

BMJ Open Skills and competencies in health data analytics for health professionals: a scoping review protocol

Akhona Nkwanyana ,¹ Verona Mathews,² Iris Zachary,³ Vishwa Bhayani³

To cite: Nkwanyana A, Mathews V, Zachary I, *et al.* Skills and competencies in health data analytics for health professionals: a scoping review protocol. *BMJ Open* 2023;**13**:e070596. doi:10.1136/bmjopen-2022-070596

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2022-070596>).

Received 28 November 2022
Accepted 25 October 2023



© Author(s) (or their employer(s)) 2023. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Department of Psychology, University of the Western Cape, Cape Town, Western Cape, South Africa

²School of Public Health, University of the Western Cape, Cape Town, Western Cape, South Africa

³Department of Health Management and Informatics, University of Missouri, Columbia, Missouri, USA

Correspondence to

Akhona Nkwanyana;
akhona.nkwanyana@gmail.com

ABSTRACT

Introduction Healthcare data analytics is a methodological approach to the systematic analysis of health data, and it provides opportunities for healthcare professionals to improve health system management, patient engagement, budgeting, planning and performing evidence-based decision-making. Literature suggests that certain skills and/or competencies for health professionals working with big data in health care would be required. A review of the skills and competencies in health data analytics required by health professionals is needed to support the development or re-engineering of curriculum for health professionals to ensure they develop the abilities to make evidence-based decisions that ultimately can lead to the effective and efficient functioning of a healthcare system.

Methods Using Arksey and O'Malley's framework, this study will review literature published in English from January 2012 to December 2022. The database search includes Academic Search Complete, CINAHL, and MEDLINE via EBSCOhost, PubMed, Science Direct, Scopus, and Taylor and Francis. The reference lists of key studies will be searched to identify additional appropriate studies to include. The review will be conducted using an inclusion and exclusion criteria. Iterative processes will be involved at the various stages of search strategy piloting, screening and data extraction. Articles will be reviewed through a two-step process (title and abstract, and full-text review) by at least two reviewers. Data will be described quantitatively and/or qualitatively and presented in diagrams and tables.

Ethics and dissemination Ethical clearance has been received, and strict protocol measures will be followed to ensure the data reported is of quality and relevant to the review purpose. The results will be disseminated through a peer-reviewed scientific journal, presentation at national and/or international conferences, and other platforms such as social media (eg, LinkedIn, Twitter), and relevant stakeholders.

INTRODUCTION

Health data analytics and health information systems

Health information management (HIM) is considered a field specialising in the collecting, analysis and careful management of patient health data, and is considered a combination of business, science and

STRENGTHS AND LIMITATIONS OF THIS STUDY

- ⇒ This review will be the first of its kind to review the skills and competencies required for health professionals in health data analytics, on a global scale.
- ⇒ The review will be used in the construction of a health data analytics skills and competencies framework and making recommendations for the development of a health data analytics course.
- ⇒ The search for literature includes a search in various databases to ensure all formats of literature (published, unpublished, commentaries, etc) are covered.
- ⇒ The review is limited to English-only literature, which may bias the evidence.

information technology.¹ Health informatics, which involves the use of technology to improve human health,² is a subset of informatics but a discipline related to HIM.³ HIM is continuously evolving first by incorporating health informatics to now having to address the use of secondary data also known as Big Data and concepts of data science. A data scientist is described as someone involved in the acquisition and interpretation of data for business intelligence thus has to analyse and interpret data into usable information, using innovative mechanisms and sources from available data whereas a data analyst uses data and information for predictive modelling to inform clinical care and as a management tool for healthcare.⁴ A health information manager on the other hand is one who performs similar roles to that of a data scientist and data analyst.⁵ There is a drive for evidence-based decision-making and adding business value that is spearheaded by the fourth Industrial Revolution. A revolution that is positioned in an era of an explosion of data and the increase in health platforms and sources of information accompanied by the increased development and adoption of e-health solutions to address clinical and management challenges.⁶ Stepping into this scene, we see healthcare data analytics



as a methodological approach to the systematic analysis of health data and provides opportunities for health-care professionals to improve health system management, patient engagement, budgeting and planning.^{7,8} It combines real-time data with historical data to predict trends and patterns revealing potential interventions and actions for achieving healthcare goals and outcomes. Healthcare data analytics provides the following benefits to the healthcare system⁹:

- ▶ Identifying important trends in diagnosis, treatment, and continued care to improve the quality of clinical care.
- ▶ Develop insights and forms an evidence base that informs decision-making and planning.
- ▶ Provide information to public health organisations to improve the management of diseases, predict outcomes and allocate human resources for health-related activities.

