

Leveraging data science to understand and address multimorbidity in sub-Saharan Africa: the MADIVA protocol

Kerry Glover ¹, Tabitha Osler ^{1,2}, Kayode Adetunji ^{1,3}, Tanya Akumu,⁴ Gershim Asiki,^{5,6} Diana Awuor,⁷ Palwendé Boua,^{1,8} Victoria Bronstein,⁹ Joan Byamugisha,¹⁰ Jacques D Du Toit ^{11,12,13}, Barry Dwolatzky,³ Jaya George,^{14,15} Paul A Harris,¹⁶ Kobus Herbst ^{17,18}, Karen Hofman,¹⁹ Celeste Holden,¹⁹ Samuel Iddi ^{20,21}, Damazo T Kadengye,²² Kathleen Kahn,¹¹ Michelle Kamp ^{1,23}, Nhlamulo Khoza,^{1,2} Faith Kimongo ¹¹, Isaac Kisiangani ^{5,24}, Dekuwin E Kogda ⁸, Michael Klipin,^{25,26} Stephen P Levitt ³, Dylan Maghini ^{1,27}, Karabo Maila,¹ Eric Maimela,^{28,29} Daniel Maina Nderitu ^{5,30}, Ndivhuwo Makondo,^{10,31} Molulaqhoa Linda Maoyi ¹⁷, Reineilwe Given Mashaba,²⁸ Nkosinathi Gabriel Masilela ¹¹, Theophilous Mathema,¹ Phelelani Thokozani Mpangase ¹, Daphine T Nyachowe ^{1,32}, Daniel Ohene-Kwofie,^{3,11} Helen Robertson ³¹, Skyler Speakman,⁴ Evelyn Thsehla,¹⁹ Siphwiwe A Thwala ¹⁰, Roy Zent,³³ Francesc Xavier Gómez-Olivé ¹¹, Chodziwadziwa W Kabudula,¹¹ Patrick Opiyo Owili ⁷, Catherine Kyobutungi,²² Michèle Ramsay ¹, Stephen Tollman,^{11,34} Scott Hazelhurst ^{1,3}

To cite: Glover K, Osler T, Adetunji K, *et al.* Leveraging data science to understand and address multimorbidity in sub-Saharan Africa: the MADIVA protocol. *BMJ Health Care Inform* 2025;**32**:e101294. doi:10.1136/bmjhci-2024-101294

KG and TO contributed equally.

Received 10 September 2024
Accepted 25 June 2025



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to
Professor Scott Hazelhurst;
Scott.Hazelhurst@wits.ac.za

ABSTRACT

Introduction Multimorbidity (MM), defined as two or more chronic diseases in an individual, is linked to adverse outcomes. MM is increasing in sub-Saharan Africa due to rapidly advancing epidemiological and social transitions. The *Multimorbidity in Africa: Digital Innovation, Visualisation and Application* Research Hub (MADIVA) aims to address MM by developing data science solutions informed by stakeholder engagement.

Methods and analysis MADIVA uses complex, individual-level datasets from research centres in rural Bushbuckridge, South Africa and urban Nairobi, Kenya. These datasets will be harmonised, linked and curated, and then used to develop MM risk prediction models, novel data science methods and interactive dashboards for research and clinical use. Pilot projects and mentorship programmes will support data science capacity development.

Ethics and dissemination Ethics approval has been granted. Dissemination will occur through scientific meetings and publications. MADIVA is committed to making data FAIR: findable, accessible, interoperable and reusable.

INTRODUCTION

Multimorbidity (MM), the presence of ≥ 2 chronic health conditions in an individual, is increasing in Sub-Saharan African (SSA) countries.^{1 2} The rise in MM is driven by

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Data science methods, like machine learning, have been developed for a range of healthcare applications, from risk and prognosis prediction to diagnosis; however, limited knowledge is available about the use of these methods to tackle the challenges that MM presents in sub-Saharan Africa (SSA).

