

A data quality assessment and estimation of infant and child mortality among Palestinian refugees

Selin Köksal, Oona Campbell, Zeina Jamaluddine, Akihiro Seita and José Manuel Aburto

Abstract

Violent conflict disproportionately affects vulnerable populations, particularly children, yet conflict-affected populations remain understudied due to challenging living circumstances that make traditional data collection difficult. Electronic data linkages offer a cost-effective alternative to traditional birth cohort studies for examining health outcomes in these populations, though the quality of such data in conflict settings requires careful assessment. This study conducts a comprehensive data quality assessment and mortality estimation using a unique linked dataset of nearly 1 million Palestinian refugee children born between December 2001 and December 2020 across United Nations Relief and Works Agency (UNRWA) service settings in the Gaza Strip, Jordan, Lebanon, Syria, and the West Bank. We systematically assess key quality indicators including age heaping, undercounting and conduct survival analysis to further evaluate data quality by various sociodemographic subgroups. We will then estimate infant and child mortality rates, examining variations across geographic settings, time periods, and by key demographic characteristics. The findings will inform future research using administrative data in conflict settings and contribute to knowledge on child health outcomes among hard-to-reach populations.

Introduction

Violent conflict has extensive repercussions that extend beyond direct fatalities, influencing well-being and human capital in many ways (Vesco et al., 2025). The impacts of conflict, and ensuing economic and food-security crises, can disproportionately affect refugees and other vulnerable populations, especially children (Bendavid et al., 2021; Jürges et al., 2022; Kadir et al., 2019). However, the long-term effects of conflict exposure among refugee populations remain understudied partly due to their living circumstances which make them difficult to reach and collect data from.

Some of us (Jamaluddine et al. (2023)) built a unique birth cohort dataset of Palestinian refugee children drawing on the United Nations Relief and Works Agency for Palestine Refugees in the Near East (UNRWA) electronic health records. UNRWA provides Palestinian refugees with free primary health and elementary-school services, the use of which is recorded in individual-based electronic databases. By linking mothers' obstetric records with child health and education records, we constructed an extremely rich dataset containing nearly 1 million Palestinian refugee children who were born between January 2010 and December 2020 and living in five settings: Gaza, Jordan, Lebanon, Syria, and the West Bank. Linkage between electronic records (e.g. birth, healthcare, education records) offers an efficient and cost-effective alternative to traditional longitudinal birth cohort studies (Jamaluddine et al., 2023). This approach brings together different sources of information enabling a comprehensive dataset to investigate complex associations and establish causal links for policy-relevant research

While electronic data linkages provide crucial resources, especially for studying hard-to-reach populations, they often come with important challenges. In general, electronic data linkages

may include errors, unclear notation or categorization, missing data, and duplicates (Jamaluddine et al., 2023). Specific to the Palestinian refugee children birth cohort, UNRWA has a civil and vital registration system which has been assessed and found to be of poor quality and to underreport deaths. For instance, we have identified that child deaths, especially very early, neonatal deaths, were not recorded in health records and are missing from the system. When newborns die early, they are not brought in to use child health services so are not recorded via health records. Overall, the child health records appear to miss around 80% of child deaths, most of which occur on the obstetric records, and most of which occur very early. To capture these neonatal and infant deaths, Jamaluddine et al. (2023) relied on obstetric records where complications during delivery and mortality reason and timing for the infant are indicated. However, these data are often in free text and it is necessary to assess data quality and validity. While this is a unique, and important dataset examining key questions on the impact of conflict and many other risk factors on Palestinian refugee children understanding and quantifying these data quality issues is essential for producing high-quality and reliable research on this population.

In this paper, we conduct a comprehensive data quality assessment of the UNRWA birth cohort dataset to characterize the nature and extent of data quality challenges and their implications for research using this resource. Our objectives are (i) assessing the quality of infant and child mortality data across age, gender, region and other sociodemographic groups (ii) apply and develop techniques for a systematic data quality assessment (iii) provide reliable infant and child mortality estimates following quality checks.

