


Fig. 6 The Jensen-Shannon distance (JSD) over time for race for (blue data markers) all unique patients, (gold data markers) all unique COVID-19 positive patients in the MIDRC data, and (white data markers) the JSD for comparing the CDC data to the US general population (for reference). The difference in JSD over time between all unique patients and all unique COVID-19 positive patients in the MIDRC data is also shown (black line). The distribution of unique patients in the MIDRC data has recently reached similar levels of representativeness to their comparison groups (difference in JSD approaching zero). MIDRC, The Medical Imaging and Data Resource Center; US, United States; CDC, Centers for Disease Control and Prevention; and C19+, cases positive for COVID-19.

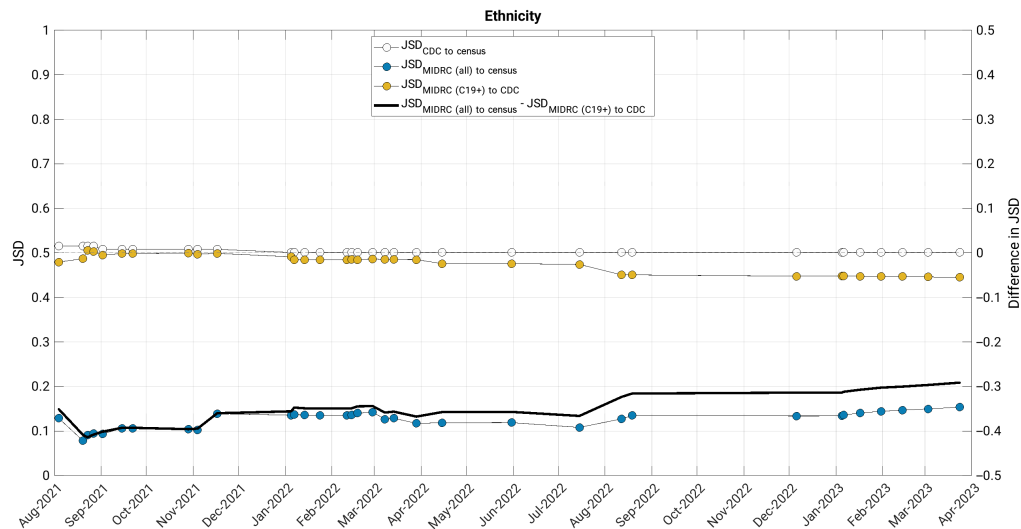


Fig. 7 The Jensen-Shannon distance (JSD) over time for ethnicity for (blue data markers) all unique patients, (gold data markers) all unique COVID-19 positive patients in the MIDRC data, and (white data markers) the JSD for comparing the CDC data to the US general population (for reference). The difference in JSD over time between all unique patients and all unique COVID-19 positive patients in the MIDRC data is also shown (black line). The distribution comparisons are substantially similar for all unique patients in MIDRC to the United States general population than all unique COVID-19 positive patients to the case count distributions from the Centers for Disease Control and Prevention (CDC) due to the high percentage of cases within the CDC data for which ethnicity is not available (over 40%). MIDRC, The Medical Imaging and Data Resource Center; US, United States; CDC, Centers for Disease Control and Prevention; and C19+, cases positive for COVID-19.

