JACC: ADVANCES VOL. 4, NO. 3, 2025

ORIGINAL RESEARCH

HEART FAILURE AND CARDIOMYOPATHIES

Diversity and Inclusion Within Datasets in Heart Failure

A Systematic Review

Elinor Laws, BSc, MBBCh,[a,b] Maria Charalambides, BSc, MBChB, MRCP,[c] Sonam Vadera, BS, MBBS,[d] Eva Keller, MD, MSc,[e] Joseph Alderman, MBChB,[a,b] Breanna Blackboro, BSc, MBChB,[b] Jeffry Hogg, MBBS, MSc, PhD,[b] Thomas Salisbury, BSc, MBChB,[f] Joanne Palmer, PhD,[a] Melanie Calvert, PhD,[b,g,h] Maxine Mackintosh, PhD,[i] Rubeta Matin, BSc, MBChB, PhD, MRCP,[j] Elizabeth Sapey, BSc, MBBS, PhD,[a,k,l] Johan Ordish, BA, MA,[b] Melissa McCradden, PhD,[m] Bilal Mateen, BSc, MBBS, MPH,[e] Jacqui Gath,[n] Adewale Adebajo, PhD,[o] Stephanie Kuku, MBChB, MD,[e] William Bradlow, MBBS,[a,b] Alastair K. Denniston, PhD, MRCP,[a,b,g,h] Xiaoxuan Liu, PhD, MBChB[a,b]

ABSTRACT

BACKGROUND Heart failure (HF) is a life-threatening disease affecting 64 million people worldwide. Artificial intelligence (AI) technologies are being developed for use in HF to support early diagnosis and stratification of treatment. The performance characteristics of AI technologies are influenced by whether the data used during the AI lifecycle reflects the populations for which the AI is used.

OBJECTIVES The aim of the study was to identify and characterize datasets used across the lifecycle of AI technologies for HF, focusing on data diversity and inclusivity.

METHODS MEDLINE and Embase were systematically searched from January 1, 2012, until August 30, 2022, to identify articles relating to the development of AI in HF. Articles were independently screened by 2 reviewers to identify datasets. Dataset documentation was analyzed with a focus on accessibility, geographical origin, relevant metadata reporting, and dataset composition.

RESULTS The 72 datasets identified represented 23 countries and over 2 million individuals. In total, 62 (86%) datasets reported "age," 61 (85%) reported sex or gender, 21 (29%) reported race and/or ethnicity, and 8 (11%) reported socioeconomic status. In the 21 datasets that reported race and/or ethnicity, 89% of individuals represented were reported within the "White" or "Caucasian" category. Only 20 (28%) datasets were fully accessible.

CONCLUSIONS Reporting of sex, gender, and socioeconomic status in HF datasets is inconsistent. There is a need to generate datasets that are transparently reported and accessible. Although collecting and reporting demographic attributes is complex and needs to be undertaken with appropriate safeguards, it is also an essential step toward building equitable AI-based health technologies. (JACC Adv. 2025;4:101610) © 2025 The Authors. Published by Elsevier on behalf of the American College of Cardiology Foundation. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

From the [a]University Hospitals Birmingham NHS Foundation Trust, Birmingham, United Kingdom; [b]Institute of Applied Health Science, University of Birmingham, Birmingham, United Kingdom; [c]University of Southampton, Southampton, United Kingdom; [d]University Hospitals Leicester, Leicester, United Kingdom; [e]University College London, London, United Kingdom; [f]South Tyneside and Sunderland NHS Foundation Trust, Sunderland, United Kingdom; [g]NIHR Birmingham Biomedical Research Centre, Birmingham, United Kingdom; [h]Birmingham Health Partners Centre for Regulatory Science and Innovation, University of Birmingham, Birmingham, United Kingdom; [i]Alan Turning Institute, London, United Kingdom; [j]University of Oxford Hospitals NHS Foundation Trust, Ocford, United Kingdom; [k]NIHR Birmingham Patient Safety Research Collaborative, Birmingham, United Kingdom; [l]Institute of Inflammation and Ageing, University of Birmingham, Birmingham, United Kingdom; [m]Bioethics Department, The Hospital for Sick Children, Toronto, Ontario, Canada; [n]Independent Cancer Patients' Voice, United Kingdom; and the [o]University of Sheffield, Sheffield, United Kingdom.

ISSN 2772-963X https://doi.org/10.1016/j.jacadv.2025.101610

ABBREVIATIONS AND ACRONYMS

AI = artificial intelligence

HF = heart failure

Over 64 million people live with heart failure (HF) worldwide. This syndrome has a wide range of etiologies, and diagnosis and treatment often rely on imaging and complex physiological assessments.[1] Prevalence of HF is increasing, but there are significant geographic disparities in access to specialist services for its management.[2] Artificial intelligence (AI) health technologies are increasingly seen as a means to improve access to specialist-level HF care, improve diagnostic accuracy, and reduce inequitable outcomes for diverse populations.[3,4] As of January 2025, there are 104 tools targeting cardiovascular diseases authorized by the U.S. Food and Drug Agency.[5]

Despite their potential benefits, there is a growing literature highlighting the risk associated with bias caused by AI health technologies.[6-8] While algorithmic bias can be multifaceted, a key driver is the misalignment between populations represented in development datasets and those for whom a model is intended to be used.[6] In the Good Machine Learning Practice for Medical Device Development published by the Food and Drug Agency alongside Health Canada and the UK's Medicines and Health products Regulatory Agency, 1 of the 10 guiding principles is that "Clinical Study Participants and Data Sets Are Representative of the Intended Patient Population."[7] However, datasets collected during a clinical trial inherit the trial's eligibility criteria and any barriers to participation as selection biases, limiting their generalizability to a wider health care context.