Methodology

Using these mortality indicators available at the dataset, we first revisited the initial data cleaning process by Jamaluddine et al. (2023). Subsequently, we plotted several descriptive graphs showing the distribution of neonatal and infant deaths by gender, field (Gaza, Jordan, Lebanon, Syria or West Bank) and mother's education level and level of detail of reporting of the age at death. This allowed us to detect where there is any heaping in reported age (day or month) at death.

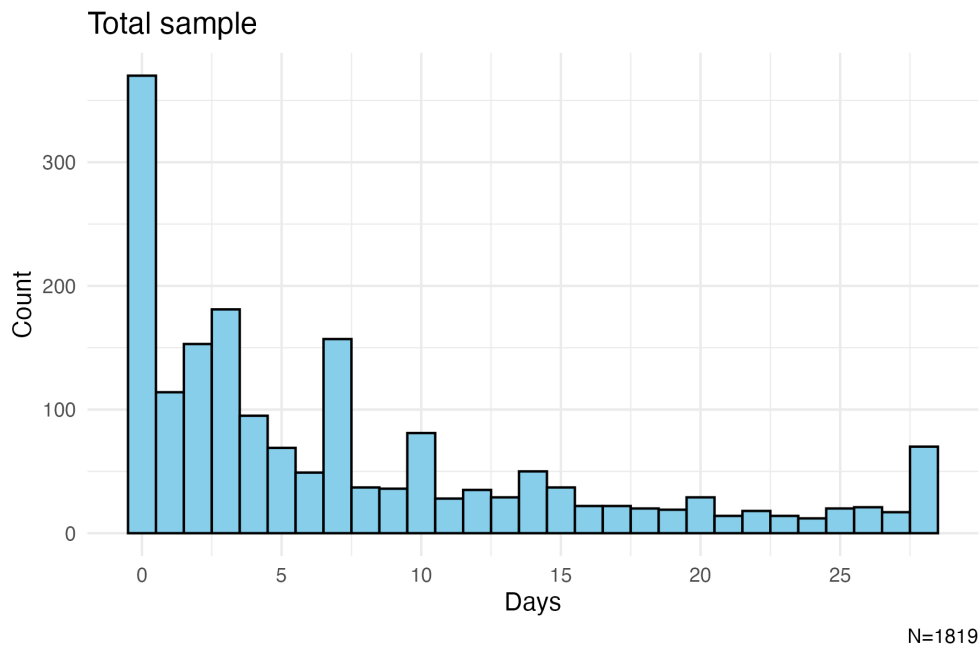
In future steps we will conduct a comprehensive analysis of data quality (e.g. Whipple index), undercounting, and a series of demographic analyses (e.g., survival analysis) to further assess data quality. We will develop strategies to estimate age-specific risk of child death (month-specific risk of infant death and day-specific risk of neonatal death where data permit) and examine whether these rates differ by gender, field and mother's education using Bayesian hierarchical models. We will compare these risk estimates to existing figures in the literature and/or with other data sources for Palestinian refugees (e.g., Multiple Indicator Cluster Surveys (MICS)).

In doing so, we will identify the subpopulations at risk of under- or misreporting by assessing whether our estimates of neonatal, infant and child mortality vary by socioeconomic and demographic indicators including maternal age, maternal education, birthweight, parity and region/field. Finally, we describe levels of neonatal, infant and child mortality by field and over time, examining maternal age, education and region as risk factors for mortality as well as assessing changes in data quality over time.

