A Case for Action: India's National Family Health Survey Datasets Await Exploration of Big Data Applications Toward Evidence-Informed Public Health Decision-Making to Tackle Malnutrition

Sir,

Malnutrition, especially undernutrition, remains a key public health challenge having intergenerational and macroeconomic consequences. The government envisages evidence-based implementation strategies which can be updated as new evidence emerges related to health, nutrition, and development.[1] Recent National Family Health Surveys (NFHS-4 & 5) allow district-level insights into the prevalence of key undernutrition indicators. However, to enable data-driven/evidence-based interventions, it requires complex analysis to identify and prioritize underlying determinants at the district-level. Given the constraints of India's public health system, participation from for-profit and nonprofit organizations is inevitable. The lack of statistical expertise with local health service organizations, academia and district health offices, deprives leveraging NFHS data for planning evidence-based, contextually appropriate interventions.

Hence, there is a perceived need for design and implementation of open-access/user-friendly online portal for advanced statistical analysis and easy-to-comprehend visualization of key health indicators and their determinants at district/state/national level using data analytics, machine learning, and data visualization packages. Our conceptualization attempts to build a case that how the proposed open-access portal for NFHS datasets could enable local public health stakeholders to initiate evidence-informed interventions and citizen's participation in public health dialogue.

A review of the literature studying NFHS-4 data helps enlist various determinants of child undernutrition. A study analyzing NFHS-4 data for regional prevalence and determinants of exclusive breastfeeding, categorized key determinants into child (sex, birth weight, birth order, etc.), maternal (age at marriage, height, education, etc.), household (family size, wealth quintile, etc.), health service (mode of delivery and type of delivery assistance), and community factors (urban/rural and district of residence).[2] A technological approach investigating the nutritional status of children using NFHS-3 (2005–06) data found machine learning technique identifies some important features not identified in extant literature with sensitivity of ~70%–80%.[3] Such efforts provide building block for exploring opportunities of using machine learning approaches for identifying possible correlates/determinants of undernutrition for specific districts. If we could provide determinants of stunting, wasting, severe wasting, and underweight in different districts through an open-access portal without the difficulties of data cleaning, sorting, and analysis; it may help local authorities and organizations to develop and implement data-driven, evidence-informed and targeted nutrition interventions. Advances in big data ecosystem, data collation, data analytics, artificial neural network, machine learning, and data visualization allow tremendous opportunities to automate district-level analysis for key health indices.[4]

Our conceptualization is about designing an open-access web-portal which can automate prevalence estimation and risk factor identification for specific health indicators for the selected district/state level using NFHS-4 and future datasets. The use of advanced visual analytics applications can demonstrate easy-to-comprehend analysis from available data for different end-users, who may be medical and public health professionals at the district/state/national level, program planners, policymakers, etc. End-users can understand the risk factors for their given districts/states without the knowhow of statistics and can use this information for the design of data-driven, evidence-based, and context-specific interventions.

An open-source programming language can be used to develop a web portal and predictive machine learning model. The use of Python programming language for building a machine learning model can help us to reduce the overall project cost. An accurate undernutrition predictive machine learning model can be investigated as a promising avenue to identify heterogeneity in prevalence and determinants of child undernutrition across districts/states. Such a modeling may not require parametric assumptions, which can reduce the risk of model misspecification, and it can capture effects nonparametrically while downgrading variables that are less important in predicting the outcome. The proposed portal can democratize access to NFHS data analysis and understanding key determinants at district-/state-level and compare inter-district variations.