Data generated during delivery of both clinical trials and routine clinical care (including electrocardiography, echocardiography, and other medical investigations) are increasingly available for research and development, including the creation of AI health technologies. However, communities that experience health inequity are frequently under-represented, even within routinely collected health data, and so are at risk of being further disadvantaged by under-performance of AI health technologies.[8] Recent reviews of datasets relating to ophthalmology, dermatology, mammography, COVID-19, and clinical free-text have demonstrated that under-representation is common, and that demographic attributes are often not reported even where these attributes are clinically relevant.[9-13] Demographic attributes (including race, ethnicity, sex, gender, socioeconomic status, and others) are associated with disparate health outcomes in the context of cardiovascular medicine. People experiencing socioeconomic deprivation are at increased risk of developing HF, and people from minoritized racial and/or ethnic groups are more likely to experience worse HF outcomes.[14,15] It is therefore important that AI health technologies for HF are developed using datasets that are representative of these populations to reduce the risk of exacerbating pre-existing health inequity.

This review aimed to investigate the datasets available for development of AI health technologies targeting HF, particularly those generated from investigations used early in the diagnostic pathway.[16,17] We assess the documentation and composition of HF datasets, in particular focusing on demographic attributes to characterize "who" is represented, "how" they are represented, and any groups left behind.

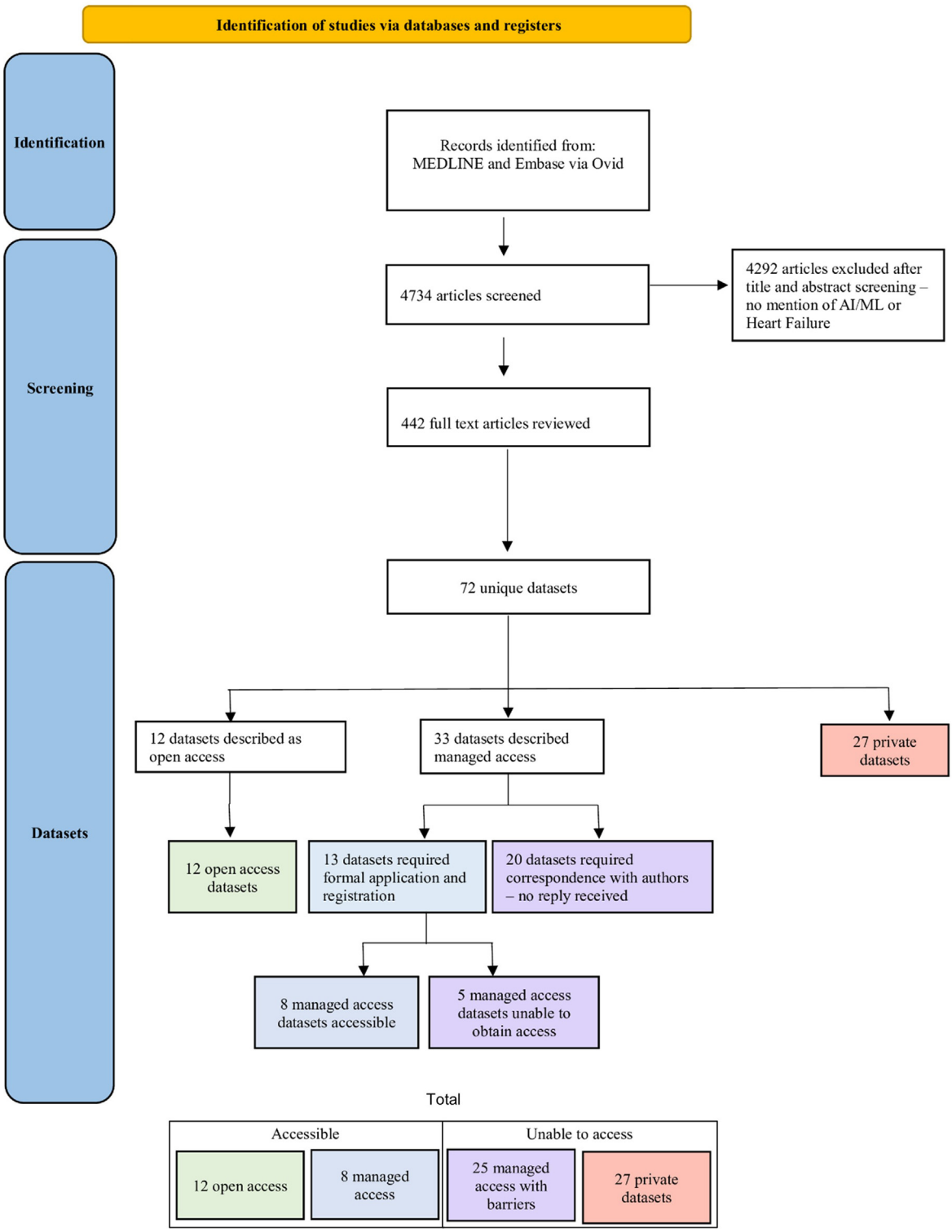
METHODS

The purpose of this systematic review was to access the datasets used during the development of AI health technologies related to HF, rather than to identify the publications from which they were identified. Accordingly, we adapted methodology used in previous reviews to conduct article screening and data extraction at 2 levels: initially to identify relevant publications and then to identify relevant datasets cited within these publications.[9-11] Searches were conducted on 30 August, 2022 using MEDLINE and Embase to identify articles that referred to datasets used during development of AI health technologies in HF. A pilot search including all investigations relating to HF was conducted and yielded an unfeasible number of papers. The search was therefore focused on investigations used early in the diagnostic pathway, including echocardiography and electrocardiography, but excluding those investigations used predominantly in secondary care (such as cardiac magnetic resonance imaging). The full search strategy is provided in Supplemental Appendix 1. Only texts written in the English language since January 1, 2012, relating to HF were included. See **Figure 1** for the PRISMA flowchart, showing the

The authors attest they are in compliance with human studies committees and animal welfare regulations of the authors' institutions and Food and Drug Administration guidelines, including patient consent where appropriate. For more information, visit the Author Center.

Manuscript received October 15, 2024; revised manuscript received January 17, 2025, accepted January 19, 2025.

FIGURE 1 PRISMA Diagram Outlining the Process of Dataset Identification



Literature search completed via MEDLINE and Embase identified 4,734 articles. After title/abstract screening, 442 full texts were reviewed to identify referenced datasets. The 72 identified datasets were classified as 1) open access (n = 12); 2) managed access (defined as datasets that required formal application and/or payment for access n = 8); 3) managed access with barriers (defined as datasets that were reported as managed access but had practical barriers to access n = 25); and 4) private (defined as datasets that were not intended to be accessible n = 27). AI = artificial intelligence.

process of screening articles and datasets. The review protocol did not qualify for PROSPERO as it is not directly related to health outcomes.