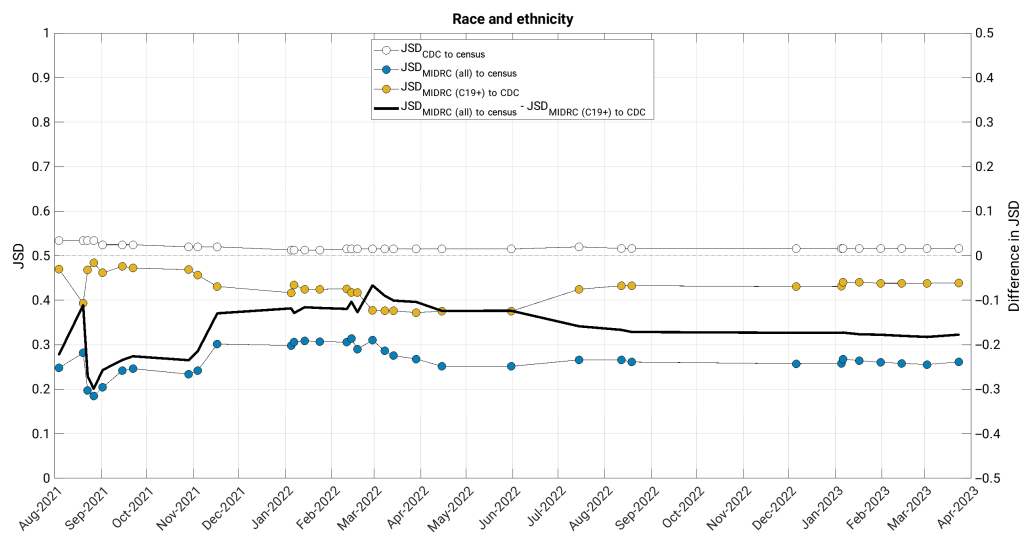


Fig. 8 The Jensen-Shannon distance (JSD) over time for race and ethnicity for (blue data markers) all unique patients, (gold data markers) all unique COVID-19 positive patients in the MIDRC data, and (white data markers) the JSD for comparing the CDC data to the US general population (for reference). The difference in JSD over time between all unique patients and all unique COVID-19 positive patients in the MIDRC data is also shown (black line). The lower representativeness (higher JSD) of all unique COVID-19 positive patients in the MIDRC data to the CDC data and the CDC data to the US general population is impacted by the substantial percentage of cases within the data from the Centers for Disease Control and Prevention for which race and ethnicity is not reported (over 40%). MIDRC, The Medical Imaging and Data Resource Center; US, United States; CDC, Centers for Disease Control and Prevention; and C19+, cases positive for COVID-19.

We believe the Jensen-Shannon distance to be useful in AI/ML investigations in medical imaging, in part due to its intuitive nature (especially in terms of its bounds) and its relationship to information theory, which is an important foundation to other AI/ML performance measures such as receiver operating characteristic analysis.⁵² Jensen-Shannon measures have also been used in some biology studies.⁵³ There are other methods for comparing the distributions of populations, such as the Hellinger distance,⁵⁴ population matching discrepancy,⁵⁵ and matching quantiles estimation.⁵⁶ The ratio of patient identity to disease prevalence, termed the participant to prevalence ratio (PPR),⁵⁷ is used in some clinical trials (in which subjects are termed “participants”) to measure representation within demographic subgroups. A measure such as the JSD is desirable for our definition of representativeness due to its ability to summarize across a demographic category, but measures such as the PPR would be complementary for analysis of individual subgroups. It would be advantageous for future studies to quantify the impact of ranges of PPR on adequate representation or lack thereof and to establish more specific criteria for such levels. On the whole, it will be useful to conduct a comprehensive comparison of different measures of representativeness (both across an entire demographic category and by subgroups) and their relationship to fairness of AI in medical imaging; this will be the topic of future work.

The work described here uses COVID-19 positive case counts collected by the CDC before the declared end of the COVID-19 public health emergency, on May 11, 2023. After this date, COVID-19 data reporting by the CDC will change, impacting the reporting of case counts.⁵⁸ We will continue to monitor the representativeness of the MIDRC open data commons as batch ingestion continues, using hospitalization rates⁵⁹ as a comparison group for COVID-19 positive cases.

