BMJ Public Health

Pooling data (or NOT): an empirical examination of the relationship between maternal education and child morbidity in Central Africa using Demographic and Health Surveys

Zacharie Tsala Dimbuene ,[1] Elizabeth T Nyirenda,[2] Philippe Kabore ,[3] Janet Kalinga ,[4] Negussie Boti Sidamo ,[5] Francisca Achaliwie [6]

To cite: Tsala Dimbuene Z, Nyirenda ET, Kabore P, *et al*. Pooling data (or NOT): an empirical examination of the relationship between maternal education and child morbidity in Central Africa using Demographic and Health Surveys. *BMJ Public Health* 2026;**4**:e004892. doi:10.1136/bmjph-2026-004892

► Additional supplemental material is published online only. To view, please visit the journal online (https://doi.org/10.1136/bmjph-2026-004892).

Received 6 January 2026
Accepted 23 September 2026

Check for updates

© Author(s) (or their employer(s)) 2026. Re-use permitted under CC BY-NC. Published by BMJ Group.

For numbered affiliations see end of article.

Correspondence to
Prof Zacharie Tsala Dimbuene; zacharie.tsala.dimbuene@gmail.com

ABSTRACT

Introduction Pooling datasets is increasingly used in global health research to increase statistical power, improve precision and enable cross-country comparisons. Yet, combining data from heterogeneous settings may obscure important contextual differences and lead to conclusions that diverge from those obtained through country-specific analyses. Using maternal education and child morbidity as a case study, this paper examines how pooled vs individual datasets can yield different interpretations, highlighting the implications for research and policy.

Methods We conducted a secondary analysis of Demographic and Health Surveys (DHS) from five Central African countries: the Democratic Republic of Congo, Cameroon, Republic of Congo, Gabon and Chad. The study population included children aged 0–59 months. Child morbidity outcomes—fever, diarrhoea and cough in the 2 weeks preceding the survey—were analysed using descriptive statistics and multivariate logistic regression models applied separately to pooled and country-specific datasets.

Results The final analytic sample comprised 58 079 children under 5 years of age. Educational attainment varied substantially across countries, with 66.0% (96% CI 64.0% to 67.9%) of women in Chad reporting no formal education—far higher than in the other countries. Overall prevalence of cough, diarrhoea and fever was 6.6% (95% CI 6.2% to 7.1%), 19.3% (95% CI 18.6% to 19.9%) and 26.1% (95% CI 25.2% to 26.9%), respectively, with notable cross-country variation. Multivariate analyses revealed that the association between maternal education and child morbidity differed between the pooled dataset and the individual country models, demonstrating how aggregation can mask local heterogeneity.

Conclusion Pooling datasets can be a powerful analytical strategy, particularly for increasing statistical power, improving the precision of estimates and enabling cross-country comparisons. Using DHS data from five Central African countries, this study demonstrates that while pooled analyses identify broad regional patterns, they can mask important country-specific heterogeneity, which is of paramount importance to devise national and local policies.

WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Pooling multiple datasets increases statistical power, enabling more precise estimates and greater ability to detect meaningful associations that may be missed in smaller, isolated samples.

⇒ Larger combined samples also improve the stability of model estimates, especially in subgroup analyses or geographically disaggregated settings.

WHAT THIS STUDY ADDS

⇒ By pooling datasets while explicitly addressing heterogeneity and contextual policy differences, this study provides new insights to better support evidence-based decision-making tailored to diverse local policy environments.

HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ When datasets span multiple jurisdictions, failing to incorporate policy context may obscure true effects or exaggerate geographic disparities.

INTRODUCTION

Data pooling from multiple countries into a single dataset provides opportunities to increase *statistical power*.[1] Demographic and Health Surveys (DHS) are very common in developing countries. Essentially, they have even become the main source of sociodemographic and health data in sub-Saharan Africa (SSA) as a replacement of traditional data sources like civil registration and censuses. The wealth of comparable information gathered from DHS provides promising avenues to further our understanding about social and health issues within and across countries in SSA.[2–5] Also, it has become common that scholars adopt 'pooling approaches' to study social and health issues in SSA.[6] Instances are[5 7–9] to study child morbidity, modern

BMJ Public Health: first published as 10.1136/bmjph-2026-004892 on 27 October 2026. Downloaded from https://bmjpublichealth.bmj.com on 11 November 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

contraceptive, mortality among twins and stillbirths in SSA. This paper argues that the first limitation of many studies using pooled DHS datasets to answer research questions is the lack of novelty since most research questions in these papers could be answered using data from a single country. Indeed, most studies using pooled DHS data overlooked the main motivation behind pooling strategy.

