

## Data resource profile

Manne-Goehler, Jennifer; Theilmann, Michaela; Flood, David ; Marcus , Maja-Emilia ; Andall-Brereton, Glennis; Agoudavi, Kokou; Lopez Arboleda , William Andres ; Aryal, Krishna K; Bicaba, Brice Wilfried; Bovet, Pascal; Brant, Luisa Campos Caldeira; Brian, Garry; Chamberlin, Grace ; Chen, Geoffrey ; Damasceno , Albertino ; Dorobantu, Maria; Dunn, Matthew; Ebert, Cara; Farzadfar, Farshad; Singh Gurung, Mongal

DOI:

[10.1093/ije/dyac125](https://doi.org/10.1093/ije/dyac125)

License:

None: All rights reserved

*Document Version*

Peer reviewed version

*Citation for published version (Harvard):*

Manne-Goehler, J, Theilmann, M, Flood, D, Marcus , M-E, Andall-Brereton, G, Agoudavi, K, Lopez Arboleda , WA, Aryal, KK, Bicaba, BW, Bovet, P, Brant, LCC, Brian, G, Chamberlin, G, Chen, G, Damasceno , A, Dorobantu, M, Dunn, M, Ebert, C, Farzadfar, F, Singh Gurung, M, Guwatudde, D, Houehanou, C, Houinato, D, Hwalla , NC, Jorgensen, JMA, Karki, KB, Labadarios, D, Lunet, N, Carvalho Malta, D, Martins, JS, Mayige, MT, Wong-McClure, R, Moghaddam, SS, Mwangi, JK, Mwalim, O, Norov, B, Quesnel-Crooks, S, Rohde, S, Seigle , JA, Sibai , AM, Silver, BK, Sturua, L, Stokes, A, Supiyev , A, Tsabedze, L, Zhumadilov, Z, Jaacks, LM, Atun, R, Davies, J, Geldsetzer, P, Vollmer, S & Bärnighausen, TW 2022, 'Data resource profile: the Global Health and Population Project on Access to Care for Cardiometabolic Diseases (HPACC)', *International Journal of Epidemiology*. <https://doi.org/10.1093/ije/dyac125>

[Link to publication on Research at Birmingham portal](#)

### **Publisher Rights Statement:**

This is a pre-copyedited, author-produced version of an article accepted for publication in *International Journal of Epidemiology* following peer review. The version of record, Manne-Goehler, J., et al., Data Resource Profile: The Global Health and Population Project on Access to Care for Cardiometabolic Diseases (HPACC), *International Journal of Epidemiology*, 2022, dyac125, is available online at: <https://doi.org/10.1093/ije/dyac125>

### **General rights**

Unless a licence is specified above, all rights (including copyright and moral rights) in this document are retained by the authors and/or the copyright holders. The express permission of the copyright holder must be obtained for any use of this material other than for purposes permitted by law.

- Users may freely distribute the URL that is used to identify this publication.
- Users may download and/or print one copy of the publication from the University of Birmingham research portal for the purpose of private study or non-commercial research.
- User may use extracts from the document in line with the concept of 'fair dealing' under the Copyright, Designs and Patents Act 1988 (?)
- Users may not further distribute the material nor use it for the purposes of commercial gain.

Where a licence is displayed above, please note the terms and conditions of the licence govern your use of this document.

When citing, please reference the published version.

### **Take down policy**

While the University of Birmingham exercises care and attention in making items available there are rare occasions when an item has been uploaded in error or has been deemed to be commercially or otherwise sensitive.

If you believe that this is the case for this document, please contact [UBIRA@lists.bham.ac.uk](mailto:UBIRA@lists.bham.ac.uk) providing details and we will remove access to the work immediately and investigate.

**TITLE: Data Resource Profile: The Global H Health and P Population Project on A Access to C Care for C Cardiometabolic diseases (HPACC)**

Jennifer Manne-Goehler<sup>1,2,10</sup>†, Michaela Theilmann<sup>3</sup>†, David Flood MD MSc<sup>4</sup>, Maja E. Marcus<sup>23</sup>, Glennis Andall-Brereton<sup>29</sup>, Kokou Agoudavi<sup>5</sup>, William Andres Lopez Arboleda<sup>3</sup>, Krishna K. Aryal<sup>6</sup>, Brice Bicaba<sup>7</sup>, Pascal Bovet<sup>54,55</sup>, Luisa Campos Caldeira Brant MD<sup>8</sup>, Garry Brian<sup>9</sup>, Grace Chamberlin<sup>10</sup>, Geoffrey Chen<sup>10</sup>, Albertino Damasceno<sup>50,51,52</sup>, Maria Dorobantu<sup>53</sup>, Matthew Dunn<sup>4</sup>, Cara Ebert<sup>11</sup>, Farshad Farzadfar<sup>12</sup>, Mongal Singh Gurung<sup>15</sup>, David Guwatudde<sup>16</sup>, Corine Houehanou<sup>17</sup>, Dismand Houinato<sup>17</sup>, Nahla Hwalla<sup>18</sup>, Jutta M. Adelin Jorgensen<sup>19</sup>, Khem B. Karki<sup>21</sup>, Demetre Labadarios<sup>22</sup>, Nuno Lunet<sup>50,51</sup>, Deborah Carvalho Malta MD<sup>24,25</sup>, João S. Martins<sup>26</sup>, Mary T. Mayige<sup>20</sup>, Roy Wong McClure<sup>27</sup>, Sahar Saeedi Moghaddam MSc<sup>28</sup>, Kibachio J. Mwangi<sup>13,30</sup>, Omar Mwalim<sup>31</sup>, Bolormaa Norov<sup>32</sup>, Sarah Quesnel-Crooks<sup>29</sup>, Sabrina Rohde<sup>3</sup>, Jacqueline A. Seiglie<sup>1,2</sup>, Abla Sibai<sup>33</sup>, Bahendeka K. Silver<sup>34,56</sup>, Lela Sturua<sup>35,36</sup>, Andrew Stokes<sup>37</sup>, Adil Supiyev<sup>47</sup>, Lindiwe Tsabedze<sup>38</sup>, Zhaxybay Zhumadilov<sup>48</sup>, Lindsay M Jaacks<sup>39,40,41</sup>†, Rifat Atun<sup>42,43</sup>†, Justine I. Davies<sup>44,45,57</sup>†, Pascal Geldsetzer<sup>3,14</sup>†, Sebastian Vollmer<sup>23</sup>†, Till W. Bärnighausen<sup>3,42,46</sup>†