There were some limitations to this study. First, the demographic categories described here were limited to one combination of demographic categories (race and ethnicity). Other combinations of demographic categories are important to consider (such as that of age and race or sex and race) and will be the topic of future study. Second, this study did not include other factors which may be relevant in studying health inequities, such as patient residence (e.g., urban versus rural), healthcare institution type (e.g., community versus academic), patient education level attainment, patient income or experienced income equities, and patient employment status. Third, there were some limitations in the data reported by the CDC: (1) it includes both probable and lab-confirmed COVID-19 cases, while COVID-19 positive cases in the MIDRC data commons include lab confirmation; (2) it includes non-unique case counts (i.e., the counts include some individuals who have tested positive for COVID-19 at different times) while the MIDRC data counts patients only once; and (3) it includes substantial proportions of data for which race and ethnicity are not reported. We are currently conducting related studies on the impact to AI/ML algorithm development and performance evaluation when representativeness is impacted by sizable proportions of missing data. While the purpose of this work is to report on the representativeness of the MIDRC open data commons *prima facie*, we also note the limitations of measures when using CDC data, including its missing data. In the future, we will investigate using methods the CDC is currently proposing to address missing data, such as assuming individuals with no reported ethnicity are non-Hispanic.⁶⁰ Finally, MIDRC works with data contributors to receive imaging study donations that have been de-identified using the Safe Harbor method, in compliance with the Health Insurance Portability and Accountability Act of 1996 privacy rule. Thus, while each patient’s timeline is preserved, the MIDRC data commons does not provide the actual date of image acquisition and COVID test. This means that the imaging studies within a given ingestion date can include imaging studies acquired theoretically at any time before the ingestion date, necessitating our comparison of cumulative case counts in both the MIDRC data and the COVID-19 case counts from the CDC, rather than a potential comparison for cases imaged within a given month to cases reported by the CDC within a given month.

In summary, the demographic characteristics of the MIDRC data in the categories of age at imaging, sex, race, ethnicity, and the combination of race and ethnicity and their similarity to comparison groups were measured using the Jensen-Shannon distance. Overall, the JSD indicated more representativeness for all unique patients than for COVID-19 positive patients when compared to their respective comparison groups. These measures can be used by investigators in

developing unbiased and generalizable AI/ML algorithms using the MIDRC data, including when building cohorts.

Disclosures

Dr. Kalpathy-Cramer has no funding to report for this article but funding for other work unrelated to what is presented here includes a research grant from GE, research support from Genentech, consultant/stock options from Siloam Vision LLC, and technology licensed to Boston AI. Dr. Koyejo has received funding support from the US National Science Foundation (NSF) (Grant Nos. NSF IIS 2205329 and NSF IIS 2046795). Dr. Myers works as an independent technical and regulatory consultant as principal for Puente Solutions LLC. Dr. Wawira-Gichoya has received funding support from the NSF Division of Electrical, Communication and Cyber Systems (Grant No. 1928481).

Data and Materials Availability Statement

The data presented in this article are publicly available from Ref. 20.

Acknowledgments

The authors are grateful to the staff at Gen3, the staff of ACR, RSNA, and AAPM, and the data contributors. The MIDRC is funded by the National Institute of Biomedical Imaging and Bioengineering (NIBIB), part of the National Institutes of Health (Grant Nos. 75N92020C00008 and 75N92020C00021).