Abstracts were independently screened in duplicate by E.L. and B.B. using the Rayyan web platform (http://rayyan.qcri.org) with discrepancies resolved by X.L.[18] Full texts were then reviewed by a single author (E.L., E.K., T.S., or X.L.) to identify datasets for further analysis. Review articles were excluded, as were datasets that were entirely composed of another included dataset.

Identified datasets were reviewed, and their access requirements were initially classified into 1) open access; 2) managed access (defined as datasets that required formal application with or without payment for access); 3) managed access with barriers (defined as datasets that were reported as managed access but had practical barriers to access, for example, where no hyperlinks were provided or no response from data curators was received); and 4) private (defined as datasets that were not intended to be accessible). Where possible, we sought approvals from managed access datasets, but if no response was received within 2 months of initial request, datasets were classified as "managed access with barriers."

Documentation accompanying accessible datasets was analyzed, and information was extracted from these sources using a standardized data extraction form (Supplemental Appendix 2), informed by the STANDING Together recommendations.[16] In brief, this included data accessibility, version history, content (population, data collection setting, modifications), and metadata (labels, type of clinical data). For datasets with documentation hosted in multiple locations (for instance, a readme file, a dataset description document, written information on a website, and an associated publication), we limited our review to 4 successive sources of documentation. Data extraction was completed by a single author (E.L., S.V., M.Ch., E.K., or J.H.). Where necessary, we contacted data custodians via email or teleconference to discuss data availability, documentation, and sharing. Where datasets were inaccessible, we attempted to extract the relevant information from any linked research papers that described the dataset.

Reviewers adopted agreed definitions for race, ethnicity, sex, and gender during the conduct of this study, which are as follows: Race is a social construct used as a proxy for human variation, often based on skin tone and other physical characteristics.[19] Ethnicity broadly refers to a person's identity in relation to their origin and background.[20] Sex is defined as a label assigned at birth (either intersex, female, or male) based on biological characteristics (eg, reproductive anatomy, genetics, and physiology). Gender refers to a person's psychological, emotional, social, and overall sense of identity, which classifies them as woman, man, or neither.[21]

PATIENT AND PUBLIC INVOLVEMENT AND ENGAGEMENT. Patient and public involvement and engagement members were consulted during funding application, problem definition, and manuscript review. Quarterly meetings were held to allow for contributions from the patient and public involvement and engagement members on how the project should be conducted and how concepts should be discussed, including the choice of demographic attributes within the analysis.

RESULTS

DATASETS IDENTIFIED FROM THE LITERATURE SEARCH. The full dataset identification process is outlined in a PRISMA diagram (**Figure 1**). In total, 4,734 articles were identified by the MEDLINE/Embase literature search. After abstract screening, 442 articles were included for full-text review. There were 72 unique datasets identified within the literature that fulfilled the inclusion criteria and underwent complete review.

DATASET ACCESSIBILITY. Of the 72 datasets identified, 20 were fully accessible to us during this review—their details are provided in **Table 1**. Of these 20 accessible datasets, 12 had data that could be freely reviewed, and so were classified as "open access." The remaining 8 datasets required data access requests to be made accompanied by justification for access, and so were classified as "managed access." We found a further 52 datasets that were unable to fully access: 20 of these were reported to be accessible in the scientific paper describing them; however, we did not receive a response to our attempted correspondence with the papers' authors. Five datasets were reported to be accessible after registering an account via an online platform; however, we were unable to access any of these. The final 27 datasets were private and not intended to be accessible by others. Characteristics of the 52 datasets that were unable to fully access are provided in Supplemental Appendix 3. Some of these datasets provided details of their composition in accompanying documentation, but we were unable to confirm this via inspection of individual-level data as these were inaccessible to us.

DATASETS CHARACTERISTICS. Across all 72 datasets, there were over 2 million individuals represented, nearly 50% of whom were from the United States. Some datasets did not report number of

individuals represented and data were anonymized meaning it was not possible to determine whether there were individuals represented in more than 1 dataset. Sixty-three (88%) datasets provided the country of data collection—the majority of these were collated from a single source, but 5 datasets were derived from more than 1 country (**Central Illustration**). Overall, 59 of the 72 datasets (81.9%) were collected from clinical research studies; of these 33 (55.9%) were derived from prospective studies, 24 (40.7%) from retrospective studies, and 2 (3.3%) from both prospective and retrospective studies. The oldest dataset was published in 1989, and the newest datasets were added continuously rather than being updated at intervals. In total, 66 of 72 (91.7%) datasets were curated after 2015. Of the 72 datasets reviewed, 33 (45.8%) reported that consent was obtained from individuals, 23 (31.9%) reported that individuals' consent had been waived by ethics committees, and 16 (22.2%) did not report consent considerations.

The datasets comprised a variety of different data types, most commonly electrocardiogram data. The data types represented within the accessible datasets are outlined in **Table 1**, and the total number of datasets providing each data type is presented in **Central Illustration**.

Reporting of individuals' attributes. The number of datasets reporting different individuals' attributes, number of individuals, country of origin, consent, ground truth, missingness, and dataset modifications are presented in **Central Illustration**. Sixty-two of 72 datasets (86.1%) reported age, and 61 of 72 (84.7%) reported sex or gender. However, of the 61 datasets reporting sex or gender, only 3 (4.9%) provided a definition for those labels—all 3 specified "sex assigned at birth." While 61 datasets reported female/male categories, only 47 provided information about how many individuals were represented within those female/male categories. There were 1,170,793 records of individuals represented within the 47 datasets reporting numbers within each sex or gender category. Within those 47 datasets, there were 588,648 records relating to people categorized as female (50.2%), 582,115 records relating to people categorized as male (49.8%) with 1 dataset also reporting 21 records as "unknown." The number of individuals within the demographic categories did not always add up to the total number of individuals reported within the dataset (as was the case with female/male categories). Supplemental Appendix 4 demonstrates the number of individual records represented within each sex/gender category for each dataset.