WHAT THIS STUDY ADDS

⇒ This study aims to develop and apply data science methods to gain a deeper understanding of MM in SSA using integrated datasets from multiple sources and provide clinically relevant and actionable insights.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ Data integration will provide insights into demographic, health and behavioural trends in different communities and enable richer predictive models for disease, while enhancing understanding of intervention pathways.
⇒ Data curated by *Multimorbidity in Africa: Digital Innovation, Visualisation and Application* will be made available for ongoing research endeavours.

ageing populations, rising obesity levels and risk factors related to urbanisation.^{2 3} MM is associated with premature mortality, lower

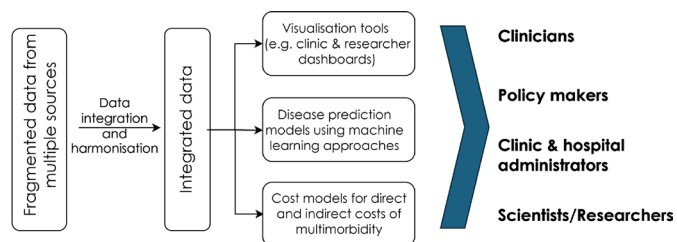


Figure 1 Action diagram of the *Multimorbidity in Africa: Digital Innovation, Visualisation and Application* process.

quality of life and increased use of healthcare services.⁴ This reduces productivity and increases healthcare costs for the state and affected individuals, adversely affecting the economy.¹⁵

Managing MM is complex, partly because clinical guidelines and primary healthcare systems address chronic conditions in isolation. This results in competing treatment priorities and complex management regimens, which can be burdensome for patients and increase the risk of adverse drug events and poor adherence.⁴ In SSA, MM management is further complicated by fragmented healthcare services and scarcity of systematic healthcare data collection and retrieval systems, especially at outpatient and community levels.^{6,7} The use of technology to tackle healthcare challenges is in its infancy.⁸ We hypothesise that well-integrated datasets, innovative data science analyses, visualisation and dissemination of information among stakeholders will enable significant change in the care of individuals with MM (figure 1).

To this end, we have initiated the *Multimorbidity in Africa: Digital Innovation, Visualisation and Application* Research Hub (MADIVA). MADIVA has access to rich longitudinal population datasets from the Health and Demographic Surveillance Systems (HDSS), clinical and laboratory records, ultrasound images and genomic data from two study sites: Bushbuckridge, South Africa and Nairobi, Kenya (table 1). MADIVA aims to: (1) curate, harmonise and build integrated datasets from both sites; (2) use data science techniques to develop MM risk and stratification models; (3) develop and implement data science methods, including human computer interaction (HCI) tools such as clinic and researcher dashboards to assist in the discovery of disease patterns and management of patients with MM and (4) strengthen capacity for integrated data science in healthcare in SSA.

METHODS

MADIVA is led from the University of the Witwatersrand (Wits), Johannesburg, South Africa with collaborators from the African Population and Health Research Center (APHRC) in Nairobi, Kenya; IBM Research Africa in Johannesburg and Nairobi; the South African Population Research Infrastructure Network hosted by the South African Medical Research Council (SAMRC) and the Vanderbilt University Medical Center in Nashville, USA. The multidisciplinary team includes professionals

with expertise in genetics, public health, bioinformatics, health economics, computing, law and clinical medicine.

Data sources

MADIVA will primarily use data generated at two sites (table 1). The SAMRC/Wits Rural Public Health and Health Transitions Research Unit is an HDSS in the Agincourt-Bushbuckridge subdistrict, Mpumalanga Province, South Africa comprising 31 contiguous villages, ~117 000 individuals undergoing rapid transition from primarily infant-maternal and infectious diseases to non-communicable conditions.⁹ APHRC runs the Nairobi Urban Health and Demographic Surveillance System, an urban-based HDSS covering two high-density informal settlements, Koro-gocho and Viwandani, in Nairobi (~90 000 individuals).¹⁰ More detail can be found in online supplemental material 1.

MADIVA cores and projects

MADIVA is organised into cores and projects (figure 2).

The Data Management and Analysis Core (DMAC)

The Data Management and Analysis Core (DMAC), alongside Project 1, is responsible for integrating and harmonising the longitudinal population-based datasets containing demographic and socioeconomic information with risk and morbidity surveys, clinic patient registries and verbal autopsy (VA) data (structured interviews with family or carers to determine cause-specific mortality¹¹). Data linkage within and across the two study sites uses both deterministic and probabilistic methods (see online supplemental materials).