References

1. J. A. W. Gold, "Race, ethnicity, and age trends in persons who died from COVID-19—United States, May–August 2020," *Morb. Mortal Wkly. Rep.* **69**, 1517–1521 (2020).
2. J. R. Goldstein and R. D. Lee, "Demographic perspectives on the mortality of COVID-19 and other epidemics," *Proc. Natl. Acad. Sci.* **117**(36), 22035–22041 (2020).
3. E. J. Marquez et al., "The lethal sex gap: COVID-19," *Immun. Ageing* **17**(1), 13 (2020).
4. E. G. Price-Haywood et al., "Hospitalization and mortality among black patients and white patients with Covid-19," *N. Engl. J. Med.* **382**(26), 2534–2543 (2020).
5. S. L. Harrison et al., "Comorbidities associated with mortality in 31,461 adults with COVID-19 in the United States: a federated electronic medical record analysis," *PLoS Med.* **17**(9), e1003321 (2020).
6. N. T. Nguyen et al., "Male gender is a predictor of higher mortality in hospitalized adults with COVID-19," *PLoS ONE* **16**(7), e0254066 (2021).
7. T. Bhowmik et al., "A comprehensive analysis of COVID-19 transmission and mortality rates at the county level in the United States considering socio-demographics, health indicators, mobility trends and health care infrastructure attributes," *PLoS ONE* **16**(4), e0249133 (2021).
8. A. M. Navar et al., "The impact of race and ethnicity on outcomes in 19,584 adults hospitalized with COVID-19," *PLoS ONE* **16**(7), e0254809 (2021).
9. M. T. Pagano et al., "Predicting respiratory failure in patients infected by SARS-CoV-2 by admission sex-specific biomarkers," *Biol. Sex Differ.* **12**(1), 63 (2021).
10. A. Tejpal et al., "Sex-based differences in COVID-19 outcomes," *J. Womens Health* **30**(4), 492–501 (2021).
11. D. Quan et al., "Impact of race and socioeconomic status on outcomes in patients hospitalized with COVID-19," *J. Gen. Intern. Med.* **36**(5), 1302–1309 (2021).
12. L. H. Unruh et al., "Health disparities and COVID-19: A retrospective study examining individual and community factors causing disproportionate COVID-19 outcomes in Cook County, Illinois," *PLoS ONE* **17**(5), e0268317 (2022).
13. L. Hol et al., "The effect of age on ventilation management and clinical outcomes in critically ill COVID-19 patients—insights from the PRoVENT-COVID study," *Aging-US* **14**(3), 1087–1109 (2022).
14. C. G. Arnold et al., "Immune mechanisms associated with sex-based differences in severe COVID-19 clinical outcomes," *Biol. Sex Differ.* **13**(1), 7 (2022).
15. L. Hill and S. Artiga, "COVID-19 cases and deaths by race/ethnicity: current data and changes over time," in KFF (2022).
16. A. C. Danielsen et al., "Sex disparities in COVID-19 outcomes in the United States: quantifying and contextualizing variation," *Social Sci. Med.* **294**, 114716 (2022).
17. J. J. Naidich et al., "Imaging utilization during the COVID-19 pandemic highlights socioeconomic health disparities," *J. Am. Coll. Radiol.* **18**(4), 554–565 (2021).