A health information system is one that allows for the collection and processing of the health data of people and other activities of a healthcare system.¹⁰ In the last 10 years, the impetus has been placed on the implementation of National Health Information Systems. However, though systems and processes have been placed on the production of health information, little attention has been given to the analysis of the data to provide an evidence base for interventions, planning and management. Recently, Big Data analytics has become increasingly important with new tools and technology emerging to capture and make use of the vast information on diseases, patients and procedures that health systems process daily.¹¹ Big data analytics provides the opportunity for healthcare data analytics to attain a holistic understanding of patients. It allows the making of strategic improvements to operations, interventions, and public health analytics, and monitors disease patterns and trends in populations to guide planning and management. Furthermore, it allows for setting priorities for allocating health resources.⁶

Healthcare professionals and health data analytics, skills and competencies

Big data in healthcare refers to electronic health data sets so large and complex that they are difficult (or impossible) to manage with traditional software and/or hardware; nor can they be easily managed with traditional or common data management tools and methods.¹² Big data in healthcare is overwhelming not only because of its volume but also because of the diversity of data types and the speed at which it must be managed.⁸ Furthermore, with health data increasing daily and more health information going digital, specific skills are required to organise, make sense of and manage this data.¹³

Literature suggests that certain skills and/or competencies for health professionals working with big data would be required. A qualitative study by Meyer,¹⁴ using content analysis, highlighted the demand for data scientists in the USA and explored the skills and qualifications required by healthcare data scientist job postings. The

study identified specialist skills that would make a worthy candidate to hire: statistics, applying machine learning techniques, story-telling, delivering actionable results, data-driven problem solving, creating visualisations and working in multidisciplinary teams (to name a few). This study is contrasted with that by Whittaker *et al.*¹⁵ The authors argued that competencies are required among general workers rather than specialist roles, especially in low-income and middle-income countries¹⁵ Competencies at four levels are suggested: service provider level (competencies include—interpretation of data, analysis of data, etc); faculty-based supervisor (competencies include checking quality of data and interpreting results, etc); district/provincial level (competencies include collection and checking quality of data, etc); and regional-national level (competencies include creating data summaries, analysing data, etc).¹⁵

While there is a demand for health professionals with data science skills, there exists a lack of health professionals who possess the necessary skills and competencies, at present, to meet the demand.¹³ Furthermore, a skills training framework for health informatics is still not well defined.¹⁶ The development of skills and competency frameworks can inform the HIS skills and abilities that the health data analytics job market requires, and enables the education and training sectors to align their learning outcomes to meet the demand.¹⁷

Taking the highlighted literature into consideration, it would be of great benefit to understand broadly the skills and competencies that currently exist for health professionals in health data analytics given the demand. A preliminary search of MEDLINE, the Cochrane Database of Systematic Reviews and JBI Evidence Synthesis was conducted and on assessment, it appears there are no current or underway systematic reviews or scoping reviews on the topic.

Purpose and study objective(s)

The objective of this scoping review is to describe health professionals skills and competencies required in health data analytics as detailed in literature across the globe. The purpose of this study is to inform curriculum development of health professional qualifications by identifying the skills and competencies (and any gaps) of health data analytics.

METHODS AND ANALYSIS

This scoping review will use the methodology framework for scoping reviews proposed by Arksey and O'Malley,^{6,18} to scope and map the literature on health data analytic skills and competencies for health professionals. Using the proposed framework will allow for a clear methodological process, that is, reliable and replicable. The review will be reported, and screening guided by the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Extension Statement for Scoping Reviews).¹⁹

Step 1: identifying research questions

Using the population–concept–context (PCC) framework (see online supplemental file I) recommended by the Joanna Briggs Institute (JBI) for scoping reviews, the research question and subquestions for the study are as follows:

Review question

1. What are the skills and competencies required in health data analytics for health professionals?

Subquestions

2. Are there different skills and competencies required in health data analytics for health professionals in high-income versus low-income and middle-income countries?
3. In what geographical contexts have the skills and competencies identified been found?
4. What are the gaps identified in the skills and competencies of health professionals in health data analytics?