Projects 1 and 2 use the integrated data to achieve their objectives with technical and development support provided by the DMAC. The DMAC has developed documentation and a data management plan to ensure reproducibility and transparency.

Project 1: data integration and visualisation

HCI tools: Project 1 and the DMAC are leading dataset integration and the design of dashboards to facilitate HCI. Using participatory design, key stakeholders—clinicians, public health officials and policymakers—are engaged to gather feedback and refine the tools. This process allows the development/validation of dashboards to ensure usability and alignment with stakeholder needs.

Clinic dashboards: integrated individual-level health information will be displayed for healthcare practitioners. Dashboards will provide aggregated data aimed at operational managers to assist in planning. A risk management stream will be developed to identify and manage errors that may impact on patient care.

Researcher dashboards: enable researchers, internal and external to MADIVA, to access MADIVA's integrated data, facilitating the generation of hypotheses for further studies.

Costing of MM: these will be estimated through a worksheet-based model. Inputs come from data collected at primary healthcare facilities and public data (eg,

Table 1 Secondary data* sources used in MADIVA

Project	Agincourt-Bushbuckridge, South Africa ¹⁴			Nairobi, Kenya		
	Data collected	Data frequency	Data volume	Data collected	Data frequency	Data volume
Health and socio-demographic surveys ^{9,10}	Demographic and social factors, health behaviours, medical history and exposures	Annually since 1992	~22 000 households, ~1 17 000 individuals, 31 villages	Demographic and social factors, health behaviours, medical history and exposures	Began 2002 with two to three rounds per year	33 462 households, 88 974 individuals
	Verbal autopsy	Deaths registered	93% of deaths from 1992 to 2020	Verbal autopsy	Deaths registered	~85% of deaths from 2002 to 2020
	NCD surveys	2002, 2007, 2014, 2015, 2018, 2020	2000–5000 participants	NCD surveys	2008, 2010, 2012, 2014	2000–5000 participants
African Research on Kidney Disease ¹⁵	Chronic kidney disease prevalence and risk factors in Africans	2018–2019	2021 adults			
	Genotype data		630 participants			
Africa Wits-INDEPTH study of genomic and environmental factors for cardiometabolic risk ^{16,17}	Demographic and social factors, health behaviours, medical history and exposures, anthropometric measurements, biomarkers	2015/2016 and 2020/2021	~2500 adults aged 40–70 years	Demographic and social factors, health behaviours, medical history and exposures, anthropometric measurements, biomarkers	2015/2016 and 2020/2021	~2100 adults aged 40–60 years
	Whole genome sequencing		100 participants	Whole genome sequencing		
	Gut microbiome	2020/2021	533 participants	Gut microbiome	2020/2021	237
Health facility records	Health data collected during illness and follow-up	2014–ongoing	Eight public health facilities, two district hospitals, ~1.5 million clinic visits from ~95 000 patients	Health data collected during illness and follow-up. Clinic-based studies for hypertension, diabetes, mental health, epilepsy		Three public health facilities and 31 private health facilities
Health and Ageing in Africa: A Longitudinal Study of an INDEPTH Community in South Africa ¹⁴	Demographic and social factors, health behaviours, social networks, anthropometric measurements, biomarkers, cognition, mental health and self-reported health	2014, 2015, 2018, 2019, 2021, 2022	5059 adults over 40 years			

*All data used in MADIVA is in digital format.

MADIVA, *Multimorbidity in Africa: Digital Innovation, Visualisation and Application*; NCD, non-communicable disease.

standard treatment guidelines and official population data¹²). The first model will estimate the cost of managing MM from a public healthcare perspective, considering only direct costs such as medication, investigations and health worker wages. A second model will estimate MM cost from a societal perspective (including indirect costs

like transport and lost productivity). This study will assist policy makers with resource allocation and priority setting.