18. R. Lacson et al., “Exacerbation of inequities in use of diagnostic radiology during the early stages of reopening after COVID-19,” *J. Am. Coll. Radiol.* **18**(5), 696–703 (2021).
19. D. Y. Jarrett et al., “Advanced imaging of disease unrelated to the coronavirus disease 2019 (COVID-19) during the pandemic: effect of patient demographics in a pediatric emergency department,” *Pediatr. Radiol.* **52**(9), 1756–1764 (2022).
20. “The Medical Imaging and Data Resource Center,” MIDRC, <https://www.midrc.org> (accessed 1 June 2023).
21. The Medical Imaging and Data Resource Center, “MIDRC data use agreement,” MIDRC, <https://www.midrc.org/midrc-data-use-agreement> (accessed 1 June 2023).
22. The Medical Imaging and Data Resource Center, “The Medical Imaging and Data Resource Center open data commons,” MIDRC, <https://data.midrc.org/> (accessed 30 December 2022).
23. The Medical Imaging and Data Resource Center, “GitHub - MIDRC/midrc_dictionary: a data dictionary for the Medical Imaging Resource Data Commons (MIDRC),” https://github.com/MIDRC/midrc_dictionary (accessed 1 January 2023).
24. N. Baughan et al., “Sequestration of imaging studies in MIDRC: a multi-institutional data commons,” *Proc. SPIE* **12035**, 91–98 (2022).
25. Office of Management and Budget, *OMB Directive 15: Race and Ethnic Standards for Federal Statistics and Administrative Reporting* (1977).
26. US Census Bureau, “Census.gov,” <https://www.census.gov/en.html> (accessed 30 December 2022).
27. Centers for Disease Control and Prevention, “COVID-19 case surveillance public use data,” <https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data/vbim-akqf> (accessed 30 December 2022).
28. D. Endres and J. Schindelin, “A new metric for probability distributions,” *IEEE Trans. Inf. Theory* **49**(7), 1858–1860 (2003).
29. J. E. Dalton, W. A. Benish, and N. I. Krieger, “An information-theoretic measure for balance assessment in comparative clinical studies,” *Entropy* **22**(2), 218 (2020).
30. J. Corander, U. Remes, and T. Koski, “On the Jensen-Shannon divergence and the variation distance for categorical probability distributions,” *Kybernetika* **57**(6), 879–907 (2021).
31. J. Lin, “Divergence measures based on the Shannon entropy,” *IEEE Trans. Inf. Theory* **37**(1), 145–151 (1991).
32. National Institutes of Health, Institute on Minority Health and Health Disparities, *Strategic Plan 2021–2025*, <https://nimhd.nih.gov/about/strategic-plan/index.html> (accessed 1 June 2023).
33. United States Food and Drug Administration, *Collection, Analysis, and Availability of Demographic Subgroup Data for FDA-Approved Medical Products*, United States Department of Health and Human Services (2013).
34. Center for Biologics Evaluation and Research, Center for Devices and Radiological Health, and United States Food and Drug Administration, “*Evaluation of Sex-Specific Data in Medical Device Clinical Studies: Guidance for Industry and Food and Drug Administration Staff*,” United States Department of Health and Human Services (2014).
35. Center for Biologics Evaluation and Research, Center for Devices and Radiological Health, and United States Food and Drug Administration, “*Evaluation and Reporting of Age-, Race-, and Ethnicity-Specific Data in Medical Device Clinical Studies: Guidance for Industry and Food and Drug Administration Staff*,” United States Department of Health and Human Services (2017).
36. G. S. Collins et al., “Protocol for development of a reporting guideline (TRIPOD-AI) and risk of bias tool (PROBAST-AI) for diagnostic and prognostic prediction model studies based on artificial intelligence,” *BMJ Open* **11**(7), e048008 (2021).
37. E. Colak, R. Moreland, and M. Ghassemi, “Five principles for the intelligent use of AI in medical imaging,” *Intensive Care Med.* **47**(2), 154–156 (2021).
38. P. Rouzrokh et al., “Mitigating bias in radiology machine learning: 1. Data handling,” *Radiol. Artif. Intell.* **4**(5), e210290 (2022).
39. K. Zhang et al., “Mitigating bias in radiology machine learning: 2. Model development,” *Radiol. Artif. Intell.* **4**(5), e220010 (2022).
40. S. Faghani et al., “Mitigating bias in radiology machine learning: 3. Performance metrics,” *Radiol. Artif. Intell.* (2022).
41. H. Estiri et al., “An objective framework for evaluating unrecognized bias in medical AI models predicting COVID-19 outcomes,” *J. Am. Med. Inf. Assoc.* **29**(8), 1334–1341 (2022).
42. M. A. Ricci Lara, R. Echeveste, and E. Ferrante, “Addressing fairness in artificial intelligence for medical imaging,” *Nat. Commun.* **13**, 4581 (2022).
43. K. Drukker et al., “Toward fairness in artificial intelligence for medical image analysis: identification and mitigation of potential biases in the roadmap from data collection to model deployment,” *J. Med. Imaging* **10**(6), 061104 (2023).