**Affiliations:**

<sup>1</sup> Division of Infectious Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA

<sup>2</sup> Harvard Center for Population and Development Studies, Harvard University, Boston, MA, USA

<sup>3</sup> Heidelberg Institute of Global Health, Faculty of Medicine and University Hospital, Heidelberg University, Heidelberg, Germany

<sup>4</sup> University of Michigan, Ann Arbor, United States of America

<sup>5</sup> Togo Ministry of Health, Lome, Togo

<sup>6</sup> Nepal Health Sector Programme 3, Monitoring Evaluation and Operational Research Project, Abt Associates, Kathmandu, Nepal

<sup>7</sup> Institut Africain de Santé Publique, Ouagadougou, Burkina Faso

<sup>8</sup> Internal Medicine Department, School of Medicine, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

<sup>9</sup> The Fred Hollows Foundation New Zealand, Auckland, New Zealand

<sup>10</sup> Medical Practice Evaluation Center, Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA

<sup>11</sup> RWI-Leibniz Institute for Economic Research, Essen (Berlin Office), Germany

<sup>12</sup> Non-Communicable Diseases Research Center, Endocrinology and Metabolism Population Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran

<sup>13</sup> Division of Non-Communicable Diseases, Ministry of Health, Nairobi, Kenya

<sup>14</sup> Division of Primary Care and Population Health, Department of Medicine, Stanford University, Stanford, CA, USA

<sup>15</sup> Health Research and Epidemiology Unit, Ministry of Health, Bhutan

<sup>16</sup> Department of Epidemiology and Biostatistics, School of Public Health, Makerere University, Kampala, Uganda

<sup>17</sup> Laboratory of Epidemiology of Chronic and Neurological Diseases, Faculty of Health Sciences, University of Abomey-Calavi, Benin

<sup>18</sup> Faculty of Agricultural and Food Sciences, American University of Beirut, Beirut, Lebanon

- <sup>19</sup> Department of Public Health, University of Copenhagen, Copenhagen, Denmark
- <sup>20</sup> National Institute for Medical Research, Dar es Salaam, Tanzania
- <sup>21</sup> Department of Community Medicine and Public Health, Institute of Medicine, Tribhuvan University, Kathmandu, Nepal
- <sup>22</sup> Faculty of Medicine and Health Sciences, Stellenbosch University, South Africa
- <sup>23</sup> Department of Economics and Centre for Modern Indian Studies, University of Goettingen, Göttingen, Germany
- <sup>24</sup> Department of Maternal-Child Nursing and Public Health, School of Nursing, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
- <sup>25</sup> Graduate Program in Nursing, School of Nursing, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil
- <sup>26</sup> Faculty of Medicine and Health Sciences, National University of East Timor, Rua Jacinto Candido, Dili, Timor-Leste
- <sup>27</sup> Epidemiology Office and Surveillance, Caja Costarricense de Seguro Social, San Jose, Costa Rica
- <sup>28</sup> Endocrinology and Metabolism Research Center, Endocrinology and Metabolism Clinical Sciences Institute, Tehran University of Medical Sciences, Tehran, Iran
- <sup>29</sup> Caribbean Public Health Agency, Port of Spain, Trinidad and Tobago
- <sup>30</sup> Faculté de médecine, Université de Genève, Geneva, Switzerland
- <sup>31</sup> Zanzibar Ministry of Health, Mnazi Mmoja, Zanzibar
- <sup>32</sup> National Center for Public Health, Ulaanbaatar, Mongolia
- <sup>33</sup> Epidemiology and Population Health Department, Faculty of Health Sciences American University of Beirut, Beirut, Lebanon
- <sup>34</sup> St. Francis Hospital, Nsambya, Kampala, Uganda
- <sup>35</sup> Non-Communicable Diseases Department, National Center for Disease Control and Public Health, Tbilisi, Georgia
- <sup>36</sup> Petre Shotadze Tbilisi Medical Academy, Georgia
- <sup>37</sup> Department of Global Health, Boston University School of Public Health, Boston, MA
- <sup>38</sup> Ministry of Health, Mbabane, Eswatini
- <sup>39</sup> Global Academy of Agriculture and Food Security, The University of Edinburgh, Easter Bush Campus, Midlothian, EH25 9RG, UK
- <sup>40</sup> Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA, USA
- <sup>41</sup> Public Health Foundation of India, New Delhi, India
- <sup>42</sup> Department of Global Health and Population, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA
- <sup>43</sup> Department of Global Health and Social Medicine, Harvard Medical School, Harvard University, Boston, MA, USA
- <sup>44</sup> MRC/Wits Rural Public Health and Health Transitions Research Unit, School of Public Health, University of Witwatersrand, Johannesburg, South Africa
- <sup>45</sup> Institute of Applied Health Research, University of Birmingham, Birmingham, UK
- <sup>46</sup> Africa Health Research Institute, Somkhele, South Africa
- <sup>47</sup> Laboratory of Epidemiology and Public Health, Center for Life Sciences, Nazarbayev University, Astana, Kazakhstan
- <sup>48</sup> Nazarbayev University School of Medicine, Nur-Sultan city, Kazakhstan
- <sup>49</sup> Ministry of Health, Monrovia, Liberia
- <sup>50</sup> EPIUnit, Institute of Public Health, University of Porto, Porto, Portugal
- <sup>51</sup> Department of Public and Forensic Health Sciences and Medical Education, Faculty of Medicine, University of Porto, Porto, Portugal
- <sup>52</sup> Faculty of Medicine, Eduardo Mondlane University, Maputo, Mozambique
- <sup>53</sup> Department of Cardiology, Emergency Hospital of Bucharest, Bucharest, Romania
- <sup>54</sup> Ministry of Health, Victoria, Seychelles
- <sup>55</sup> University Center for Primary Care and Public Health (Unisanté), Lausanne, Switzerland