Biomedical text processing. MADIVA will use a knowledge-based natural language processing approach to extract cause of death information from VA reports, leveraging

MADIVA Research Hub Organization

Site 1: Agincourt, South Africa | Site 2: Nairobi, Kenya | H3Africa | Other
 Africa Wits-INDEPTH study of genetic and environmental factors for cardiometabolic disease risk (AWI-Gen) | African Research on Kidney Disease (ARK)
 Health and Aging in Africa: longitudinal studies of an INDEPTH community (HAALS) | Health and socio-demographic surveys | Verbal autopsy
 HIV testing | Health facility records | NCD surveys

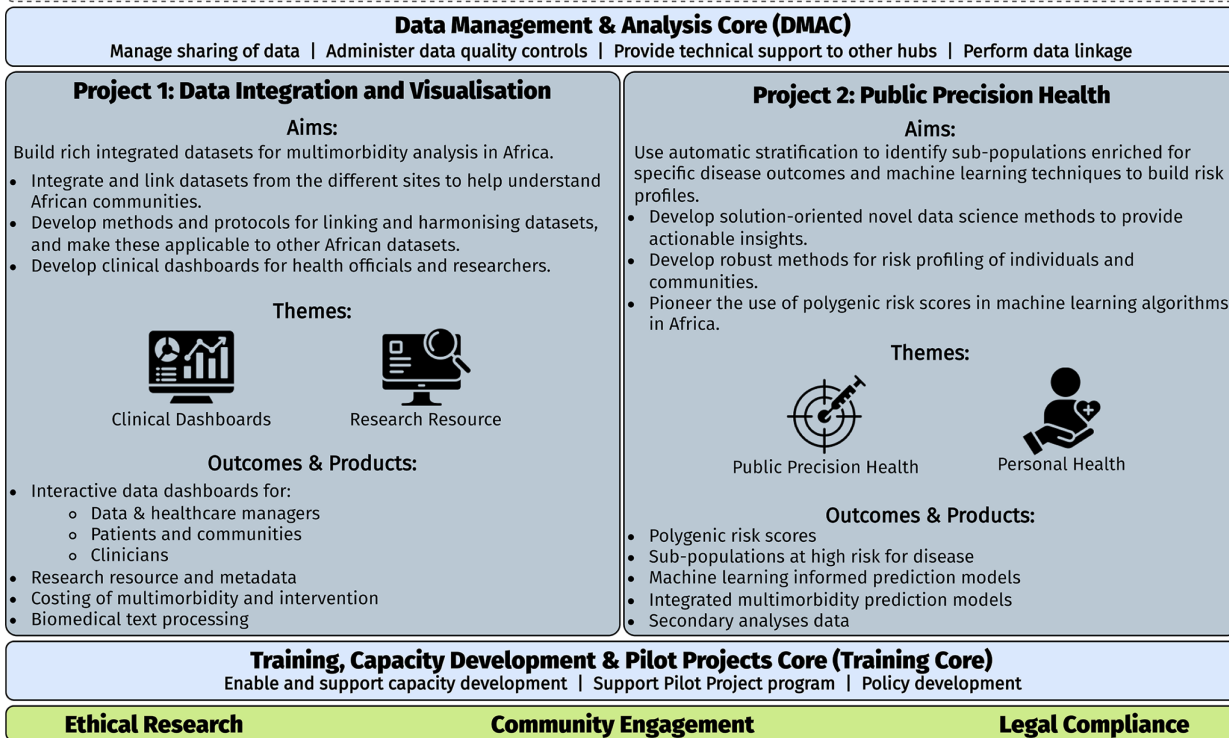


Figure 2 Project structure of the *Multimorbidity in Africa: Digital Innovation, Visualisation and Application* (MADIVA) consortium.

external medical knowledge bases. Other approaches, including the use of Large Language Models, may also be used.

Project 2: public precision health

Project 2 applies data-driven discovery to individual-level health-related data to predict outcomes. First, automatic stratification (AS) will be used to detect anomalous patterns and identify population subsets with an over-representation of individuals with specific health outcomes, including MM.¹³ By scanning multiple risk factors, AS identifies the combination of features associated with these outcomes. Second, machine learning models will be applied and assessed for the development of risk-prediction algorithms.

These studies begin with understanding strengths and limitations of the data, followed by a decision-making process involving domain experts - clinicians and scientists- to identify and include relevant features. Models will be built up from a limited number of demographic and clinical variables and some models will include genetic data, such as polygenic scores. Emphasis is placed on the explainability of outcomes at population and individual levels. Project 2 will further explore scientifically sound and ethically justifiable ways of communicating results to stakeholders and individuals through community and stakeholder engagement. The results of these

models could be used in the dashboards and feedback to clinicians.