Step 2: identifying relevant studies (search strategy)

This stage will involve an iterative process of searching the literature, refining the search strategy and reviewing articles for study inclusion. A systematic search strategy using a combination of keywords, search terms and Boolean operators AND/OR will be developed. The search strategy was developed in consultation with the research team. Using an initial limited search of the title and abstract of articles in a chosen database, the main concepts were used to develop the search string to be tailored for each database (see online supplemental file II). The initial search strategy (see online supplemental file III) will be piloted in the databases and adapted accordingly, to check its suitability to selected databases and keywords. This initial search and refinement will take place from 12 August 2022 and the proposed final review and analysis will take place at the end of June/July 2023. All reviewers will review the piloted search strategy for commentary or edition, and the proposed search strategy will be checked by searching the selected databases and the first 100 search results will be reviewed to ensure validity of the search strategy. The inclusion and exclusion criteria will be refined and agreed on by all four reviewers. The literature search will include published and unpublished literature. The English literature included must be published between January 2012 and December 2022. The chosen databases include Academic search complete, CINAHL, and MEDLINE via EBSCOhost, PubMed, Taylor, and Francis, Science Direct, and Scopus. All the retrieved resources at all phases of the study will be uploaded onto Covidence (via Mendeley) for the screening and eventual charting process.

Step 3: selection of relevant articles

On agreement of the search strategy, an official search of the databases will be conducted by at least two reviewers, independently. Following the search, all identified citations will be collated and uploaded into Covidence (via Mendeley), and duplicates removed. At least two reviewers will screen a random sample of 25 titles/abstracts for assessment against

the inclusion criteria for the review. Once agreement on selected studies reaches 75%, then official screening can begin. The full text of selected citations will be assessed in detail against the inclusion criteria by two or more independent reviewers. Reasons for exclusion of sources that do not meet the inclusion criteria will be recorded and reported. Any disagreements that arise between the reviewers at each stage of the selection process will be resolved through regular meetings and discussions. If consensus cannot be reached, a third reviewer will be invited to mitigate the disagreement. The search will be documented detailing the date, database, keywords and the number of articles retrieved. Searches will continue until saturation is reached and no new studies meeting inclusion and exclusion criteria are identified. A PRISMA diagram (see online supplemental file IV) will be completed to show the numbers of articles identified, screened, eligible and included for full-text review. The reference list of all included sources of evidence will be screened for additional potential studies.

Inclusion criteria: the scoping review will include published and unpublished literature that either uses a qualitative, quantitative or mixed methods approach. Furthermore, the review will include systematic and scoping reviews, opinion papers and commentaries that include discussions on health professionals' skills and competencies related to health data analytics. The literature will not be gender or culture specific. Literature published in English between January 2012 and December 2022, to cover the last 10 years of relevant published literature.

Exclusion criteria: studies will be excluded if the target audience does not include health professionals, that is, the general population, they are not published in English, do not cover the concepts of the review nor are published in the review's time frame.

Patient and public involvement

Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Step 4: charting the data

Charting the data includes data extraction. Continuing with the PCC framework, the draft data extraction form to be used for this study is that provided by the JBI.²⁰ Data will be extracted from 5 to 10 papers included in the scoping review by two or more independent reviewers using the data extraction form, which will be modified and revised as necessary during the process of data extraction from each included evidence source. Modifications will be detailed in the scoping review. Any disagreements that arise between the reviewers will be resolved through discussion, or with an additional reviewer(s). The data extracted will include specific details about the participants, concept, context, study methods and key findings relevant to the review question/s (ie, skills and competencies found, health professionals identified and settings where skills and competencies have been identified). If necessary (and time permits), authors of papers will

be contacted to request missing or additional data, where required.

Step 5: collating, summarising and reporting results

As the breadth of the data is not yet clearly known, it is not possible to predetermine the best method of collating and reporting the study results. However, it is proposed, that the data will first undergo basic descriptive statistics to highlight (numerically) the most presented or identified health data analytics skills and/or competencies for health professionals, and the geographical settings where these skills and competencies are generally most concentrated (vs the least), and how distributed the skills and competencies are across the globe (in terms of income classification of the countries). Furthermore, the data will be tabulated to list the skills and competencies identified in the literature and the contexts where the skills and competencies were identified, followed by a narrative summary of the tabulated results to highlight the patterns, similarities and differences that exist in the data to address the objectives of the study. Critical appraisal for risk of bias assessment is not an objective of this study, nor is it required in scoping reviews,²¹ thus no appraisal will be covered in this review.