Ethnicity, race, and other related characteristics were infrequently reported—7 datasets (10%) reported ethnicity, 11 (15%) reported race, 1 reported "race/ethnicity" (1%), and 1 reported both race and ethnicity (1%); an additional dataset (1%) reported categories similar to those within race or ethnicity (ie, "White" and "non-White") but were not labeled as such. Of the 21 (29%) datasets that reported race or ethnicity, no definitions for these terms were provided. These 21 datasets represent 922,445 individual records. While the race/ethnicity categories varied between datasets (Supplemental Appendix 5), all included a category of "White" or "Caucasian." In total, 89% (820,327 of 922,445) of the individual records represented within the 21 datasets that reported race/ethnicity breakdowns were classified within the "White" or "Caucasian" categories. While no explicit definitions were given for what the groupings represented, in some cases, information was given about where the data derived from, which gave some context to the data, for example, self-reported questionnaire as part of a prospective research study.

Seven datasets (10%) reported socioeconomic status. The methods for summarizing socioeconomic status also varied greatly between the 7 (10%) datasets that reported this information, from insurance status to bespoke questionnaires being completed by individuals relating to household income, education level, employment, and insurance status. An overview of all methods used to summarize socioeconomic status is provided in Supplemental Appendix 6.

DATASET DOCUMENTATION FORMAT. The design and layout of HF dataset documentation was variable. Dataset documentation was presented in peer-reviewed journal articles, downloadable files alongside the data, and information written within tabs on dataset host sites. Fourteen of the datasets were hosted within "repositories" like Physionet (https://physionet.org/) and BioLINCC (https://biolincc.nhlbi.nih.gov/home/) where they had more structured, consistent formats describing how dataset documentation was displayed between different datasets. One dataset was derived from a clinical research network called PCORnet, and data were aligned to a common data model.[22]

DISCUSSION

Within this systematic review, we assessed datasets being used to develop AI health technologies for HF, focusing on demographic attributes and representation. We assessed 72 datasets representing over 2 million individuals from 23 countries. The most common data types were electrocardiography, blood results (including brain natriuretic peptide), and data

TABLE 1 Characteristics of Accessible Datasets

Dataset Name	Accessibility	Clinical Focus	Number of Individuals	Date of Release	File Format (Comma Separated)	Country of Origin (Individual)
ARIC	Managed access	Cardiac	15,792	Ongoing	sas7bdat, xls, pdf	USA
BIDMC CHF	Open access	HF	15	2000	hea, electrocardiogram, dat, txt	USA
CAMUS	Open access	Cardiac	450	2022	_gt.nii.gz, cfg	France
ChestX-ray8	Open access	General	30,805	2017	pdf, txt, csv	USA
CHF RR Interval	Open access	HF	29	2003	Electrocardiogram, hea	Not reported
EchoNet-Dynamic	Managed access	Cardiac	10,030	2020	Xls	USA
eICU Collaborative Research Database 2.0	Managed access	General	139,000	2019	csv.gz	USA
Fantasia	Open access	General	40	2003	dat, hea, Electrocardiogram	Not reported
HF proteomics	Open access	HF	654	Not reported	txt, tar.gz	New Zealand, Singapore
Japanese database	Open access	HF	2,413	2020	Xls	Japan
MESA	Managed access	Cardiac	6,800	Ongoing	sas7bdat, xls, pdf, sas	USA
MIMIC-III Clinical Database	Managed access	General	59,098	2016	csv.gz	USA
MIMIC-IV	Managed access	General	299,712	Version: 2020, 2021, 2022, 2023	csv.gz	USA
MIT-BIH Arrhythmia	Open access	Cardiac	47	Digitized in 1989	dat, hea, xws, atr, txt	USA
MIT-BIH NSR	Open access	Cardiac	18	2016	hea, xws, atr, dat	USA
NSR RR Interval	Open access	Cardiac	54	2003	hea, Electrocardiogram	USA
Pulmonary edema severity grades based on MIMIC-CXR	Managed access	HF	1,916	2021	txt, csv	USA
SaMi-Trop Study	Open access	HF	1,304	2022	csv, rar	Brazil
STAFF III	Open access	Cardiac	104	2017	dat, hea, ods, event, png, xls, pdf	USA
UK Biobank	Managed access	General	500,000	Ongoing	pdf, txt, tar, bed, fam, bgem, bgi, sample, bin, batch	UK

BNP = brain natriuretic peptide; HF = heart failure.

Continued on the next page

relating to patients' comorbidities. Only 28% of datasets were accessible, and many were derived from clinical trials, posing a risk of selection bias. There were significant barriers to dataset access and incomplete reporting of relevant demographic metadata, with limited inclusion and diversity. Here, we explore these issues in the context of wider literature and explore their implications for the development of inclusive and equitable AI health technologies.

MOST DATASETS ARE DERIVED FROM CLINICAL TRIALS. A key finding was that most of the datasets were created from data collected during research studies rather than from routine clinical care. Rigid eligibility criteria and barriers to participating in