Furthermore, there are analytic and methodological issues to consider before diving into pooling data (see, for more details, ref.[10]). Wendt[10] provides a practical guidance on how to approach pooling data in the context of two identical surveys in Canada. Scholars mentioned the sample sizes as the critical element towards pooling data[11 12]: "The most common reason for researchers to integrate separate cycles is to increase sample sizes of small domains." This is not the case for DHS data which are nationally representative with large sample sizes. Therefore, when pooling DHS data, studies are increasing statistical power without novelty, and consequently, they are masking heterogeneity across countries. DHSs collect information on women of reproductive ages (15—49 years); however, the distribution of the samples might differ across surveys in the same country or they might be different across countries. Of paramount importance is the 'statistical heterogeneity', which states that individual trials or surveys that appear similar may report results that are not consistent with each other.[13]

This brings us to another programmatic issue. When pooling datasets, the analyst should be aware of the study focus. Oftentimes, the objective is to estimate the "effect of the average country in the sample".[14] For instance, when pooling data from 31 countries,[5] the researcher estimates the average effects of a putative variable on the outcome of interest for the 31 pooled datasets. By doing so, studies ignore *heterogeneity* across countries; therefore, findings become meaningless to design more effective local policies. Indeed, what is the practical value of generated evidence using pooled data when it is well established that SSA countries are heterogeneous and, additionally, economic and health policies could not ignore heterogeneity within SSA countries? Can these findings lead to local effective interventions to mitigate inequities within and across countries?

This paper contributes to the existing literature on pooling datasets and argues that findings generated using this approach might be less appealing in designing more effective economic and health interventions in SSA countries.

METHODS

Study design

Data from DHS of five countries (Democratic Republic of Congo (DRC), Cameroon, Republic of Congo, Gabon and Chad) in Central Africa were used in the study. These surveys are nationally representative and suitable to generate reliable subnational estimates. The datasets were from surveys conducted between 2011 and 2015 in the countries under consideration. Although Cameroon has more recent DHS data in 2018, preliminary analyses showed that the sample size of children under 5 years was slightly small compared to other countries. To keep sample sizes from all countries a bit comparable, 2011 Cameroon DHS data were used in the current study.

Sampling technique

The sample was selected for the surveys using multistage sampling technique. Each country was divided into clusters or communities. From each cluster, urban and rural areas were selected. Households were thereafter selected from the chosen urban and rural areas. Women within the reproductive age of 15–49 years in each household were eligible for interview. Information about the children was obtained from the mothers of such children.

Data collection

The use of questionnaire was adopted as a means of information gathering in the surveys. Questionnaires were administered by interviewers who engaged respondents in a face-to-face interview. Information obtained from each respondent related to sociodemographic characteristics, reproductive issues, breastfeeding, immunisation, childhood diseases, delivery and postnatal care, and health services use for children, among others.

Outcome variable

The primary outcome variable in this study was child morbidity, assessed among children under 5 years of age in Central Africa. Following the DHS approach, morbidity was measured based on mothers' reports of whether the child experienced fever, cough and diarrhoea in the 2weeks preceding the survey. Each morbidity indicator was coded as a binary variable ('1=yes' if the child had the illness, and '0=no' otherwise). These three conditions were selected because they are among the most common childhood illnesses in SSA and represent leading causes of under-five morbidity and mortality. For analysis, child morbidity was examined both as individual outcomes (fever, cough and diarrhoea).

Independent variables

Key independent variable

In this paper, maternal education is the key independent variable. Indeed, previous studies showed that maternal education is a strong predictor of child health.[15] DHS data provide the opportunities to test the effect of an additional year of maternal education on child health.

