EBioMedicine 74 (2021) 103740

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.elsevier.com/locate/ebiom

Commentary

Using big data for risk stratification of childhood pneumonia in low-income and middle-income countries (LMICs): Challenges and opportunities

Maheen Sheikh, Fyezah Jehan*

Department of Pediatrics and Child Health, Aga Khan University, Karachi, 74800, Pakistan

ARTICLE INFO

Article History:
Received 15 November 2021
Accepted 24 November 2021
Available online 13 December 2021

Pneumonia is a major cause of death in children younger than five years of age, accounting for 15% of all deaths in this age range worldwide[1]. The burden of disease has shown a steady decline, yet inequities still drive mortality linked to air pollution, low immunisation coverage, low socioeconomic settings, malnutrition, and poor water sanitation and hygiene (WASH) in low-income and middle-income countries (LMICs) [2].

The World Health Organization (WHO) categorises pneumonia into mild features (with tachypnoea or retraction of the chest wall) and severe with danger signs (stridor, hypoxia, inability to feed, persistent vomiting, convulsions, and altered consciousness) [3]. Key problems in the current evaluation in LMICs come from the lack of accessibility to, and atypical presentation on, chest X-rays, along with the lack of a gold-standard test to determine etiology. Even in circumstances where patient samples can be sent for testing, *Streptococcus pneumoniae,* one of the most common culprit organisms of pneumonia and a fastidious bacterium, is very difficult to isolate on routine culture. Newborns may present with poor feeding, irritability, and grunting, while infants may experience cough, tachypnoea, hypoxemia, congestion, and fever [4]. Constitutional symptoms in adolescents, such as headache, pleuritic chest pain, and non-specific abdominal pain, all denote the varying presentations clinicians must consider.

Algorithmic evaluations recommended by the WHO, such as those listed above, are relevant for children up to 5 years of age, however, limited evidence-based data exists for adolescents. Arbitrary measures for diagnosis including respiratory rate, disregard consideration of alternative diagnoses such as congenital cardiac illness or concurrent fever as a reason for tachypnoea. Beyond diagnosis, identifying those at risk of dying has been the subject of several investigations, but the clinical prediction indices tested failed to perform when externally applied [5], which may be due to the lack of incorporation of contextual risk factors, such as prematurity, air pollution, malnutrition, WASH and sociodemographic status.

In this milieu, employing big data and machine learning technologies to incorporate diverse clinical, sociodemographic, and other factors have a very promising scope in LMICs. It could give way to the identification of patients who are at high-risk of developing severe disease and mortality, and gauge the need for in-hospital admission and management. Previous work on the use of deep learning to identify pneumonia in chest X-rays showed 96% to 98% accuracy in pneumonia detection [6,7], and this strategy could be valuable for patients in remote areas where low-cost chest X-ray machines could be deployed. This may potentially lead to task shifting for community and other healthcare workers, who could instead focus on counselling and management of those requiring hospitalisation, resulting in curtailment in unnecessary referral and an overall reduction in healthcare costs.

The challenge of utilising big data in LMICs lies in the six Vs of big data: volume, variety, velocity, value, veracity, and variability. While there is ample systematically collected data in LMICs through the establishment of global child health surveillance networks [8] and years of clinical trial research in pneumonia in children that is accurate and reliable [9], the datasets lack variety and variability because homogeneity is desired in such experiments by design. Having quality datasets from prospective registries in programme settings and healthcare facilities, using electronic health records and/or population-level data repositories is called for. Associated with collection are the inherent ethical challenges, such as privacy, security and third-party access, for which governance is generally lacking in these settings. Beyond data availability and collection, subsequent phases of the data science pipeline, such as data organisation, analysis, interpretation, and application, are all key aspects. In resource-limited settings, the absence of information technology and data specialists, poor infrastructure, and limited technical capacity all add to the obstacles faced in employing big data to its full potential in the settings where it is most urgently needed.

Declaration of Competing Interest

The authors declare no conflicts of interest.

* Corresponding author at: Aga Khan University, Department of Pediatrics and Child Health, STADIUM ROAD, Karachi 74800, Pakistan.
E-mail address: fyezah.jehan@aku.edu (F. Jehan).

https://doi.org/10.1016/j.ebiom.2021.103740

References

[1] Marangu D, Zar HJ. Childhood pneumonia in low-and-middle-income countries: An update. Paediatr Respir Rev 2019 Nov;32:3–9.
[2] Sultana M, Alam NH, Ali N, Faruque ASG, Fuchs GJ, Gyr N, et al. Household economic burden of childhood severe pneumonia in Bangladesh: a cost-of-illness study. Arch Dis Child 2021 Jun 1;106(6):539–46.
[3] Organization WH. Revised WHO classification and treatment of pneumonia in children at health facilities: evidence summaries. 2014;
[4] Scott JAG, Wonodi C, Moïsi JC, Deloria-Knoll M, DeLuca AN, Karron RA, et al. The Definition of Pneumonia, the Assessment of Severity, and Clinical Standardization in the Pneumonia Etiology Research for Child Health Study. Clin Infect Dis 2012 Apr 1;54(suppl_2):S109–16.
[5] Rees CA, Hooli S, King C, McCollum ED, Colbourn T, Lufesi N, et al. External validation of the RISC, RISC-Malawi, and PERCH clinical prediction rules to identify risk of death in children hospitalized with pneumonia. J Glob Health 2021 Oct 9;11:04062.
[6] Almaslukh B. A lightweight deep learning-based pneumonia detection approach for energy-efficient medical systems. Wireless Commun Mobile Comput 2021 Apr 22;2021:e5556635.
[7] Chagas JVS das, de A. Rodrigues D, Ivo RF, Hassan MM, de Albuquerque VHC, Filho PPR. A new approach for the detection of pneumonia in children using CXR images based on an real-time IoT system. J Real-Time Image Proc 2021 Aug 1;18(4):1099–114.
[8] Taylor AW, Blau DM, Bassat Q, Onyango D, Kotloff KL, Arifeen SE, et al. Initial findings from a novel population-based child mortality surveillance approach: a descriptive study. Lancet Global Health 2020 Jul 1;8(7):e909–19.
[9] Hooli S, Colbourn T, Lufesi N, Costello A, Nambiar B, Thammasitboon S, et al. Predicting Hospitalised Paediatric Pneumonia Mortality Risk: An External Validation of RISC and mRISC, and Local Tool Development (RISC-Malawi) from Malawi. PLoS One 2016;11(12):e0168126.