The open-access portal may allow public dissemination of scientific findings from big data collected through public financing. This will open up newer avenues for public dialog on critical public health indicators, public scrutiny of big data, foster critical thinking and social innovations at all levels, and most importantly stem the flow of scientific misinformation in

the media. This, in turn, may foster citizen contribution to big data economy and realizing its analytical findings for greater public good.[5] Open access, to not just big data but big data analytics and easy-to-understand visual analytics, can also empower citizens to charter political dialog around crucial public health issues and draw attention of local politicians and legislative representatives on such health indicators.[6] There is an attempt to represent the burden of malnutrition among under-5 children by Loksabha constituencies to bring more political accountability, and authors have shared methodology to enable scientific community to explore indicators and its determinants by political constituencies.[7] This study also recommended understanding the burden of key health indicators by monitoring legislative assembly level constituencies, where the attention of local politicians can be brought to public health status in their constituency.[7]

Hence, it is imperative upon academia and professional associations representing public health and/or community medicine to facilitate consortium and bring together experts from public health/community medicine and computer sciences and data sciences. These experts can brainstorm possible collaboration opportunities for interdisciplinary research and identify key public health domains (such as malnutrition and noncommunicable diseases) for technology-led research and development. Master's in Public Health, M. D. in Community Medicine, and Ph. D. in Public Health/Community Medicine candidates shall be encouraged to undertake joint projects, dissertations, or research work together with data/computer science experts to design technological platforms, which could help identify key social determinants of health and predict future trends based on available big data such as NFHS. Going forward, such platforms could be programmed as an expert system to suggest context-specific interventions and guide uniform data collection for tracking intervention fidelity, its evaluation and real-time decision support. Advances in data analysis can be leveraged with AI applications for evidence-based decision-making and problem-solving by government, nongovernment, and private health service providers. Public leadership in such a platform may facilitate data collection, aggregation, and validation from different sources and guide future strategies for NFHS based on emerging trends.[8] Such a platform could also have global health implications and could help evidence-informed decision-making in many countries, which participate in the Demographic and Health Surveys Program.

Financial support and sponsorship

Conflicts of interest

There are no conflicts of interest.

Kandarp Narendra Talati[1,2], Geetika D Madan-Patel[3], Rajiv Kumar Gurjwar[4], Arvind R Yadav[5]

[1]Department of Interdisciplinary Research, Foundation for Diffusion of Innovations, [2]Center of Research for Development, Parul University, [3]Department of Community Medicine, Parul Institute of Medical Sciences and Research, Parul University, [4]Department of Computer Science Engineering, Parul Institute of Technology, Parul University, [5]Department of Electronics and Communications Engineering, Parul Institute of Engineering and Technology, Parul University, Vadodara, Gujarat, India

Address for correspondence: Mr. Kandarp Narendra Talati, #106, Center of Research for Development, Parul Institute of Medical Sciences and Research, Parul University, Limda, Vadodara - 391 760, Gujarat, India. E-mail: kandarp.talati@paruluniversity.ac.in

References

1. National Institution for Transforming India (NITI) Aayog. Nourishing India: National Nutrition Strategy, Government of India. New Delhi: NITI Aayog; 2017.
2. Ogbo FA, Dhami MV, Awosemo AO, Olusanya BO, Olusanya J, Osuagwu UL, *et al.* Regional prevalence and determinants of exclusive breastfeeding in India. Int Breastfeed J 2019;14:20.
3. Khare S, Kavyashree S, Gupta D, Jyotishi A. Investigation of nutritional status of children based on machine learning techniques using Indian demographic and health survey data. Procedia Comput Sci 2017;115:338-49.
4. Raghupathi W, Raghupathi V. Big data analytics in healthcare: Promise and potential. Health Inf Sci Syst 2014;2:3.
5. Eagleman DM. Why public dissemination of science matters: A manifesto. J Neurosci 2013;33:12147-9.
6. Scheufele DA. Science communication as political communication. Proc Natl Acad Sci U S A 2014;111 Suppl 4:13585-92.
7. Khan J, Mohanty SK. Spatial heterogeneity and correlates of child malnutrition in districts of India. BMC Public Health 2018;18:1027.
8. Talati K, Madan-Patel G, Swaminarayan P. India's national family health survey (2015-16) big data awaits advanced analytics and artificial intelligence applications towards strengthening public health outreach and outcomes. Isr J Health Policy Res 2019.

Access this article online

Quick Response Code:

Website:
www.ijcm.org.in

DOI:
10.4103/ijcm.ijcm_698_21