44. A. Rajkomar et al., “Ensuring fairness in machine learning to advance health equity,” *Ann. Intern. Med.* **169**(12), 866–872 (2018).
45. L. H. Clemmensen and R. D. Kjærsgaard, “Data representativity for machine learning and AI systems,” arXiv:2203.04706 (2022).
46. V. Azimi and M. A. Zaydman, “Optimizing equity: working towards fair machine learning algorithms in laboratory medicine,” *J. Appl. Lab. Med.* **8**(1), 113–128 (2023).
47. N. V. Chawla et al., “SMOTE: synthetic minority over-sampling technique,” *J. Artif. Intell. Res.* **16**, 321–357 (2002).
48. D. Dablain, B. Krawczyk, and N. V. Chawla, “DeepSMOTE: fusing deep learning and SMOTE for imbalanced data,” *IEEE Trans. Neural Netw. Learn. Syst.* (2022).
49. K. C. Santosh, “AI-driven tools for coronavirus outbreak: need of active learning and cross-population train/test models on multitudinal/multimodal data,” *J. Med. Syst.* **44**(5), 93 (2020).
50. S. A. Harmon et al., “Artificial intelligence for the detection of COVID-19 pneumonia on chest CT using multinational datasets,” *Nat. Commun.* **11**(1), 4080 (2020).
51. J. M. Johnson and T. M. Khoshgoftaar, “Survey on deep learning with class imbalance,” *J. Big Data* **6**(1), 27 (2019).
52. C. E. Metz, “Basic principles of ROC analysis,” *Sem. Nucl. Med.* **8**(4), 283–298 (1978).
53. N. Ramakrishnan and R. Bose, “Analysis of healthy and tumour DNA methylation distributions in kidney-renal-clear-cell-carcinoma using Kullback–Leibler and Jensen–Shannon distance measures,” *IET Syst. Biol.* **11**(3), 99–104 (2017).
54. E. Hellinger, “Neue Begründung der Theorie quadratischer Formen von unendlichvielen Veränderlichen,” *J. für die reine und angew. Mathematik* **136**, 210–271 (1909).
55. J. Chen et al., “Population matching discrepancy and applications in deep learning,” in *Adv. Neural Inf. Process. Syst.*, Vol. 30, Curran Associates, Inc. (2017).
56. N. Sgouropoulos, Q. Yao, and C. Yastremiz, “Matching a distribution by matching quantiles estimation,” *J. Am. Stat. Assoc.* **110**(510), 742–759 (2015).
57. R. Poon et al., “Participation of women and sex analyses in late-phase clinical trials of new molecular entity drugs and biologics approved by the FDA in 2007-2009,” *J. Womens Health-Larchmt* **22**(7), 604–616 (2013).
58. B. J. Silk, “COVID-19 surveillance after expiration of the public health emergency declaration—United States, May 11, 2023,” *MMWR Morb. Mortal. Wkly. Rep.* **72**(19), 523–528 (2023).
59. CDC, “COVID Data Tracker: COVID-NET Laboratory-confirmed COVID-19 hospitalizations,” Centers for Disease Control and Prevention, 28 March 2020, <https://covid.cdc.gov/covid-data-tracker/#covidnet-hospitalization-network> (accessed 6 June 2023).
60. P. Yoon, “Alternative methods for grouping race and ethnicity to monitor COVID-19 outcomes and vaccination coverage,” *MMWR Morb. Mortal. Wkly. Rep.* **70**(32), 1075–1080 (2021).

Heather M. Whitney, PhD, is a research assistant professor of radiology at the University of Chicago. She received her PhD in physics from Vanderbilt University, conducting research at the Vanderbilt University institute of Imaging Science. She is interested in investigating the effects of the physical basis of imaging on radiomics, the repeatability and robustness of radiomics, the development of methods for task-based distribution, and bias and diversity of medical imaging datasets.

Natalie Baughan, PhD, is a recent graduate in the University of Chicago Graduate Program in medical physics. After receiving her BS degree in nuclear engineering and radiological sciences from the University of Michigan in 2019, her research has focused on breast cancer risk assessment in mammography and statistical methods for AI. She is continuing to residency in medical physics at University of North Carolina–Chapel Hill.

Kyle J. Myers, PhD, served as a research scientist and manager in the FDA’s Center for Devices and Radiological Health for over 30 years. She coauthored *Foundations of Image Science*, winner of the First Biennial J.W. Goodman Book Writing Award from OSA and SPIE. She is a fellow of AIMBE, Optica, SPIE, and a member of the National Academy of Engineering. She received her PhD in optical sciences from the University of Arizona in 1985.

Karen Drukker, PhD, is a research associate professor of radiology at the University of Chicago, where she has been involved in medical imaging research for 20+ years. She received her PhD in physics from the University of Amsterdam. Her research interests include machine learning applications in the detection, diagnosis, and prognosis of disease, focusing on rigorous

training/testing protocols, generalizability, performance evaluation, and bias and fairness of AI. She is a fellow of SPIE.