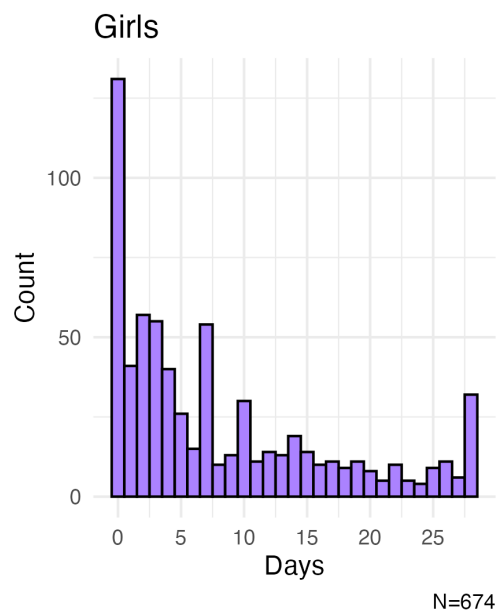
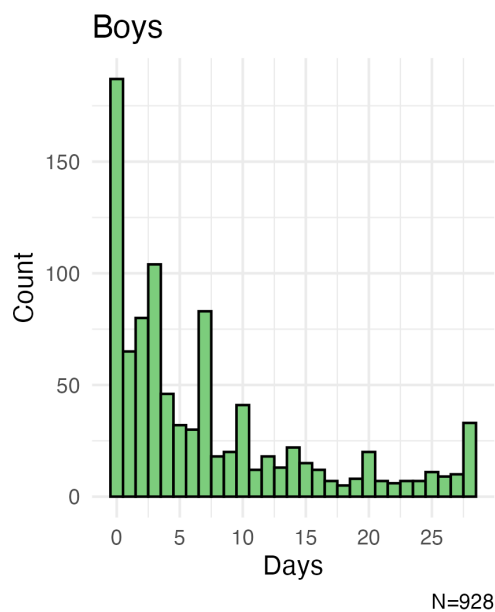
Preliminary findings

Below we present preliminary graphs summarizing the reported day of death for neonatal deaths and reported month of death for infant deaths, disaggregated by key sociodemographic indicators including gender, field, and mother's education level.

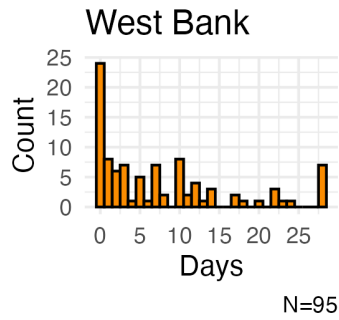
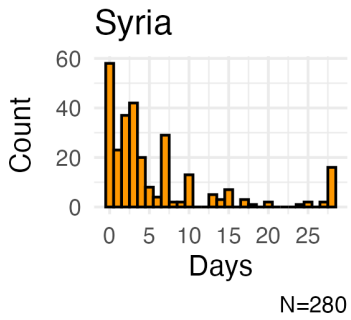
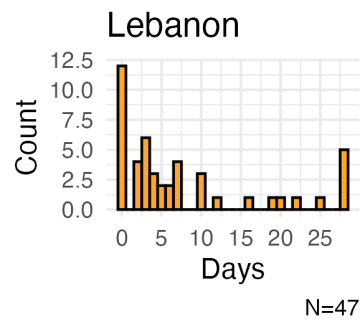
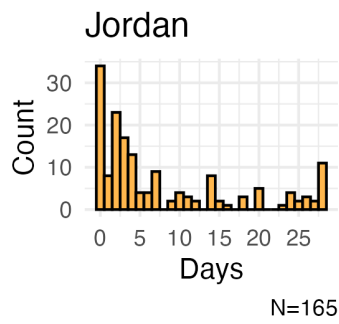
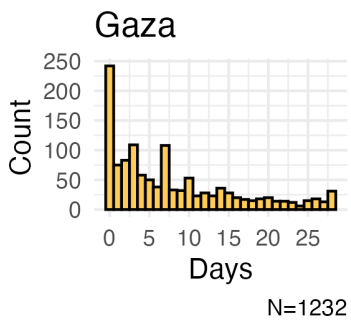
Neonatal deaths