<sup>56</sup> Uganda Martyrs University, Kampala, Uganda

<sup>57</sup> Centre for Global Surgery, Department of Global Health, Stellenbosch University, Cape Town, South Africa

WORD COUNT: 3435 (incl. abstract, excl. tables and figures)

NUMBER OF TABLES AND FIGURES: 4

Corresponding Author: Division of Infectious Diseases, Brigham and Women's Hospital, Harvard Medical School, 75 Francis Street, Boston, MA 02115, Tel: 754-224-6060, Email: JMANNE@POST.HARVARD.EDU

†Equal contribution

## Key Features

- Though more than 4 in 5 deaths due to cardiovascular disease (CVD) occur in low- and middle-income settings, there have been few data sources that allow for empirical estimation of key relationships relevant to the epidemiology, health behavior, and health services of CVD risk factors at the level of the individual. The Global **H** Health and **P**opulation Project on **A**ccess to **C**are for **C**ardiometabolic Diseases (HPACC) is a novel data resource that fills this gap.
- The HPACC dataset consists of 77 nationally representative, population-based surveys conducted after 2005 of adults 15 years and older.
- The dataset includes 1,269,542 participants from 76 countries; 46 of these have a repeat survey available.
- The main categories of data collection include demographic and socioeconomic data, anthropometry, biological measures of disease status, healthcare service utilization for hypertension, diabetes, and hyperlipidemia and self-reported information on health behaviors relevant to CVD. The dataset is currently being expanded to include mental health, cervical cancer, HIV and injuries where available.
- The collaboration includes a global team of physicians, economists, and health policymakers. Most surveys contained in the dataset are publicly available; the harmonized, deidentified data and an accompanying dictionary can be requested from the corresponding author.

**KEYWORDS:** cardiovascular disease, diabetes, hypertension, global health, epidemiology, access to care

## Data Resource Basics

Cardiovascular disease (CVD) is the leading cause of morbidity and mortality globally.(1) In recent decades, a rising prevalence of major CVD risk factors including diabetes, hypertension, and hyperlipidemia has been observed in many low- and middle-income countries (LMICs) where CVD was not previously considered a major health priority. As economic development drives urbanization and changes in lifestyle habits in many LMICs, this trend is expected to continue.(2-4) While approximately 80% of CVD deaths now occur in LMICs,(5) there are fewer data sources from which *empirical* estimation of key indicators and relationships relevant to epidemiology, health behavior, and health services can be undertaken at the level of the individual. In this profile, we provide an overview of a novel data resource for the study of CVD risk factors in LMICs, entitled the Global **H**ealth and **P**opulation Project on **A**ccess to **C**are for **C**ardiometabolic Diseases (HPACC).

HPACC was initiated in 2016 by researchers at the University of Göttingen, Harvard and Heidelberg Universities, in collaboration with in-country partners, with the objective to gather and harmonize household survey data on CVD and its risk factors. The HPACC dataset is comprised of nationally representative, population-based, individual level surveys conducted in 2005 or later that are collated and then harmonized. The harmonization process is guided by the World Health Organization STEPwise approach to Surveillance (STEPS) survey instrument. This process is complex and includes aligning variable definitions across surveys such as income and asset data, checking data quality and ensuring consistency especially for biomarker measures, documenting variable skip patterns and confirming sample weights. This results in a dataset covering 76 LMICs (Figure 1) and including data from 1,269,542 participants from 77 surveys (the Republic of Tanzania, while one country, conducted two separate surveys for Tanzania and for Zanzibar). These 76 countries represent 4.4 billion adults, which is 79% of the LMIC and 65% of the global adult population share. The dataset is dynamic and new surveys are added as they become available. Furthermore, in many countries, eligible surveys are conducted

periodically, at least once every 5-10 years. For the most part, these repeat surveys consist of a subsequent cross-section of the population but do not follow the same participants longitudinally. The current version of the HPACC dataset includes the most recent data in a cross-sectional harmonization but there is an ongoing effort to expand the dataset to include repeated cross-sections where available. Availability of repeated survey waves by country is provided in detail in Table 2. The dataset is not publicly available, but access can be granted upon request.

The dataset was collated by an international group of collaborators to meet the need for rigorous, nationally representative evidence on key epidemiological relationships and health system performance relating to CVD risk factors in these understudied country contexts (Table 1). Its structure allows for the comparison of population groups across the socio-economic and demographic spectra within and across countries. Furthermore, it yields results that are representative of the entire population of the countries covered by the included surveys.

## **Data collected**

### Survey inclusion criteria and search methodology

A systematic search and inclusion procedure is used to identify eligible surveys for harmonization and inclusion in the HPACC dataset. The inclusion criteria are detailed in Table 1. In brief, the survey needs to have been conducted in a LMIC after 2005. Furthermore, individual-level data from surveys representative of a country's adult population need to be available and include either a blood pressure, blood glucose, or lipid measurement. The first systematic search was conducted in 2016. The search was updated in 2021, identifying additional surveys conducted in 2010 or later.