Judy Gichoya, MD, MS, is an assistant professor at Emory University in interventional radiology and informatics. Her career focus is on validating machine learning models for health in real clinical settings, exploring explainability, fairness, and a specific focus on how algorithms fail. She is heavily invested in training the next generation of data scientists through multiple high school programs, serving as the program director for the *Radiology: Artificial Intelligence* trainee editorial board and the medical student machine-learning elective.

Brad Bower, PhD, is a Data and Technology Advancement National Service Scholar with the NIH Office of Data Science Strategy (ODSS) and the National Institute of Biomedical Imaging and Bioengineering. He has a background in developing and commercializing imaging and data-driven devices in healthcare. At NIBIB he is leading efforts for improving FAIR data use and bench-to-bedside AI healthcare product development.

Weijie Chen, PhD, is a research physicist in the Division of Imaging, Diagnostics, and Software Reliability, Office of Science and Engineering Laboratories, CDRH, US FDA, where he conducts research and provides consults to regulatory review of medical devices. He earned his PhD in Medical Physics in 2007 from the University of Chicago. His research interests include performance characterization and assessment methodologies for medical imaging and AI/ML/CAD devices.

Nicholas Gruszauskas, PhD, is currently the technical director of the University of Chicago's Human Imaging Research Office and a faculty member in the Department of Radiology's Clinical Imaging Medical Physics Residency Program. He earned both an MS and PhD in biomedical engineering from the University of Illinois Chicago, where he investigated computer-aided diagnosis and artificial intelligence methods for breast imaging. His team is currently responsible for facilitating clinical trial medical imaging at the university.

Jayashree Kalpathy-Cramer, PhD, is the chief of the Division of Artificial Medical Intelligence in the Department of Ophthalmology at the University of Colorado. Previously, she was an associate professor of radiology at Harvard Medical School where she was actively involved in data science activities with a focus on medical imaging. Her research spans the spectrum from novel algorithm development to clinical deployment. She has authored over 200 peer-reviewed publications and has written over a dozen book chapters.

Sanmi Koyejo, PhD, is an assistant professor in the Department of Computer Science at Stanford University. Koyejo's research interests are in developing the principles and practice of trustworthy machine learning, including fairness and robustness. Additionally, Koyejo focuses on applications to neuroscience and healthcare. Koyejo has received several awards, including a best paper award from the Conference on Uncertainty in Artificial Intelligence, a Skip Ellis Early Career Award, and a Sloan Fellowship.

Rui C. Sá, PhD, is an assistant professor of physiology at the University of California, San Diego. His research focuses on functional imaging of the human lung in health and disease. As a Data and Technology Advancement National Service Scholar at NIBIB/NIH, he supported the first two years of MIDRC and other NIH medical imaging data-centric initiatives.

Berkman Sahiner, PhD, is a senior biomedical research scientist with the Division of Imaging, Diagnostics and Software Reliability, Center for Devices and Radiological Health, US FDA. He has a PhD in electrical engineering and computer science from the University of Michigan, Ann Arbor. His research is focused on the evaluation of medical imaging and computer-assisted diagnosis devices, including devices that incorporate machine learning and artificial intelligence. He is a fellow of SPIE and AIMBE.

Zi Zhang, MD, is an assistant professor of radiology at the Sidney Kimmel Medical College Thomas Jefferson University. She completed her imaging informatics fellowship and breast imaging fellowship at the University of Pennsylvania. She is a member of the Bias and Diversity Working Group at the Medical Imaging and Data Resource Center. She also serves as a

committee member of the RSNA Asia/Oceania International Advisory Committee and the Society of Breast Imaging Social Media Committee.

Maryellen L. Giger, PhD, is the A.N. Pritzker distinguished service professor at the University of Chicago. Her research involves computer-aided diagnosis/machine learning in medical imaging for cancer and now COVID-19, and is contact PI on the NIBIB-funded Medical Imaging and Data Resource Center (midrc.org), which has published more than 100,000 medical imaging studies for use by AI investigators. She is a member of the National Academy of Engineering, a recipient of the AAPM Coolidge Gold Medal, SPIE Harrison H. Barrett Award, and RSNA Outstanding Researcher Award, and is a fellow of AAPM, AIMBE, SPIE, and IEEE.