TABLE 1 Continued

Consent Model	Dataset Composition	Dataset Origin	Dataset URL
Obtained from individuals	Electrocardiogram, BNP, observations, eg, blood pressure, admission data, comorbidities, death certificates, out-of-hospital cardiac arrest, physical examination, blood results, family history	Prospective research study	https://aric.cscc.unc.edu/aric9/researchers/cohort_description
Not reported	Electrocardiogram, blood results, pharmacotherapy, NYHA functional class	Prospective research study	https://physionet.org/content/chfdb/1.0.0/
Not reported	Echocardiography	Routine clinical practice	https://humanheart-project.creatis.insa-lyon.fr/database/#collection/6373703d73e9f0047faa1bc8
Not reported	Chest X-rays	Routine clinical practice	https://nihcc.app.box.com/v/ChestXray-NIHCC
Not reported	Electrocardiogram, NYHA functional class	Not reported	https://archive.physionet.org/physiobank/database/chf2db/
Not reported	Echocardiography, electrocardiogram, image annotations, image reports	Routine clinical practice	https://echonet.github.io/dynamic/
Not reported	Observations, eg, blood pressure, clinical care noting, blood results	Routine clinical practice	https://physionet.org/content/eicu-crd/2.0/
Not reported	Electrocardiogram, observations, eg, blood pressure	Prospective research study	https://archive.physionet.org/physiobank/database/fantasia/
Obtained from individuals	Echocardiography, proteomics, comorbidities	Prospective research study	https://github.com/ArisStefanosSn/HFproteomics
Waived	Echocardiography, BNP, observations, eg, blood pressure, comorbidities, blood results, NYHA functional class	Routine clinical practice	https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0234294
Obtained from individuals	Electrocardiogram, observations, eg, blood pressure, genetics (whole genome sequencing), comorbidities (eg, hypertension, diabetes, etc), social factors, reproductive history, mental health history, CT chest, cardiac MRI, carotid USS/Doppler, arterial wave form, ABPI, blood results, demographic data, socioeconomic status, pharmacotherapy, psychosocial assessment, diet assessment, family history, anthropometry, electrocardiography, spot urine collection for microalbuminuria, endothelial function	Prospective research study	https://biolincc.nhlbi.nih.gov/studies/mesa/
Waived	Echocardiography, clinical care noting, blood results, comorbidities, admission data, pharmacotherapy	Routine clinical practice	https://physionet.org/content/mimiciii/1.4/
Waived	Observations, eg, blood pressure, blood results, admission data, pharmacotherapy, microbiology, clinical care noting	Routine clinical practice	https://physionet.org/content/mimiciv/2.2/
Not reported	Electrocardiogram, pharmacotherapy	Prospective research study	https://physionet.org/content/mitdb/1.0.0/
Not reported	Electrocardiogram	Routine clinical practice	https://archive.physionet.org/physiobank/database/nsrdb/
Not reported	Electrocardiogram	Routine clinical practice	https://archive.physionet.org/physiobank/database/nsr2db/
Waived	Chest X-rays, X-ray reports	Routine clinical practice	https://physionet.org/content/mimic-cxr-pe-severity/1.0.1/
Obtained from individuals	Electrocardiogram	Prospective research study	https://github.com/samitrop/AI-Electrocardiogram-Chagas
Obtained from individuals	Electrocardiogram, artery location of balloon inflation, occlusion and contrast injections, nuclear imaging data	Prospective research study	https://www.physionet.org/content/staffiii/1.0.0/
Obtained from individuals	Electrocardiogram, observations, eg, blood pressure, admission data, exercise testing data, genetics (whole genome and exome sequencing), comorbidities, blood results, physical activity, saliva assays	Prospective research study	https://www.ukbiobank.ac.uk/

health research make it less likely that these datasets are representative of the wider population seeking medical treatment. It is well documented that women, pregnant people, children, people with disabilities, and those experiencing poverty are frequently excluded from research.[23,24] The use of these data for the development of AI health technologies risks compounding selection bias and pre-existing inequity. AI health technologies developed on data with inherent selection bias may work less well for those who are under-represented.

Transparent documentation is needed. There are many available recommendations to guide transparent data practices, but many datasets in our review had limited dataset documentation.[16,25-27] Dataset documentation should describe the decisions made during the dataset curation process including data collection methods used (eg, imaging devices,

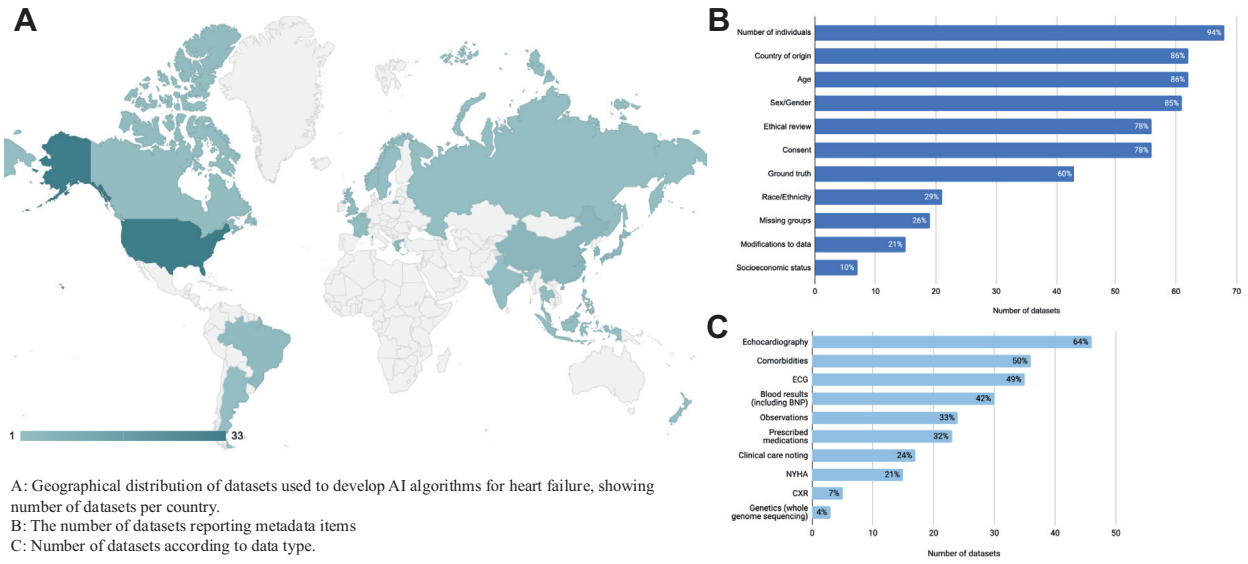
biochemistry reference ranges, processes for inputting data), ethical considerations (consent, ethical/internal review board approval), and among others. Transparent documentation is particularly relevant for concepts that are variable such as demographic attributes like race, ethnicity, and socioeconomic status—the definition and interpretation of which are inextricably linked with culture and economies. Without description and definition of these data, there is a risk of misinterpretation and misuse during development of associated AI health technologies. To ensure the usability and interoperability of health data, there needs to be transparent documentation defining and describing their context.[28,29]