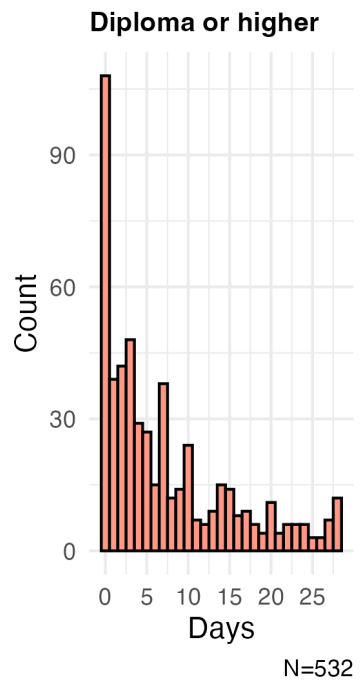
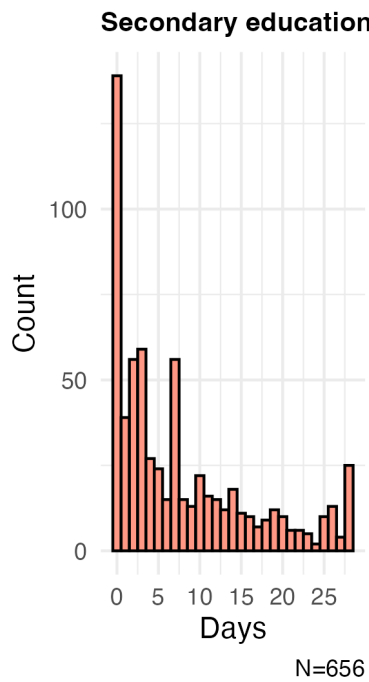
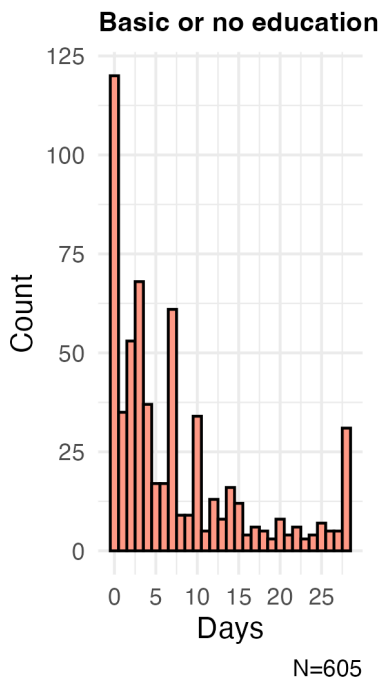
Neonatal deaths by gender