ETHICAL CONSIDERATIONS AND DISSEMINATION

Although there are not any specific ethical considerations recommended for scoping reviews, strict protocols for conducting the review will be followed to ensure its replicability, such as ensuring that the data reported in the review are relevant to the review purpose. The results of the review are expected to be disseminated through a peer-reviewed scientific journal, presentations at national and/or international conferences, and relevant stakeholders—other platforms such as social media (eg, LinkedIn, Twitter) will be considered. Possible limitations to the study include missing out on potentially relevant articles due to the use of AND instead of OR combination of ‘competencies’ and ‘skills’. Time constraints influencing the direction and conduct of the study may work to become a possible limitation to this review. Further limitations to the study will be reported and recommendations on gaps identified and spaces for health data analytics curriculum development for health professionals in low-income and middle-income countries will be made in a final report.

Contributors VM and IZ conceived the idea for the scoping review. VM summarised the idea of the protocol, while AN led the design of the protocol and methodology. AN wrote and edited the drafts of the manuscript. AN and VM provided input to the drafts, and all authors (AN, VM, IZ and VB) contributed to development of the search strategy and approval of the final draft for submission.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

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 iD

Akhona Nkwanyana <http://orcid.org/0000-0001-6722-551X>

REFERENCES

- Al-dossary H, Alumran A, Al-rayes S, et al. An overview of health information management education in Saudi Arabia. *Informatics in Medicine Unlocked* 2021;23:100530.
- Veinot TC, Ancker JS, Bakken S. Health informatics and health equity: improving our reach and impact. *J Am Med Inform Assoc* 2019;26:689–95.
- Sherifi D, Ndanga M, Hunt TT, et al. The Symbiotic relationship between health information management and health Informatics: opportunities for growth and collaboration. *Perspect Heal Inf Manag* 2021;18.
- Fenton SH, Low S, Abrams KJ, et al. Health information management: changing with time. *Yearb Med Inform* 2017;26:72–7.
- American Health Information Managers Association. Embracing the future: new times, new opportunities for health information managers. *Embrac Futur New Times, New Oppor Heal Inf Manag / AHIMA, Am Heal Inf Manag Assoc* 2005;307:1211–2.
- Pastorino R, De Vito C, Migliara G, et al. Benefits and challenges of big data in healthcare: an overview of the European initiatives. *Eur J Public Health* 2019;29(Supplement_3):23–7.
- Murdoch TB, Detsky AS. The inevitable application of big data to health care. *JAMA* 2013;309:1351–2.
- Raghupathi W, Raghupathi V. Big data Analytics in healthcare: promise and potential. *Health Inf Sci Syst* 2014;2:1–10.
- Islam M, Hasan M, Wang X, et al. A systematic review on healthcare analytics: application and theoretical perspective of data mining. *Healthcare* 2018;6:54.
- Renee E, Thulani M, Peter Barron AS. 7 health information systems in South Africa SAHR 2011. 2011.
- Razzak MI, Imran M, Xu G. Big data analytics for preventive medicine. *Neural Comput Appl* 2020;32:4417–51.
- Dash S, Shakyawar SK, Sharma M, et al. Big data in healthcare: management, analysis and future prospects. *J Big Data* 2019;6:1–25.
- Ambigavathi M, Sridharan D. Big data Analytics in Healthcare. In: *2018 Tenth International Conference on Advanced Computing (ICoAC)*. Institute of Electrical and Electronics Engineers Inc, 2018: 269–76.
- Meyer MA. Healthcare data scientist qualifications, skills, and job focus: a content analysis of job postings. *Journal of the American Medical Informatics Association* 2019;26:383–91.
- Whittaker M, Hodge N, Mares RE, et al. Preparing for the data revolution: identifying minimum health information competencies among the health workforce. *Hum Resour Health* 2015;13.
- Sapci AH, Sapci HA. Teaching hands-on informatics skills to future health informaticians: a competency framework proposal and analysis of health care informatics curricula. *JMIR Med Inform* 2020;8:e15748.
- Whittaker M, Hodge N, Mares RE, et al. Preparing for the data revolution: identifying minimum health information Competencies among the health workforce. *Hum Resour Health* 2015;13:17.
- Arksey H, O'Malley L. Scoping studies: towards a methodological framework. *International Journal of Social Research Methodology* 2005;8:19–32.
- Tricco AC, Lillie E, Zarin W, et al. PRISMA extension for scoping reviews (PRISMA-SCR): checklist and explanation. *Ann Intern Med* 2018;169:467–73.



- 20 Joanna Briggs Institute. JBI manual for evidence synthesis. In: *JBI*. 2020.
- 21 Pollock D, Davies EL, Peters MDJ, *et al*. Undertaking a Scoping review: A practical guide for nursing and Midwifery students, Clinicians, researchers, and academics. *J Adv Nurs* 2021;77:2102–13.