We use a four-part search methodology to screen surveys for inclusion in the dataset (Figure 2). First, all countries in which a World Health Organization (WHO) Stepwise Approach to Non-Communicable Disease (NCD) Risk-Factor Surveillance (STEPS) survey has been conducted are identified.<sup>(6)</sup> All eligible STEPS survey data are requested from the WHO STEPS data repository.<sup>(7)</sup> Prior to 2019, when these were not available from a repository, data requests were made directly to the country survey team. Second, a search of three well-regarded survey resources, the Demographic and Health Surveys (DHS),<sup>(8)</sup> the WHO Study on Global Ageing and Adult Health (SAGE),<sup>(9)</sup> and the Gateway to Global Aging studies,<sup>(10)</sup> is conducted. Third, we search three other commonly used summary resources, the NCD Risk Factor Collaboration (NCD RisC),<sup>(11)</sup> Global Health Data Exchange (GHDx),<sup>(12)</sup> and the International Diabetes Federation (IDF) Diabetes Atlas.<sup>(13)</sup> If these yield a potentially eligible survey, we perform a Google search to confirm and request the most recent survey data from the designated point of contact. Fourth, a systematic Google search is conducted using prespecified search terms (Figure 2), and the first 30 returned results are reviewed. If two surveys are available for a particular country, we prioritize either the most recent survey and/or the one that contains a larger number of variables of interest. To access non-publicly available survey data that we identify at any step during the search process, we contact each survey's lead investigator(s) at least twice to invite them to share data and participate as an HPACC collaborator. Repeat surveys are included based on the same search methodology and eligibility criteria, as new data become available. The dataset is updated periodically when new surveys are released and harmonized. Table 2 indicates the countries for which repeat surveys are already available and the years in which the repeat survey was conducted.

### Survey instruments

The HPACC dataset contains 57 STEPS surveys. The STEPS survey instrument is designed to help countries build and strengthen their surveillance capacity by obtaining core data on established chronic



disease risk factors. These surveys are the official approach developed by WHO for monitoring NCD risk factors in adults at the population level and include questionnaires, physical measurements, and biochemical measurements, with 'core,' 'expanded,' and 'optional' modules to allow for flexibility by country capacity and interest. The 'core' self-reported measures include data on demographic characteristics, tobacco and alcohol use, dietary behavior (fruit and vegetable intake), physical activity, and history of raised blood pressure, diabetes, raised total cholesterol, heart disease, and stroke. In addition, there are 20 non-STEPS surveys in the HPACC dataset. Among these, the DHS instrument is frequently used. The DHS, in partnership with country governments, are nationally representative household surveys that collect data on a wide range of health-related monitoring and impact evaluation indicators.(14) The standard surveys have large sample sizes (usually greater than 5,000 participants) and are conducted, on average, once every five years. While they have historically focused on maternal and child health and more recently HIV, later surveys have increasingly incorporated questions about health behavior, nutrition, and tobacco use. These recent DHS have also begun to measure blood pressure and incorporate biomarkers for diabetes, consistent with the HPACC inclusion criteria.

#### *Survey design, clustering, and weighting*

Most surveys in the HPACC dataset use multi-stage cluster sampling of respondents and almost all provide documentation on the applied sampling strategy. In all analyses, sampling weights account for item non-response and are rescaled to ensure either population-size weighting or equal weighting for each country, depending on the primary question of interest.

#### Thematic content and measures

##### *Demographic and socioeconomic characteristics*

All surveys included in the HPACC dataset contain information on participants' age and sex, along with a measure of years of schooling or educational attainment. The majority of these surveys also include

information about location of residence (rural/urban, 46 surveys), marital status (68 surveys), working status (72 surveys), pregnancy status (70 surveys), and a measure of household wealth (62 surveys). The construction of the wealth quintile within each survey depends on the available household wealth indicator. Surveys using an asset index have surveyed household ownership of a range of assets, dwelling characteristics, and other country-specific wealth indicators. Based on the standard DHS approach, we use principal component analysis to derive an asset index, from which we create household wealth quintiles.<sup>(15)</sup> Countries using a household income-based measurement mainly followed the WHO STEPS template questionnaire. In these cases, respondents were asked about the household earnings over the past year: weekly or monthly average, or year total. In cases where this question was left unanswered, respondents were asked to place their household in one of the given income brackets. These brackets usually were defined according to a country's national household income quintile thresholds. Both, continuous income and income brackets, are used in the creation of the wealth quintiles.

We assume that household incomes within a country follow a log-normal distribution and we are therefore able to combine income quintiles and categories.<sup>(16)</sup> In 11 surveys, we dismissed pre-coded income brackets as they displayed very large discrepancies with respect to the continuous income range or could otherwise not be correctly identified. However, as the pre-coded estimates were typically only asked of respondents that had not indicated a continuous income, this has resulted in only minor information losses.

#### *Anthropometric and biological measures of disease status*

The dataset includes height and weight measurements in all surveys and 69 surveys include waist and/or hip circumference. In terms of the three CVD risk factors of interest, 54 surveys measured blood pressure and diabetes and lipid biomarkers, and 9 surveys measured blood pressure and diabetes biomarkers only (Table 2). Blood pressure measurement was standardized, though the number of measurements that were taken and subsequently averaged per survey ranged from one to three. The most meaningful

differences exist in the assessment of diabetes biomarkers as a variety of different measures were used. 56 surveys measured fasting blood glucose only, four measured hemoglobin A1c only, and five measured both. We have collected extensive documentation of measurement details relating to each survey's approach to measurement of the biomarkers. For blood pressure measurements, we note the number of measurements, details of the procedure, and devices and cuff-sizes used. Documentation on the diabetes biomarker includes information on type of glucometer or measurement device, whether samples were capillary or venous, if plasma conversions are required, and the fasting status of respondents. The lipid measurements were also subject to modest variation in the approach to measurement, with countries choosing either a point-of-care approach or a lab-based measurement. In case of the former, the CardioCheck PA device was most frequently utilized. Of the surveys with lipid measurements, all measures total cholesterol and many additionally collect HDL, triglycerides, and LDL cholesterol. Details of these measurement approaches are available in Supplementary Tables S1, S2, and S3.