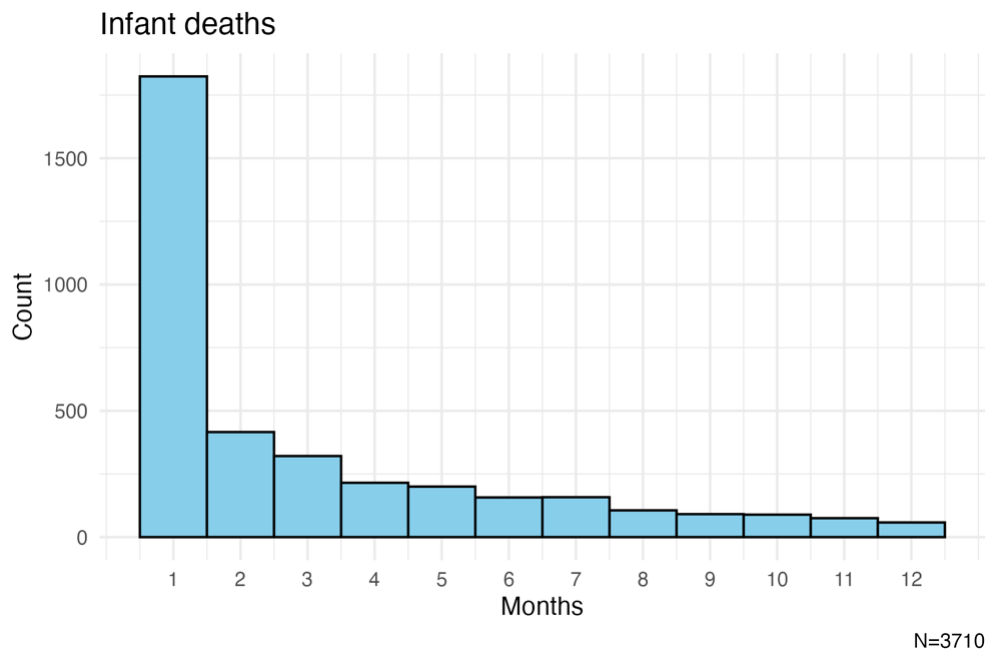
Neonatal deaths by field



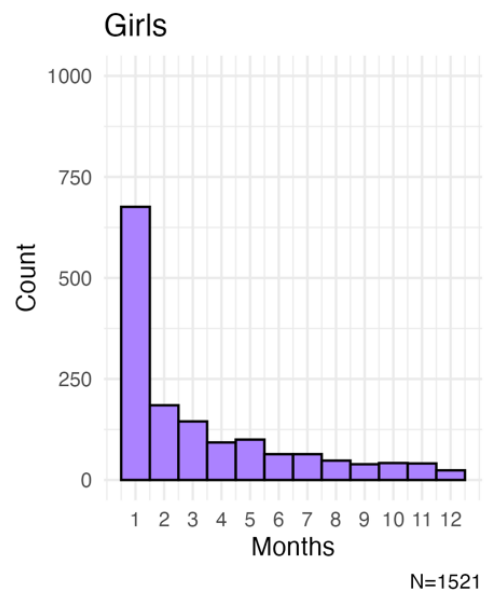
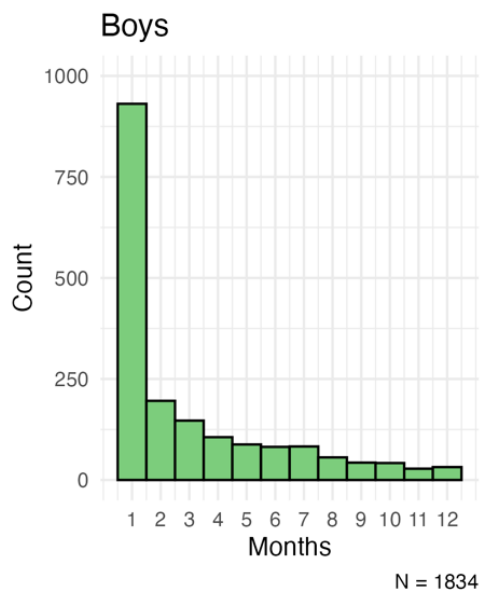
Neonatal deaths by mother's education