LIMITED DATA AVAILABILITY FROM DIVERSE POPULATIONS. Most datasets used for developing AI health technologies for HF reported age and "female"/"male" categories, but less than one-third reported race or ethnicity, and even fewer reported socioeconomic status. Where reported, there was almost equal representation of "female" and "male" individuals, which is reassuring, as historically, there have been a disproportionately low number of "female" individuals represented within research, and there is a particular focus on improving the representation of women within cardiology data.[24,30] Where race and ethnicity categories were reported, however, the majority of the individuals represented were within the "White" or "Caucasian" categories, and there was a complete lack of representation of individuals from any country within the continent of Africa. The lack of diversity within HF research has been widely reported, and while representation of women has improved, perhaps in response to initiatives like the SAGER guidelines, there is significant work to be done in improving the availability of data from diverse demographic groups.[31-34] In the United States, HF rates and clinical severity are much worse for Black people in comparison to people in all other racial or ethnic categories.[35] In order for AI developers to understand how their technologies may perform for these at-risk groups, data need to be accessible.[8,36,37] Datasets can be built with diversity and inclusivity in mind: one dataset included within this review (MESA [Multi-Ethnic Study of Atherosclerosis]) had a particular focus on obtaining a "multi-ethnic" cohort, which resulted in a diverse group of individuals being included.[38] Prespecifying which attributes are relevant given the context of a particular dataset and identifying barriers to inclusion may help achieve greater diversity, but this needs to be weighed up against the risk of selection bias.

A trade-off between need for demographic reporting and risk of identification. Accessibility of demographic data is necessary for incorporating health equity considerations in the development of AI health technologies. However, demographic attributes like race, ethnicity, sex, gender, and socioeconomic status are all imperfect representations of complex intersecting phenomena that can themselves introduce bias. In some localities, certain demographic attributes may be seen as inappropriate, culturally taboo, or even unlawful to collect or publish.[28,39] Demographic attributes within health data have themselves been a facilitator for stigmatization and discrimination within health care.[29] Here there is a trade-off: demographic data are required to assess the performance of AI health technologies for different groups known to experience worse health outcomes, but there could be situations where collecting this information could put those individuals at risk of identification and associated harms. Consent considerations, governance, and ethical review are key to protecting individuals against misuse of health data but these considerations were not widely reported within the identified HF datasets. To earn trust from communities in the responsible use of health data, it is essential that processes are transparently reported, ethical principles are adhered to, and individuals from relevant demographic groups are involved in the dataset curation and AI development processes from the outset.[40,41]

STUDY LIMITATIONS. A robust search strategy was used within this systematic review, but the list of datasets identified is unlikely to be exhaustive of all available datasets used in the development of AI health technologies for HF. Given that research articles are usually published after the datasets themselves, newer HF datasets published after August 2022 are unlikely to be included. Despite this limitation, datasets surfaced by this study highlight the cultures of practice in contemporary health care data curation and reporting. The practice of utilizing clinical research data within AI development is a novel finding, but all other conclusions drawn within this paper are consistent with similar reviews conducted in other disease areas.[9-11] While our strategy to include only papers relating to AI health technologies means there may be other datasets not included in this review, our intent was to analyze only those datasets underpinning AI health technologies which have actually been developed, rather than to assess all potentially usable data. A key finding was the large number of datasets that were effectively inaccessible

CENTRAL ILLUSTRATION **Diversity and Inclusion Within Datasets in Heart Failure: A Systematic Review**



A: Geographical distribution of datasets used to develop AI algorithms for heart failure, showing number of datasets per country.
B: The number of datasets reporting metadata items
C: Number of datasets according to data type.

Laws E, et al. JACC Adv. 2025;4(3):101610.

(A) Geographical distribution of datasets used to develop AI algorithms for heart failure, showing number of datasets per country. Some datasets included data from several countries, which is not represented in this figure. USA (n = 33), China (n = 5), Taiwan (n = 5), Japan (n = 4), South Korea (n = 4), Brazil (n = 3), UK (n = 3), Belgium, Canada, France, Greece, Singapore (all n = 2), Argentina, Hong Kong, India, Indonesia, Malaysia, New Zealand, Norway, Philippines, Russia, Sweden, and Thailand (all n = 1). (B) Number of datasets according to data type. The datasets involved data of different types including echocardiography: The dataset includes data about echocardiology (numerical, text, or image-based). Comorbidities: The dataset has information about individuals diagnoses other than heart failure, for example, diabetes, etc. ECG: The dataset includes electrocardiogram data (numerical, text, or image-based). Blood results (including BNP): The dataset includes brain natriuretic peptide data. Observations: The dataset includes clinical observation data, for example, blood pressure, heart rate, etc. Prescribed medication: The dataset includes information about patients' prescribed medication. NYHA functional class: The dataset includes NYHA functional class heart failure classification data. CXR: The dataset includes chest X-ray data (numerical, text, or image-based). Genetics (whole genome sequencing): The dataset included whole genome sequencing data. (C) The number of datasets reporting metadata items shows the number and percentage of identified datasets that reported metadata data items including: Number of individuals: The dataset reported the number of individuals included in the dataset. Country of origin: The dataset reported where the data were collected from. Age: The dataset reported the age of individuals included. Sex/Gender: The dataset reported sex and/or gender. Sex and gender are different concepts but since not all datasets defined these terms, it was difficult to determine whether sex or gender had been reported. We therefore combine these 2 categories. Ethical review: The dataset reported their ethical review process. Consent: The dataset reported whether consent was obtained from individuals or waived by an ethics committee. Ground truth: The dataset reported how ground truths were determined, for example, where direct observational measurements were used to confirm a diagnosis. Race/Ethnicity: The dataset reported race and/or ethnicity. Race and ethnicity are separate concepts, but since not all datasets defined these terms, it was difficult to determine whether race or ethnicity had been reported. We therefore combine these 2 categories. Missing groups: The dataset reported groups with missing data or groups that were not included within the dataset. Modifications to data: The dataset reported where modifications had been made to the data. Socioeconomic status: The dataset reported information relating to socioeconomic status (including insurance status). AI = artificial intelligence; BNP = brain natriuretic peptide; ECG = electrocardiography.