#### *Measures of healthcare service utilization*

A unique feature of the HPACC dataset is that underlying surveys collected information on measures of healthcare service utilization for hypertension, diabetes, and/or hyperlipidemia. For the vast majority of surveys this includes a question about (1) whether a person had ever received a screening test for the respective condition, (2) whether they have been diagnosed with this condition, and (3) whether they are treated (have received advice or currently take medication) for the condition. 60 surveys include information on ever having visited a traditional healer and 65 on whether they currently use traditional medicine for the respective condition. 44 of the included surveys asked respondents about prior history of CVD including heart attack or stroke. In addition, in the case of diabetes, of the 70 surveys with information on diabetes medication, 66 contained questions about use of oral medications and insulin separately. 68 surveys also included a suite of questions about whether a participant had received lifestyle counseling from a healthcare provider with respect to losing weight, physical activity, reducing salt intake, increasing fruit and vegetable intake, and/or smoking cessation.

### *Behavioral measures*

The majority of surveys include self-reported information on health behaviors relevant to CVD. This includes information about dietary patterns of fruit and vegetable consumption and salt intake (61 surveys). In addition, 63 include a physical activity inventory mainly using the Global Physical Activity Questionnaire (GPAQ). The GPAQ measures moderate and vigorous activity at work and during leisure time and travel to and from places. This enables the calculation of a standardized estimate of activity in Metabolic Equivalents of Tasks (METs), which is in turn a ratio of the person's working metabolic rate relative to their resting rate. Furthermore, 60 surveys capture sedentary behavior. The GPAQ has been validated for use in several LMICs previously and is a recognized tool for assessing physical activity level in resource-limited settings.(17-19) Surveys included in the HPACC dataset also ask questions about alcohol (67 surveys) and tobacco use (76 surveys), including the history of use, frequency, and intensity.

### *Novel research frontiers*

The STEPS surveys and several of the other surveys included in the HPACC dataset also have available modules on other chronic conditions that are of particular interest globally. These include sections about mental health, cervical cancer, and oral health within the context of the STEPS surveys. Furthermore, recent STEPS surveys include information on injury and violence. The DHS includes additional modules about nuanced health behaviors (sugar-sweetened beverage intake or fast/fried food consumption, for example), HIV, and expanded nutritional biomarkers (for example micronutrient status). A codebook summarizing all harmonized variables is provided in the appendix (Appendix Table IV).

### Harmonization process

To harmonize these surveys, we performed a detailed review of the survey documentation with respect to the design and questionnaires from which the data were obtained. The harmonization process is based on and guided by the STEPS survey instrument. We defined the core group of variables that were

preserved across instruments and established common definitions for variables where response items were heterogeneous. We perform detailed data quality assessments prior to harmonization and collation.

### Brief overview of the study population

The current HPACC dataset contains data from 1,269,542 individuals aged 15 years and older who were eligible for at least one of the three measurements: blood pressure, blood glucose/glycated haemoglobin (HbA1c), or lipids. Their unweighted mean age is 35.1 (SD: 13.9) and 73.9% are female. The educational attainment of the population is 21.9% with no formal schooling, 20.2% with at least some primary school education or primary school completed and 56.6% with at least some secondary education or greater. The overall weighted prevalence (derived from the respective body measurement and self-reported use of medication) of hypertension is 29.4% (95% CI 28.8-30.0) of diabetes 8.6% (95% CI: 7.9-9.4), and of hyperlipidemia 6.7% (95% CI 6.2-7.2).

### **Data Resource Use**

The HPACC dataset has already provided several important and novel insights about the epidemiological relationships and health system performance for CVD risk factors in LMICs. To date, over 15 research articles using the HPACC dataset have been published in peer-reviewed journals. These analyses cover a range of topics such as an evaluation of health system performance for the management of hypertension (in 44 LMICs),(20) the variation in eligibility for hypertension treatment depending on clinical guideline (in 50 LMICs),(21) targeting of hypertension screening through easily identifiable individual characteristics,(22) and the relationship between estimated CVD risk and hypertension undertreatment/overtreatment (in 45 LMICs).(23) Health system performance for the management of diabetes has been analyzed in three separate studies using the HPACC data, which included 12, 28, and 55 LMICs, respectively.(24-26) Furthermore, analyses of the association between diabetes and socio-

economic status,(27) as well as anthropometric measures, including body mass index (BMI)(28), and adult height(29) have been conducted. In addition to analyses focused on diabetes and hypertension, the HPACC data have been used to describe the cascade of care for hypercholesterolemia and estimate statin coverage in LMICs.(30, 31) Health systems investigation in this dataset has also included a comprehensive consideration of country preparedness indicators for management of CVD risk factors.(32)

Furthermore, the HPACC data have been used to explore important behavioral risk factors for CVD and diabetes, including the patterns of fruit and vegetable consumption,(33) as well as sex differences in dietary behaviors..(34) Another study estimated the lifetime prevalence of cervical cancer screening.(35) Furthermore, the HPACC dataset is used in modeling studies such as the impact of increased diagnosis, treatment and control of diabetes.(36)

### Participating countries institutions and researchers

The harmonization of survey data for the HPACC dataset is only possible thanks to the large number of collaborators from these countries. These researchers play an active and crucial role in all studies that have been published so far and those which are planned. We have ongoing efforts to identify, contact, and foster collaboration with the survey teams of publicly available datasets. Furthermore, the involvement of young researchers from LMICs is one of HPACC's priorities and efforts to do so are being intensified.