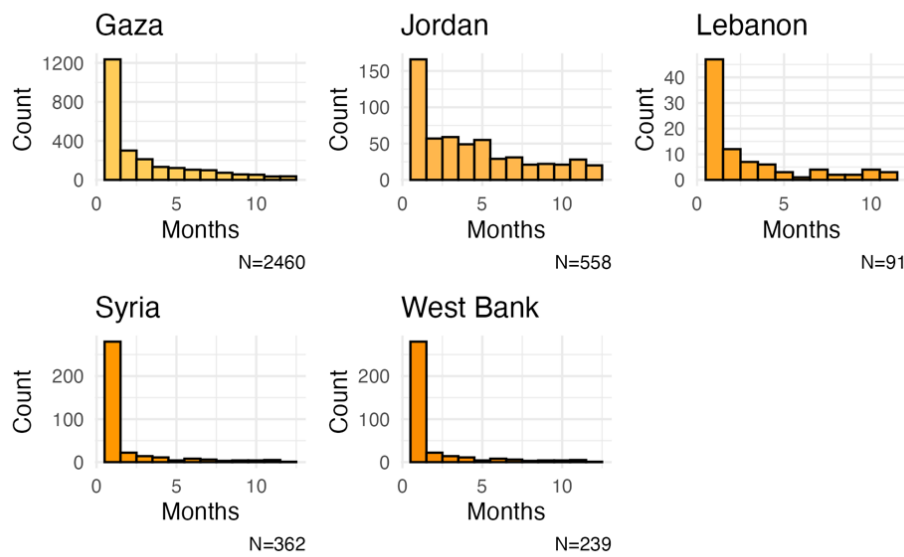
Infant deaths



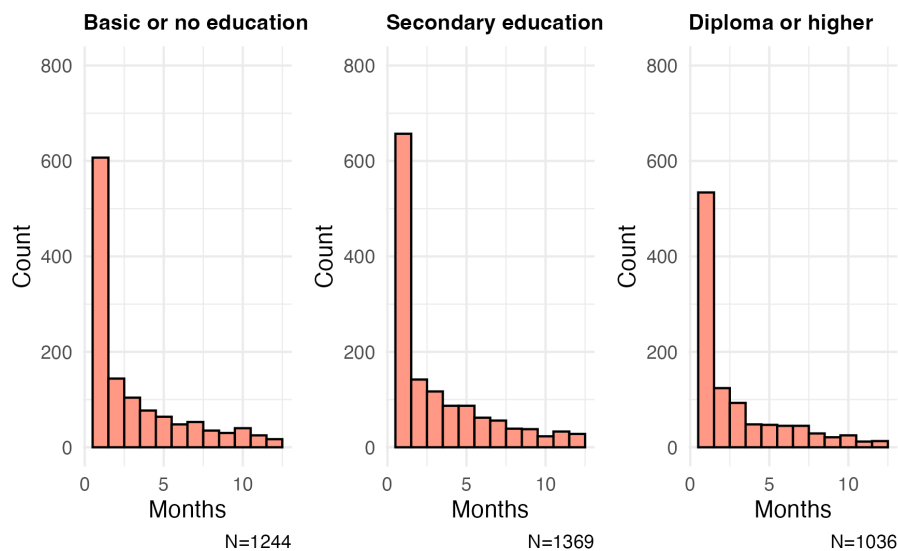
Infant deaths



Infant deaths by field



Infant deaths by mother's education



References

- Bendavid, E., Boerma, T., Akseer, N., Langer, A., Malembaka, E. B., Okiro, E. A., Wise, P. H., Heft-Neal, S., Black, R. E., Bhutta, Z. A., Bhutta, Z., Black, R., Blanchet, K., Boerma, T., Gaffey, M., Langer, A., Spiegel, P., Waldman, R., & Wise, P. (2021). The effects of armed conflict on the health of women and children. *The Lancet*, 397(10273), 522–532. [https://doi.org/10.1016/S0140-6736\(21\)00131-8](https://doi.org/10.1016/S0140-6736(21)00131-8)

- Jamaluddine, Z., Seita, A., Ballout, G., Al-Fudoli, H., Paolucci, G., Albaik, S., Ibrahim, R., Sato, M., Ghattas, H., & Campbell, O. M. (2023). Establishment of a birth-to-education cohort of 1 million Palestinian refugees using electronic medical records and electronic education records. *International Journal of Population Data Science*, 8(1), 2156.
- Jürges, H., Stella, L., Hallaq, S., & Schwarz, A. (2022). Cohort at risk: Long-term consequences of conflict for child school achievement. *Journal of Population Economics*, 35(1), 1–43.
- Kadir, A., Shenoda, S., & Goldhagen, J. (2019). Effects of armed conflict on child health and development: A systematic review. *PloS One*, 14(1), e0210071.
- Vesco, P., Baliki, G., Brück, T., Döring, S., Eriksson, A., Fjelde, H., Guha-Sapir, D., Hall, J., Knutsen, C. H., Leis, M. R., Mueller, H., Rauh, C., Rudolfsen, I., Swain, A., Timlick, A., Vassiliou, P. T. B., Von Schreeb, J., Von Uexkull, N., & Hegre, H. (2025). The impacts of armed conflict on human development: A review of the literature. *World Development*, 187, 106806. <https://doi.org/10.1016/j.worlddev.2024.106806>