Reaching consensus on essential health data science skills and knowledge gaps using a validated 'crowd consensus' methodology

Global Investigators
Prof Trudie Lang, Director of TGHN and HDR Global Director of Research
Ryan Walker, TGHN;
Sinead Whitty, TGHN Head of Training, Teaching and Career Development;
Arancha De La Horra Gozalo, Clinical Research Specialist, TGHN;
Anne Wozencraft, Director of Partnerships and Strategic Delivery, HDR Global;
Matt Retford, Technical and Programme Manager, HDR Global;
Sally Boylan, Capacity Development and Engagement Manager, HDR Global.

Regional Investigators
Dr Elvis Temfack, Senior research officer, Africa CDC
Dr Aliya Naheed, MBBS (DU), MPH (JHU), PhD (TU), Post doc (UoM), Scientist & Health Systems Specialist icddr,b;
Vanessa Jorge, Information and Communication Coordinator, Fiocruz;
Frank Kagoro, TGHN Africa regional manager;
Mercedes Rumi, TGHN Latin America and Caribbean regional manager;
Salvia Zeeshan, TGHN Asia regional manager;
Flavia Bueno, Fiocruz Hub Coordinator;
Larissa Pruner Marques, Data Science Coordinator, Fiocruz;
Saimul Islam, Data Science Coordinator icddr,b;

1 Background and Rationale
Developing training and resources to build data science capability is critical. The development of new expertise in data science and the skills to overcome the well-recognised barriers that limit the collecting, sharing, analysis and use of high-quality data, are essential to addressing the current inequalities encountered by health research teams in low-resource settings.

There is a need to determine where the barriers and enablers to engagement in data science lie and so by working with a large, global community of practice we will identify, by consensus and via an evidence-based approach, what are considered essential health data science skills. A data science curriculum can be built around those consensus themes. Resources, tools and training materials can be identified, developed or redeployed taking account of what already exists and what has proven effective. Mechanisms for dissemination of these also need to be determined.
2 Aims and objectives

To identify, by consensus, essential health data science skills to develop data science skills curriculum, pragmatic and accessible tools, training and resources.

3 Methods

The adapted Delphi process for forming a ‘crowd’ consensus

The Delphi technique is a consensus building method that gathers experts to discuss issues (Developing a Consensus Led Essential Research Skills Training Curriculum, Arancha De La Horra Gozalo et al, 2022). The process goes through a series of cycles. In each cycle, a panel of experts is presented with a set of statements to rate. The Delphi process typically starts with a round of open-ended questions to form the basis of the initial pool of topics to rate. Controlled feedback is then given that shows how each individual’s ratings compare with the whole group.

The Global Health Network has conducted several studies to reach a consensus using an adapted version of the Delphi approach (Developing a Consensus Led Essential Research Skills Training Curriculum, Arancha De La Horra et al, 2022; and COVID-19 Research Priorities Identified by the Global Research Community, Trudie Lang et al, July 2020). For example, one of these studies aimed to identify essential health research training skills gaps across Africa, Asia and Latin America in a three-stage approach:

- conducted a research training gap analysis reviewing responses of 7176 participants from 153 countries collected from 2017 to 2019, some of these were in-person workshops and some online. This generated a list of 98 research-training themes;
- a two-round Delphi study was carried out online to prioritise the outcomes of the gap analysis and find consensus on what constituted the minimum set of research skills;
- conducted a Stakeholder Review and Implementation workshops online which reviewed the global applicability of the findings and how they could be most effectively implemented.

The process resulted in an evidence-led research skills training curriculum formed by 108 themes mapped into 13 modules that any research group could use to implement and guide research training programmes. This work is now a formal TDR report.

The same Delphi approach was also used to assess the relevance of the WHO blueprint during COVID-19 (COVID-19 Research Priorities Identified by the Global Research Community, Trudie Lang et al, 15 July 2020) and more recently to determine whether a community of practice can support better pharmacovigilance across Africa.