The Training, Capacity Development and Pilot Project Core (Training Core)

The Training Core's primary role is to develop the MADIVA team—especially early career researchers—and stakeholders. The Core's training mandate is being met through workshops and teaching on data science, healthcare, genetics and innovation across MADIVA. The Training Core runs a mentorship programme to support trainees and early career researchers working within the MADIVA project.

The Training Core will assist the projects in transitioning from basic science to translational work through capacity building and sourcing appropriate training and linkages. The Vanderbilt University Medical Center and Vanderbilt Institute for Clinical and Translational Research are formal partners providing considerable assistance.

The Training Core supports pilot projects in order to extend MADIVA's work beyond its immediate scope and supporting early career researchers from SSA. Currently, MADIVA has three pilot projects—in Burkina Faso at the Clinical Research Unit of Nanoro and two in South Africa at the University of Limpopo and the Africa Health Research Institute in KwaZulu-Natal.

The Training Core has developed a Data Sharing Policy to guide data storage, management, exportability, ownership and sharing (see online supplemental materials).

OUTPUTS

Outputs from MADIVA will be disseminated through publications and presentations. Outputs include (see figure 2 for detail):

- ▶ Interactive researcher and clinic-level dashboards enabling key stakeholders to explore and use these datasets.
- ▶ Research resources (data/metadata—see online supplemental materials for availability).
- ▶ Costing models.
- ▶ Biomedical text processing.
- ▶ Risk prediction models and related results.
- ▶ Novel data science tools for automated stratification of MM.
- ▶ Policies and protocols to facilitate research practices in large collaborative studies.

Evaluation of the impact of the work will be a focus of the last 2 years of the grant.

Author affiliations

¹Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg, South Africa

²Division of Human Genetics, University of the Witwatersrand, Johannesburg, South Africa

³School of Electrical and Information Engineering, University of the Witwatersrand, Johannesburg, South Africa

⁴IBM Research – Africa, Nairobi, Kenya

⁵Chronic Disease Management Unit, African Population and Health Research Center, Nairobi, Kenya

⁶Department of Women's and Children's Health, Karolinska Institute, Stockholm, Sweden

⁷Research and Related Capacity Strengthening Unit, African Population and Health Research Center, Nairobi, Kenya

⁸Clinical Research Unit of Nanoro, Institut de Recherche en Sciences de la Sante, Nanoro, Burkina Faso

⁹School of Law, University of the Witwatersrand, Johannesburg, South Africa

¹⁰IBM Research – Africa, Johannesburg, South Africa

¹¹MRC/Wits Rural Public Health and Health Transitions Research Unit (Agincourt), University of the Witwatersrand Johannesburg, Johannesburg, South Africa

¹²Swiss Tropical and Public Health Institute, Allschwil, Switzerland

¹³University of Basel, Basel, Switzerland

¹⁴Wits Diagnostic Innovation Hub, University of the Witwatersrand, Johannesburg, South Africa

¹⁵Department of Chemical Pathology, National Health Laboratory Service, Johannesburg, South Africa

¹⁶Departments of Biomedical Informatics, Biostatistics, Biomedical Informatics & Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, Tennessee, USA

¹⁷DSI-SAMRC South African Population Research Infrastructure Network (SAPRIN), South African Medical Research Council, Durban, South Africa

¹⁸Africa Health Research Institute, Durban, South Africa

¹⁹SAMRC/Wits Centre for Health Economics and Decision Science – PRICELESS SA, University of the Witwatersrand, Johannesburg, South Africa

²⁰Data Synergy and Evaluation Unit, African Population and Health Research Center, Nairobi, Kenya

²¹Department of Statistics and Actuarial Science, University of Ghana, Accra, Ghana

²²African Population and Health Research Center, Nairobi, Kenya

²³Statistical Genetics Unit, King's College, London, UK

²⁴Department of Global Health, Amsterdam University Medical Centre, Amsterdam, The Netherlands

²⁵Department of Surgery, University of the Witwatersrand, Johannesburg, South Africa

²⁶WitsBits, Health Sciences Biomedical Informatics, University of the Witwatersrand, Johannesburg, South Africa

²⁷Department of Medicine (Hematology), Stanford University, Stanford, California, USA

²⁸DIMAMO Population Health Research Centre, University of Limpopo, Polokwane, South Africa