(including many of those that claimed managed access), but this in turn limits our ability to characterize those datasets in terms of composition and level of metadata reporting. There may be additional embedded metadata within the data files or additional dataset documentation that could not be accessed as part of this review. The limitations of this study have been mitigated where possible, and the results highlight an acute need to improve documentation and representativeness of HF datasets used for the training and testing of AI technologies.

CONCLUSIONS

We highlight a lack of demographic reporting and under-representation of diverse groups within

datasets relating to heart failure available for development of AI health technologies. Responsible curation of transparently documented heart failure datasets is an essential step in allowing the creation of AI technologies that perform well for everyone. There is work to be done in improving accessibility of diverse health data. Challenge also lies beyond achieving a perceived level of representativeness: from this review, we infer that there are privacy and data governance considerations, generalizability, and ethical factors that all influence the utility of the dataset. The curation of health care datasets can be greatly enhanced with support of communities represented, but there is also value in adhering to evidence-based guidelines and recommendations, building the foundations for the development of AI technologies that are safe, effective, and equitable.[16]

FUNDING SUPPORT AND AUTHOR DISCLOSURES

This work is funded by the Health Foundation and the NHS AI Lab UK. STANDING Together was funded by The NHS AI Lab and The Health Foundation and was supported by the National Institute for Health and Care Research (NIHR; AI_HI200014). The authors have reported that they have no relationships relevant to the contents of this paper to disclose.

ADDRESS FOR CORRESPONDENCE: Dr Xiaoxuan Liu, Institute of Applied Health Science, University of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom. E-mail: x.liu.8@bham.ac.uk.

PERSPECTIVES

COMPETENCY IN SYSTEM-BASED PRACTICE: AI health technologies are being developed for use in heart failure to aid diagnosis and outcome prediction. The performance of an AI technology is affected by the extent to which the patient population for which the technology is intended resembles the data on which it was trained. This study aimed to assess the documentation of datasets being used to develop AI technologies in the context of HF, with a particular focus on demographic attributes to assess "who" is represented but also "how" they are represented. Many datasets were inaccessible and therefore inscrutable, while those that could be examined were mainly drawn from the global north. Heart failure imposes a disproportionate burden felt among minoritized ethnic and disadvantaged socioeconomic groups, yet most datasets did not describe race, ethnicity, or socioeconomic status. Where they did, most individuals were classed within the "White" or "Caucasian" categories.

TRANSLATIONAL OUTLOOK: Clinicians using or considering deploying AI HF technologies in practice should be cognizant of how these issues can exacerbate pre-existing disparities in care. The STANDING Together recommendations can be used as guidance for those involved in creating datasets to address these issues in the future.

REFERENCES

1. Choi DJ, Park JJ, Ali T, Lee S. Artificial intelligence for the diagnosis of heart failure. *NPJ Digit Med*. 2020;3(1):1-6. https://doi.org/10.1038/s41746-020-0261-3

2. Savarese G, Moritz Becher P, Lund LH, et al. Global burden of heart failure: a comprehensive and updated review of epidemiology. *Cardiovasc Res*. 2023;118:3272-3287. https://doi.org/10.1093/cvr/cvac013

3. Haq IU, Chhatwal K, Sanaka K, Xu B. Artificial intelligence in cardiovascular medicine: current insights and future prospects. *Vasc Health Risk Manag*. 2022;18:517-528. https://doi.org/10.2147/VHRM.S279337

4. Johnson AE, Brewer LPC, Echols MR, Mazimba S, Shah RU, Breathett K. Utilizing artificial intelligence to enhance health equity among patients with heart failure. *Heart Fail Clin*. 2022;18(2):259. https://doi.org/10.1016/J.HFC.2021.11.001

5. Artificial intelligence and machine learning (AI/ML)-enabled medical devices | FDA. Accessed January 7, 2025. https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices

6. Chen IY, Pierson E, Rose S, Joshi S, Ferryman K, Ghassemi M. Ethical machine learning in healthcare. *Annu Rev Biomed Data Sci*. 2021;4(1):123-144. https://doi.org/10.1146/annurev-biodatasci-092820-114757

7. Diamond M. Good Machine Learning Practice for Medical Device Development: Guiding Principles. 2021.

8. Ibrahim H, Liu X, Zariffa N, Morris AD, Denniston AK. Health data poverty: an assailable barrier to equitable digital health care. *Lancet Digit Health*. 2021;3(4):e260-e265. https://doi.org/10.1016/S2589-7500(20)30317-4

9. Wen D, Khan SM, Xu AJ, et al. Characteristics of publicly available skin cancer image datasets: a systematic review. *Lancet Digit Health*. 2022;4(1):e64-e74. https://doi.org/10.1016/S2589-7500(21)00252-1

10. Khan SM, Liu X, Nath S, et al. A global review of publicly available datasets for ophthalmological imaging: barriers to access, usability, and generalisability. *Lancet Digit Health*. 2021;3(1):e51-e66. https://doi.org/10.1016/S2589-7500(20)30240-5

11. Wu J, Liu X, Li M, et al. Clinical text datasets for medical artificial intelligence and large language models — a systematic review. *NEJM AI*. 2024;1(6):AIra2400012. https://doi.org/10.1056/AIRA2400012

12. Laws E, Palmer J, Alderman J, et al. Diversity, inclusivity and traceability of mammography datasets used in development of Artificial Intelligence technologies: a systematic review. *Clin Imaging*. 2025;118:110369. https://doi.org/10.1016/j.clinimag.2024.110369

13. Alderman JE, Charalambides M, Sachdeva G, et al. Revealing transparency gaps in publicly available COVID-19 datasets used for medical artificial intelligence development—a systematic review. *Lancet Digit Health*. 2024;6(11):e827-e847. https://doi.org/10.1016/S2589-7500(24)00146-8

14. Dewan P, Rørth R, Jhund PS, et al. Income inequality and outcomes in heart failure: a global between-country analysis. *JACC Heart Fail*.