## **Strengths and Weaknesses**

### Strengths

The HPACC dataset has several important strengths. First, uniquely, the data are at the individual-level and population-based and are harmonized to allow granular examination of key epidemiological and

health systems relationships, including variation across demographic and socioeconomic groups within countries. Moreover, the dataset includes nationally representative survey data from 76 countries whose combined population is approximately 79% (4.4 billion people) of the adult population living in LMICs. The extensive geographic coverage has enabled analyses at the global and regional levels, along with cross-country comparisons. Second, these surveys all contain biomarkers and biological measurements, including blood pressure, fasting blood glucose/hemoglobin A1c, and/or lipids, as well as anthropometry. Given that CVD risk factors of interest are defined based on these measurements, the HPACC dataset represents a substantial advance as few population-based surveys collected blood-based measurements prior to 2005. Third, most of these surveys include at least one measure of household income or asset ownership, which have been harmonized into wealth quintiles. The availability of this measure has allowed for a unique exploration of important relationships between prevalence and care for CVD risk factors and an individual's socioeconomic status.

### Weaknesses

The limitations of the HPACC dataset center primarily on the challenge of harmonizing heterogeneous measures across many surveys from diverse geographic and cultural contexts. This includes the wealth measure which varies across surveys. In addition, there is substantial heterogeneity in the measurement of biomarkers of diabetes and hyperlipidemia. To ensure transparency, we have documented these variations, including the survey-specific approach to eliciting fasting status, biomarker measurement (point-of-care versus laboratory-based), and nuances about the specific measurement instruments that can impact the interpretation of results, such as plasma-equivalent blood glucose measures. We would like to note that surveys recording self-reported information are standardized within survey types (STEPS, DHS) and largely comparable. A second limitation is that the measures of access to care for all CVD risk factors are self-reported by survey participants and thus subject to recall bias. Moreover, self-report of access may be affected by education, or health literacy, which may lead to underestimates in the prevalence of health behaviors and barriers to health care. Other potential biases are those of the

interviewer and social desirability. Third, the included survey were conducted during the past 15 years. In particular older surveys do not necessarily describe the current situation in the respective country. However, the dataset is updated on a regular basis and, thus, includes the newest available data. Fourth, due to the cross-sectional nature of the data, it is currently not possible to describe time trends. It is planned to harmonize and include older and future waves conducted in LMICs, which will allow us to look at changes over time.

### **Data Resource Access**

Many surveys contained in the HPACC dataset are publicly available. The two most common data sources are the WHO data repository and the DHS website.(7, 8, 37) Several additional surveys have been obtained through formal requests of survey teams whose data is not already made public. The pooled, harmonized, deidentified participant-level HPACC dataset and accompanying data dictionary have been created through a partnership between Harvard University, University of Göttingen and Heidelberg University, in collaboration with all country-level survey teams. Access can be requested through the corresponding author. More information about HPACC including additional contact information for the collaboration can be found on [www.hpaccproject.org](http://www.hpaccproject.org) or by emailing Jennifer Manne-Goehler at [jmanne@post.harvard.edu](mailto:jmanne@post.harvard.edu).

### **Ethics approval**

The included population-based surveys sought ethical approval from the respective country's ethics review committee prior to data collection. All surveys followed standardized ethics procedures, such as asking for participants' informed consent to participate in the respective survey. The final collated HPACC dataset is de-identified and no investigator can contact nor re-identify subjects. The Federal Policy for the Protection of Human Subjects (45 CFR Part 46) states that studies are excluded from an IRB review if "information, which may include information about biospecimens, is recorded by the investigator in such a manner that the identity of the human subjects cannot readily be ascertained directly or through



identifiers linked to the subjects, the investigator does not contact the subjects, and the investigator will not re-identify subjects.” The HPACC dataset was designated as Non-Human Subjects Research by the Harvard T. H. Chan School of Public Health in 2018 under protocol #IRB16-1915.

### **Data availability**

See Data Resource Access, above.

### **Supplementary data**

Supplementary data are available at IJE online.

### **Author contributions**

JMG and MT wrote the original draft. DF and MEM validated the results. JMG, MT, DF, MEM, LMJ, RA, JID, PG, SV, and TWB co-conceived the HPACC project. All authors participated in the data curation and critically reviewed the article draft. JMG and MT are the guarantors of the work.

### **Funding**

Each STEPS survey is co-funded by the country’s government and the WHO. DHS are co-funded by the United States Agency for International Development (USAID) and the respective country’s government. The funding of the other surveys are mostly co-funded by a country’s government, universities, international organizations, and sometimes supported by local sponsors. The creation of the final collated dataset has been funded by the Harvard McLennan Family Fund and the Alexander von Humboldt Foundation as well as institutional funds from the Universities of Heidelberg and Göttingen.

### **Acknowledgments**

We would like to thank each of the country-level survey teams and study participants who made the compilation of this dataset possible. We also gratefully acknowledge the high performance computing

bwHPC and data storage service SDS@hd supported by the Ministry of Science, Research and the Arts Baden-Württemberg (MWK) and the German Research Foundation (DFG) through grants INST 35/1134-1 FUGG , INST 35/1314-1 FUGG, and INST 35/1503-1 FUGG.