²⁹Department of Public Health, University of Limpopo, Polokwane, South Africa

³⁰School of Public Health, University of the Witwatersrand, Johannesburg, South Africa

³¹School of Computer Science and Applied Mathematics, University of the Witwatersrand, Johannesburg, South Africa

³²Steve Biko Centre for Bioethics, School of Clinical Medicine, University of the Witwatersrand, Johannesburg, South Africa

³³Division of Nephrology, Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA

³⁴INDEPTH Network, Accra, Ghana

X Kayode Adetunji @kayvins and Scott Hazelhurst @shazeZA

Acknowledgements We thank our NIH Science Officers, Qi Duan and Keith Mintzer; our Scientific Advisory Board; our DS-I ELSI partner, DSI-LAW (Pi Donrich Thaldar); NIH programme staff, especially Brad Newsome and Gifty Dankyi; Okechinyere Achilonu and the SBIMB Project Office and laboratory staff.

Contributors Equal contributions were made to the writing of this paper by TO and KG and all authors reviewed the paper. Project conceptualisation and funding acquisition: SH, CK, MR, ST, KK; data collection, curation and management: CWK, SI, TM, DMN, SH, MLM, DO-K, JG, NGM, K Herbst; investigation, analysis: FXG, KA, TA, VB, JB, CH, GA, DTK, SI, MK, NK, NM, DMN, TM, PTM, DTN, SS, ET, SAT, SH, ST, MR, IK; dashboard development: FXG, CWK, SPL, FK, KM, SH; pilot projects: PB, RGM, EM, DEK; ethics, training, mentorship, stakeholder engagement and data governance: POO, DA, VB, JDDT, HR, DTN, DMN, MK, RZ, PAH, KK; project administration: KG, SH; supervision FXG, SH, SI, CK, POO, CWK, MR, SS, MK, K Hofman. BD led dashboard development until his death, making key contributions to conceptualisation, stakeholder engagement, functionality and technological framework. He was not able to participate in manuscript development. SH is guarantor.

Funding Research reported in this publication was supported by the Fogarty International Center and National Institute of Biomedical Imaging and Bioengineering (NIBIB) and OD/Office of Strategic Coordination (OSC) and Office of Data Science Strategy (ODSS) of the National Institutes of Health, under Award Number U54 TW 012077 and a Strategic Health Innovation Partnerships grant from the South African Medical Research Council.

Disclaimer The content is solely the responsibility of the authors and does not necessarily represent the official views of the funders.

Competing interests None declared.

Patient consent for publication Not applicable.

Ethics approval The University of the Witwatersrand Human Research Ethics Committee (Medical) (HREC M210825), and the Kenyan AMREF Research and Ethics and Scientific Review Committee (AMREF-ESRC P1206/2022) have approved MADIVA's projects.

Provenance and peer review Not commissioned; externally peer reviewed.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

**ORCID iDs**

Kerry Glover <http://orcid.org/0000-0002-3778-2611>
 Tabitha Osler <http://orcid.org/0000-0003-2966-6327>
 Kayode Adetunji <http://orcid.org/0000-0003-0504-382X>
 Jacques D Du Toit <http://orcid.org/0000-0002-0003-0168>
 Kobus Herbst <http://orcid.org/0000-0002-5436-9386>
 Samuel Iddi <http://orcid.org/0000-0002-2366-2774>
 Michelle Kamp <http://orcid.org/0000-0002-4334-6883>
 Faith Kimongo <http://orcid.org/0009-0002-7730-6246>
 Isaac Kisiangani <http://orcid.org/0000-0002-1107-4334>
 Dekuwin E Kogda <http://orcid.org/0009-0009-1745-9567>
 Stephen P Levitt <http://orcid.org/0000-0001-6054-6134>
 Dylan Maghini <http://orcid.org/0000-0001-9542-492X>
 Daniel Maina Nderitu <http://orcid.org/0000-0001-9800-2431>
 Molulaqhoa Linda Maoyi <http://orcid.org/0000-0002-0176-098X>
 Nkosinathi Gabriel Masilela <http://orcid.org/0000-0002-4718-9336>
 Phelelani Thokozani Mpangase <http://orcid.org/0000-0001-8280-8940>
 Daphine T Nyachowe <http://orcid.org/0000-0003-2544-804X>
 Helen Robertson <http://orcid.org/0000-0002-4135-0379>
 Siphwe A Thwala <http://orcid.org/0000-0002-4587-2246>
 Francisc Xavier Gómez-Olivé <http://orcid.org/0000-0002-4876-0848>
 Patrick Opiyo Owili <http://orcid.org/0000-0002-9417-8947>
 Michèle Ramsay <http://orcid.org/0000-0002-4156-4801>
 Scott Hazelhurst <http://orcid.org/0000-0002-0581-149X>