2019;7(4):336-346. https://doi.org/10.1016/J.JCHF.2018.11.005

15. Hawkins NM, Jhund PS, McMurray JJV, Capewell S. Heart failure and socioeconomic status: accumulating evidence of inequality. *Eur J Heart Fail*. 2012;14(2):138-146. https://doi.org/10.1093/EURJHF/HFR168

16. Alderman JE, Palmer J, Laws E, et al. Tackling algorithmic bias and promoting transparency in health datasets: the STANDING Together consensus recommendations. *Lancet Digit Health*. 2025;7(1):e64-e88. https://doi.org/10.1016/S2589-7500(24)00224-3

17. Alderman JE, Palmer J, Laws E, et al. Tackling algorithmic bias and promoting transparency in health datasets: the STANDING together consensus recommendations. *NEJM AI*. 2024;2(1):AIp2401088. https://doi.org/10.1056/AIP2401088

18. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. *Syst Rev*. 2016;5(1). https://doi.org/10.1186/S13643-016-0384-4

19. Cerdeña JP, Plaisime MV, Tsai J. From race-based to race-conscious medicine: how anti-racist uprisings call us to act. *Lancet*. 2020;396(10257):1125-1128. https://doi.org/10.1016/S0140-6736(20)32076-6

20. Gabbert W. Concepts of ethnicity. *Lat Am Caribb Ethnic Stud*. 2006;1(1):85-103. https://doi.org/10.1080/17486830500510034

21. American Medical Association, Center for Health Justice. Advancing Health Equity: A Guide to Language, Narrative and Concepts. 2021.

22. American Medical Association. The national patient-centered clinical research network. Accessed January 8, 2025. https://pcornet.org/

23. Camanni G, Ciccone O, Lepri A, et al. 'Being disabled' as an exclusion criterion for clinical trials: a scoping review. *BMJ Glob Health*. 2023;8(11):e013473. https://doi.org/10.1136/BMJGH-2023-013473

24. Liu KA, Dipietro Mager NA. Women's involvement in clinical trials: historical perspective and future implications. *Pharm Pract*. 2016;14(1):708. https://doi.org/10.18549/PHARMPRACT.2016.01.708

25. Data standardization – OHDSI. Accessed October 3, 2024. https://www.ohdsi.org/data-standardization/

26. Overview - FHIR v5.0.0. Accessed October 3, 2024. https://www.hl7.org/fhir/overview.html

27. Rostamzadeh N, Mincu D, Roy S, et al. Healthsheet: development of a transparency artifact for health datasets. *ACM Int Conf Proceeding Ser*. 2022;37(111):1943-1961. https://doi.org/10.1145/3531146.3533239

28. Yordanov Y, Le Louarn A, Khoshnood A. Challenges in recording race and ethnicity data in biomedical research: the French and Swedish perspectives. *Eur J Emerg Med*. 2023;30(3):153-154. https://doi.org/10.1097/MEJ.0000000000001035

29. Keane A, Islam S, Parsons S, et al. Understanding who is and isn't involved and engaged in health research: capturing and analysing demographic data to diversify patient and public involvement and engagement. *Res Involv Engagem*. 2023;9(1):1-20. https://doi.org/10.1186/S40900-023-00434-5/FIGURES/4

30. The MESSAGE Initiative: integrating sex and gender in UK research - BHF. Accessed October 2, 2024. https://www.bhf.org.uk/what-we-do/news-from-the-bhf/news-archive/2023/december/the-message-initiative

31. Sullivan LT, Randolph T, Merrill P, et al. Representation of black patients in randomized clinical trials of heart failure with reduced ejection fraction. *Am Heart J*. 2018;197:43-52. https://doi.org/10.1016/J.AHJ.2017.10.025

32. Lewsey SC, Breathett K. Racial and ethnic disparities in heart failure: current state and future directions. *Curr Opin Cardiol*. 2021;36(3):320. https://doi.org/10.1097/HCO.0000000000000855

33. Vilcant V, Ceron C, Verma G, Zeltser R, Makaryus AN. Inclusion of under-represented racial and ethnic groups in cardiovascular clinical trials. *Heart Lung Circ*. 2022;31(9):1263-1268. https://doi.org/10.1016/J.HLC.2022.06.668

34. NIH policy and guidelines on the inclusion of women and minorities as subjects in clinical research | grants.nih.gov. Accessed April 16, 2024. https://grants.nih.gov/policy/inclusion/women-and-minorities/guidelines.htm

35. Nayak A, Hicks AJ, Morris AA. Understanding the complexity of heart failure risk and treatment in black patients. *Circ Heart Fail*. 2020;13(8):E007264. https://doi.org/10.1161/CIRCHEARTFAILURE.120.007264

36. Martschenko DO, Wand H, Young JL, Wojcik GL. Including multiracial individuals is crucial for race, ethnicity and ancestry frameworks in genetics and genomics. *Nat Genet*. 2023;55(6):895-900. https://doi.org/10.1038/s41588-023-01394-y

37. Wei S, Le N, Zhu JW, et al. Factors associated with racial and ethnic diversity among heart failure trial Participants: a systematic bibliometric review. *Circ Heart Fail*. 2022;15(3):e008685. https://doi.org/10.1161/CIRCHEARTFAILURE.121.008685

38. Bahrami H, Kronmal R, Bluemke DA, et al. Differences in the incidence of congestive heart failure by ethnicity: the multi-ethnic study of atherosclerosis. *Arch Intern Med*. 2008;168(19):2138-2145. https://doi.org/10.1001/ARCHINTE.168.19.2138

39. Chan CC, Neves AL, Majeed A, Faisal A. Bridging the equity gap towards inclusive artificial intelligence in healthcare diagnostics. *BMJ*. 2024;384:q490. https://doi.org/10.1136/BMJ.Q490

40. United Nations. A Human Rights-Based Approach to Data Leaving No One Behind in the 2030 Agenda for Sustainable Development. 2018.

41. Mohan K, Kitsos P, Williamson SM, Prybutok V. Balancing privacy and progress: a review of privacy challenges, systemic oversight, and patient perceptions in AI-driven healthcare. *Appl Sci*. 2024;14(2):675. https://doi.org/10.3390/APP14020675

KEY WORDS **artificial intelligence, digital health, diversity, equity, health data, health technology, inclusion**

APPENDIX **For supplemental tables and figures, please see the online version of this paper.**