**Conflict of interest**

None declared

## References

1. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*. 2020;396(10258):1204-22.
2. Tucker KL, Buranapin S. Nutrition and aging in developing countries. *J Nutr*. 2001;131(9):2417S-23S.
3. Sudharsanan N, Geldsetzer P. Impact of Coming Demographic Changes on the Number of Adults in Need of Care for Hypertension in Brazil, China, India, Indonesia, Mexico, and South Africa. *Hypertension*. 2019;73(4):770-6.
4. United Nations DoEaSA, Population Division,. *World Population Prospects 2019, Volume II: Demographic Profiles (ST/ESA/SER.A/427) 2019* [Available from: [https://population.un.org/wpp/Publications/Files/WPP2019\\_Highlights.pdf](https://population.un.org/wpp/Publications/Files/WPP2019_Highlights.pdf)].
5. Roth GA, Mensah GA, Johnson CO, Addolorato G, Ammirati E, Baddour LM, et al. Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. *J Am Coll Cardiol*. 2020;76(25):2982-3021.
6. Riley L, Guthold R, Cowan M, Savin S, Bhatti L, Armstrong T, et al. The World Health Organization STEPwise Approach to Noncommunicable Disease Risk-Factor Surveillance: Methods, Challenges, and Opportunities. *Am J Public Health*. 2016;106(1):74-8.
7. World Health Organization. *NCD Microdata Repository 2020* [Available from: <https://extranet.who.int/ncdsmicrodata/index.php/catalog>].
8. ICF. *The DHS Program: Demographic and Health Surveys 2021* [Available from: <https://dhsprogram.com/>].
9. WHO. *WHO Study on global AGEing and adult health (SAGE) 2021* [Available from: <https://www.who.int/healthinfo/sage/en/>].
10. Program on Global Aging, Health, and Policy, Center for Economic and Social Research (CESR). *Gateway to Global Aging Data 2021* [Available from: <https://g2aging.org/>].
11. (NCD-RisC) NRFC. *Data Downloads 2021* [Available from: <http://nc-drisc.org/data-downloads.html>].
12. Institute for Health Metrics and Evaluation. *Global Health Data Exchange 2021* [Available from: <http://ghdx.healthdata.org/gbd-results-tool>].
13. International Diabetes Federation. *IDF Diabetes Atlas, 9th edition*. International Diabetes Federation; 2019.
14. ICF. *Demographic and Health Survey: the DHS Website: USAID; 2021* [Available from: <https://dhsprogram.com/Methodology/Survey-Types/DHS.cfm>].
15. Filmer D, Pritchett LH. Estimating wealth effects without expenditure data--or tears: an application to educational enrollments in states of India. *Demography*. 2001;38(1):115-32.

16. Harttgen K, Vollmer S. Using an asset index to simulate household income. *Economics Letters*. 2013;121(2):257-62.
17. Bull FC, Maslin TS, Armstrong T. Global physical activity questionnaire (GPAQ): nine country reliability and validity study. *J Phys Act Health*. 2009;6(6):790-804.
18. Cleland CL, Hunter RF, Kee F, Cupples ME, Sallis JF, Tully MA. Validity of the global physical activity questionnaire (GPAQ) in assessing levels and change in moderate-vigorous physical activity and sedentary behaviour. *BMC Public Health*. 2014;14:1255.
19. Chu AH, Ng SH, Koh D, Muller-Riemenschneider F. Reliability and Validity of the Self- and Interviewer-Administered Versions of the Global Physical Activity Questionnaire (GPAQ). *PLoS One*. 2015;10(9):e0136944.
20. Geldsetzer P, Manne-Goehler J, Marcus ME, Ebert C, Zhumadilov Z, Wesseh CS, et al. The state of hypertension care in 44 low-income and middle-income countries: a cross-sectional study of nationally representative individual-level data from 1.1 million adults. *Lancet*. 2019;394(10199):652-62.
21. Sudharsanan N, Theilmann M, Kirschbaum TK, Manne-Goehler J, Azadnajafabad S, Bovet P, et al. Variation in the Proportion of Adults in Need of Blood Pressure-Lowering Medications by Hypertension Care Guideline in Low- and Middle-Income Countries: A Cross-Sectional Study of 1 037 215 Individuals From 50 Nationally Representative Surveys. *Circulation*. 2021;143(10):991-1001.
22. Kirschbaum TK, Theilmann M, Sudharsanan N, Manne-Goehler J, Lemp JM, De Neve JW, et al. Targeting Hypertension Screening in Low- and Middle-Income Countries: A Cross-Sectional Analysis of 1.2 Million Adults in 56 Countries. *J Am Heart Assoc*. 2021;10(13):e021063.
23. Peiris D, Ghosh A, Manne-Goehler J, Jaacks LM, Theilmann M, Marcus ME, et al. Cardiovascular disease risk profile and management practices in 45 low-income and middle-income countries: A cross-sectional study of nationally representative individual-level survey data. *PLoS Med*. 2021;18(3):e1003485.
24. Manne-Goehler J, Geldsetzer P, Agoudavi K, Andall-Brereton G, Aryal KK, Bicaba BW, et al. Health system performance for people with diabetes in 28 low- and middle-income countries: A cross-sectional study of nationally representative surveys. *PLoS Med*. 2019;16(3):e1002751.
25. Manne-Goehler J, Atun R, Stokes A, Goehler A, Houinato D, Houehanou C, et al. Diabetes diagnosis and care in sub-Saharan Africa: pooled analysis of individual data from 12 countries. *Lancet Diabetes Endocrinol*. 2016;4(11):903-12.
26. Flood D, Seiglie J, Dunn M, Tschida S, Theilmann M. The state of diabetes treatment coverage in 55 low-income and middle-income countries: a cross-sectional study of nationally representative, individual-level data in 680 102 adults. *The Lancet Healthy Longevity*. 2021;2(6):E340-E51.
27. Seiglie JA, Marcus ME, Ebert C, Prodromidis N, Geldsetzer P, Theilmann M, et al. Diabetes Prevalence and Its Relationship With Education, Wealth, and BMI in 29 Low- and Middle-Income Countries. *Diabetes Care*. 2020;43(4):767-75.
28. Teufel F, Seiglie JA, Geldsetzer P, Theilmann M, Marcus ME, Ebert C, et al. Body-mass index and diabetes risk in 57 low-income and middle-income countries: a cross-sectional study of nationally representative, individual-level data in 685 616 adults. *Lancet*. 2021;398(10296):238-48.