These examples validated this approach of asking a large community of practice for their view, and this can be collected online or in person. These data can then be combined to form the themes and also to serve as stages of review, as per the Delphi approach. This ‘crowd’ approach adds the important step to the standard Delphi of returning to the community of practice, the participants, in forming this consensus to take the findings back to them and ask them for their views on how to implement and take forward the findings.

4 Process stages

All stages of the process will be carried out in collaboration with a cross-regional working group.
4.1 Setting the themes

- Landscaping and knowledge gap analysis (KGA) through a traditional and web-based literature review and stakeholder interviews and survey. A review of the results of the TGHN Essential Health Research skills Delphi to identify themes relevant to data science will also be undertaken at this stage; Working group discussions to review the literature to decide on the themes to be progressed into the crowd Delphi review.

4.2 Crowd review to achieve consensus

Survey 1

- The survey will be defined by the themes and can be disseminated through many approaches; as an online survey, but also given out at regional data science or research capacity building workshops, conferences, training sessions;
- Respondents will identify their top 3 priorities under each theme and rank them. This will generate a scoring 1-3. Essential will be defined as any theme reaching 75% consensus or above.

Survey 2

- The questions will be determined by the results of stage one to address areas where there is not a consensus, new themes are identified or there is a lack of clarity on the theme. The survey can be disseminated as an online survey but also given out at regional data science or research capacity building workshops, conferences, training sessions;
- Respondents will be asked to score each theme for relevance as either ‘essential’ or not ‘essential’ and we will also record no response. Themes scoring 75% or more for relevance will be included;
- Respondents will be asked to score each theme for clarity either ‘clear’ or ‘unclear’. We will also record no response. Themes scoring 80% or more for clarity will be included.

4.3 Stakeholder review

- Webinars and in person review of the results - this step will be undertaken to validate the themes and framework. The specific aims of stakeholder review are to:
  - consider the suitability of the themes and groupings as an accurate reflection of the content;
  - evaluate the applicability of the proposed Essential Global Health Data Science Skills Curriculum findings to the global research community.
- The stakeholder review will conclude the process of developing an essential global health data science skills curriculum framework based on consensus findings.

4.4 Choosing the panel

Led by working groups:

- Develop criteria to be met by panellists ensuring a diverse range of stakeholders and taking account of different geographical contexts and global health data science perspectives
- Define recruitment process for panellists (likely to be a combination of open and targeted recruitment with regional partners identifying key collaborators).

4.5 Implementation of consensus
Seeks guidance from the community on how to disseminate the curriculum which could include sessions such as:

- Promotional webinar and via TGHN Regional Hubs and the Global Health Data Science Hub and TGHN broader network. Investigators from Fiocruz, icddr,b and Africa CDC will lead the promotion and identification of key collaborators to be invited to participate in their region.
- **To note:** importance of communicating that this process aims to identify the *essential* data science skills, i.e. the minimum skills required for someone with limited experience to be able to conduct quality data science (not *all or advanced* data science skills).

4.6 Ensuring equity

- Language considerations (ideally, we will deliver the surveys in 4 languages - English, French, Spanish and Portuguese - but the cost and time requirement will need to be taken into account);
- Led by working groups with global representation, identify broad but appropriate criteria for panelists;
- Transparency (we will publish the process and associated documents on TGHN Regional Hubs and Global Health Data Science Hub).

5 Survey platform

- Survey Monkey for the Landscaping and Knowledge Gap Analysis and 2 crowd consensus surveys.

6 Data management and analysis

6.1 GDPR compliance and consent

- Anonymised responses – need to consider response storage and anonymisation;
- Password protection;
- Restricted access.

6.2 Analysis

- Analysis to be undertaken by a small team from across all 5 partners;
- NVivo will support qualitative analysis;
- Results to be reviewed by the working groups and investigators ahead of stakeholder review webinars and/or meetings.

7 Results

- Shared via stakeholder review webinars;
- Published following implementation webinars, led by investigators in each region.

8 Outputs

- An essential data science curriculum *framework* based on consensus which meets the needs of the global health data science community in low- middle-income settings.