REFERENCES

- 1 Afshar S, Roderick PJ, Kowal P, *et al.* Multimorbidity and the inequalities of global ageing: a cross-sectional study of 28 countries using the World Health Surveys. *BMC Public Health* 2015;15:776.
- 2 Brady E, Castelli M, Walker R, *et al.* The prevalence and social determinants of multimorbidity in South Africa. *World Med Health Policy* 2023;15:435–54.
- 3 Bradshaw D, Steyn K, Levitt N, *et al.* Non-communicable diseases - a race against time. 2011.1–4. Available: <https://www.samrc.ac.za/sites/default/files/attachments/2022-08/raceagainst.pdf>
- 4 Guthrie B, Payne K, Alderson P, *et al.* Adapting clinical guidelines to take account of multimorbidity. *BMJ* 2012;345:e6341.
- 5 Cabral GG, Dantas de Souza AC, Barbosa IR, *et al.* Multimorbidity and Its Impact on Workers: A Review of Longitudinal Studies. *Saf Health Work* 2019;10:393–9.
- 6 Basto-Abreu A, Barrientos-Gutierrez T, Wade AN, *et al.* Multimorbidity matters in low and middle-income countries. *J Multimorb Comorb* 2022;12:26335565221106074.
- 7 Wandai M, Aagaard-Hansen J, Day C, *et al.* data sources for monitoring non-communicable diseases and their risk factors in South Africa. *S Afr Med J* 2017;107:331–7.
- 8 Majnarić LT, Babić F, O'Sullivan S, *et al.* AI and Big Data in Healthcare: Towards a More Comprehensive Research Framework for Multimorbidity. *J Clin Med* 2021;10:766.
- 9 Kahn K, Collinson MA, Gómez-Olivé FX, *et al.* Profile: Agincourt health and socio-demographic surveillance system. *Int J Epidemiol* 2012;41:988–1001.
- 10 Wamukoya M, Kadengye DT, Iddi S, *et al.* The Nairobi Urban Health and Demographic Surveillance of slum dwellers, 2002–2019: Value, processes, and challenges. *Global Epidemiology* 2020;2:100024.
- 11 D'Ambruoso L, Boerma T, Byass P, *et al.* The case for verbal autopsy in health systems strengthening. *Lancet Glob Health* 2017;5:e20–1.
- 12 Statistics South Africa. Improving lives through data ecosystems. Available: <https://www.statssa.gov.za/> [Accessed 19 May 2024].
- 13 Speakman S, Tadesse GA, Cintas C, *et al.* n.d. Detecting Systematic Deviations in Data and Models. *Computer (Long Beach Calif)*56:82–92.
- 14 Gómez-Olivé FX, Montana L, Wagner RG, *et al.* Cohort Profile: Health and Ageing in Africa: A Longitudinal Study of an INDEPTH Community in South Africa (HAALS). *Int J Epidemiol* 2018;47:689–690j.
- 15 Fabian J, Kalyesubula R, Mkandawire J, *et al.* Measurement of kidney function in Malawi, South Africa, and Uganda: a multicentre cohort study. *Lancet Glob Health* 2022;10:e1159–69.
- 16 Ramsay M, Crowther N, Tambo E, *et al.* H3Africa AWI-Gen Collaborative Centre: a resource to study the interplay between genomic and environmental risk factors for cardiometabolic diseases in four sub-Saharan African countries. *Glob Health Epidemiol Genom* 2016;1:e20.
- 17 Ali SA, Soo C, Agongo G, *et al.* Genomic and environmental risk factors for cardiometabolic diseases in Africa: methods used for Phase 1 of the AWI-Gen population cross-sectional study. *Glob Health Action* 2018;11:1507133.