29. Teufel F, Geldsetzer P, Manne-Goehler J, Karlsson O, Koncz V, Deckert A, et al. Analysis of Attained Height and Diabetes Among 554,122 Adults Across 25 Low- and Middle-Income Countries. *Diabetes Care*. 2020;43(10):2403-10.
30. Marcus ME M-GJ, Theilmann M, Farzadfar F, Moghaddam SS, Keykhaei M, Hajebi A, Tschida S, Lemp JM, Aryal KK, Dunn M, Houehanou C, Bahendeka S, Rohloff P, Atun R, Bärnighausen TW, Geldsetzer P, Ramirez-Zea M, Chopra V, Heisler M, Davies JI, Huffman MD, Vollmer S, Flood D. Use of statins for the prevention of cardiovascular disease in 41 low-income and middle-income countries: a cross-sectional study of nationally representative, individual-level data. *The Lancet Global Health*. 2022;10(3):e369-e79.
31. Marcus ME, Ebert C, Geldsetzer P, et al. Unmet need for hypercholesterolemia care in 35 low- and middle-income countries: A cross-sectional study of nationally representative surveys. *PLoS Med*. 2021;18(10):e1003841.
32. Davies JI, Reddiar SK, Hirschhorn LR, Ebert C, Marcus ME, Seiglie JA, et al. Association between country preparedness indicators and quality clinical care for cardiovascular disease risk factors in 44 lower- and middle-income countries: A multicountry analysis of survey data. *PLoS Med*. 2020;17(11):e1003268.
33. Frank SM, Webster J, McKenzie B, Geldsetzer P, Manne-Goehler J, Andall-Brereton G, et al. Consumption of Fruits and Vegetables Among Individuals 15 Years and Older in 28 Low- and Middle-Income Countries. *J Nutr*. 2019;149(7):1252-9.
34. McKenzie BL, Santos JA, Geldsetzer P, Davies J, Manne-Goehler J, Gurung MS, et al. Evaluation of sex differences in dietary behaviours and their relationship with cardiovascular risk factors: a cross-sectional study of nationally representative surveys in seven low- and middle-income countries. *Nutr J*. 2020;19(1):3.
35. Lemp JM, De Neve JW, Bussmann H, Chen S, Manne-Goehler J, Theilmann M, et al. Lifetime Prevalence of Cervical Cancer Screening in 55 Low- and Middle-Income Countries. *JAMA*. 2020;324(15):1532-42.
36. Basu S, Flood D, Geldsetzer P, Theilmann M, Marcus ME, Ebert C, et al. Estimated effect of increased diagnosis, treatment, and control of diabetes and its associated cardiovascular risk factors among low-income and middle-income countries: a microsimulation model. *Lancet Glob Health*. 2021;9(11):e1539-e52.
37. World Health Organization. STEPwise Approach to NCD Risk Factor Surveillance (STEPS) 2021 [Available from: <https://www.who.int/teams/noncommunicable-diseases/surveillance/systems-tools/steps>].

## Figures and Tables

### Figure 1. Map of geographic coverage and number of conditions measured by country across 76 low-income and middle-income countries

*Note: Fiji had information on blood glucose but not blood pressure. Countries marked as high-income countries were consistently classified as high-income countries from 2005-2018 by the World Bank. The following countries are represented by dots: Cabo Verde, Comoros, Fiji, Kiribati, Marshall Islands, Samoa, São Tomé and Príncipe, Seychelles, Solomon Islands, St. Vincent and the Grenadines, Tonga, Vanuatu. Abbreviations: BP = blood pressure, BG = blood glucose, HbA1c = haemoglobin A1c, HIC = high-income country, LMIC = low- and middle-income country, CVD = cardiovascular disease*

*The map was generated using the ggplot2 package in R.*

**Table 1. Key details about the Global Health and Population Project on Access to Care for Cardiometabolic Diseases (HPACC) Data Resource**

**Figure 2. Global Health and Population Project on Access to Care for Cardiometabolic Diseases (HPACC) search strategy**

Abbreviations: DHS = Demographic and Health Survey; GHDx = Global Health Data Exchange; IDF = International Diabetes Federation; NCD RisC = NCD Risk Factor Collaboration; SAGE = Study of Global Ageing and Adult Health; STEPS = Stepwise Approach to Non-Communicable Disease Risk-Factor Surveillance, See S4 for a list of countries that were classified as low- or middle-income for at least one year from 2005-2021.

## Table 2. Survey type and year, sample size, response rate, and available measures and indicators by country

1. Abbreviations: STEPS = Stepwise Approach to Non-Communicable Disease Risk-Factor Surveillance, HH=household wealth

2. \*Median response rate

3. Albania, Egypt, Fiji, Indonesia, Mexico, Peru, Romania, Uganda, and Ukraine only collect data on tobacco use.

4. Repeat data are already pooled for the following countries and years: Bangladesh 2011 and 2018, Benin 2015 and 2018, Bhutan 2014 and 2019, Brazil 2013 and 2019, Ecuador 2012 and 2018, Chile 2009-10 and 2016-17, China 2009 and 2015, Fiji 2009 and 2011, Mexico 2009-12 and 2018-19, Kyrgyz Republic 2012 and 2013, Lebanon 2008-09 and 2017, Lesotho 2012 and 2014, Mongolia 2009 and 2013, Mozambique 2005 and 2015, Nepal 2013 and 2019, Sao Tome and Principe 2009 and 2019, South Africa 2012 and 2016, India 2015-16 and 2017-19.