Control variables

Control variables were selected from previous similar studies.[5 16–18] The control variables included in the analyses are as follows: mother's age, household wealth, place of residence, employment, media exposure, antenatal care attendance, sex of child, child's age, vaccination status, birth order, baby's size at birth, source of water and toilet facility. Mother's age was defined as 15–24,

BMJ Public Health: first published as 10.1136/bmjph-2026-004892 on 27 October 2026. Downloaded from https://bmjpublichealth.bmj.com on 11 November 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

25–34 and 35 years and above. Household wealth was recoded into three categories as poor (40% bottom), middle (20%) and rich (40% top). Residence was categorised into urban and rural. Employment was grouped into working and not working. An index of media exposure was constructed using three original variables (Listening Radio, Watching Television, Reading Newspapers) and women were recoded as exposed and not exposed. Antenatal care attendance was captured using the number of antenatal visits a woman reported. Sex of child was defined as male and female. Child's age was categorised as 0–11, 12–23 and 24–59 months. Vaccination status was measured as 'no' for children who never had vaccine and 'yes' for those who were vaccinated. Birth order was expressed as first-order birth, second-order birth and third or higher order birth. Child's size at birth was defined as large, average and small. Source of water and toilet facility were grouped into improved and not improved.

Analytical strategy

The paper is primarily interested in the associations between maternal education and child morbidity in Central Africa countries. First, data for the five countries were pooled together to estimate the average effect of maternal education on the outcomes (fever, cough and diarrhoea) while accounting for the complex survey design (CSD) in DHS data. Because pooled DHS analyses can be sensitive to the weighting strategy adopted, we considered two alternative approaches. We implemented an equal-country weighting scheme by rescaling survey weights so that each country contributed equally to the pooled estimates, regardless of its sample size. This approach estimates the average effect across countries and is appropriate when the primary objective is comparative inference at the country level. Furthermore, a population-size weighting scheme was considered, whereby country-specific survey weights were adjusted to reflect each country's share of the combined regional population. This approach estimates the average effect for children living in the pooled population and gives greater influence to more populous countries. Comparing results across these weighting strategies provides a useful sensitivity analysis to assess the extent to which substantive conclusions are driven by differences in country size, sample allocation and population composition. To achieve this goal, unique identifiers for clusters and strata were constructed to obtain unbiased estimates using logistic regression given the binary outcomes.

Second, the paper argues that pooled data might under-(over-)estimate the effect of maternal education on child morbidity. To implement this idea, the effect of maternal education on child morbidity was estimated for each country and compared with the pooled estimate to showcase the potential heterogeneity across countries. Finally, national analyses were performed to estimate the effect of maternal education on child morbidity and compared with (1) pooled and (2) national estimates to detect heterogeneity within countries.

The *svyset* option was implemented to account for CSD in DHS datasets (svyset is a Stata command that declares a dataset's CSD—primary sampling units/clusters, strata and sampling weights—so that subsequent estimation commands produce design-adjusted SEs). The analyses included descriptive and multivariable. For the descriptive analysis, χ^2 test was used to test the relationship between the outcome variables and independent categorical variables using pooled and individual datasets. As mentioned above, multivariable analysis was performed through binary logistic regression model. The net effects of maternal education on child morbidity were plotted. Here, 'net effect' refers to the association between maternal education and each morbidity outcome after adjusting for the other covariates included in the multivariable logistic regression model—as distinct from the unadjusted (gross) bivariate associations reported in the descriptive analysis above. Additionally, ORs, together with their 95% CIs, are reported in Appendices. Goodness-of-fit test was conducted to ensure that the models were well fitted.

Patient and public involvement

Patients/public were not involved in the design or implementation of this study.

FINDINGS

Descriptive results

Description of the sample: children under 5 years in Central Africa.

The final sample consisted of 58 079 children under 5 years, unequally distributed across the five countries, with DRC (29.3%) and Chad (28.8%) having the larger shares, followed by Cameroon (17.8%) and the Republic of Congo (14.5%). Gabon has a marginal share with 9.7% (see table 1).

Homogeneity versus heterogeneity: what the data tell us?

Before diving into more sophisticated analysis of potential heterogeneity, the paper builds on a simple indicator regarding the main independent variable of the study: *maternal education*. Although analyses rely on maternal education as measured in single years, the paper

Table 1 Description of the study sample: children under 5 years in Central Africa

Variable	N	Percentage
Country, year		
Democratic Republic of Congo (2013–2014)	16 999	29.3
Cameroon (2011)	10 317	17.8
Republic of Congo (2011–2012)	8426	14.5
Gabon (2012)	5603	9.7
Chad (2014–2015)	16 734	28.8
Total	58 079	100.0

BMJ Public Health: first published as 10.1136/bmjph-2026-004892 on 27 October 2026. Downloaded from https://bmjpublichealth.bmj.com on 11 November 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

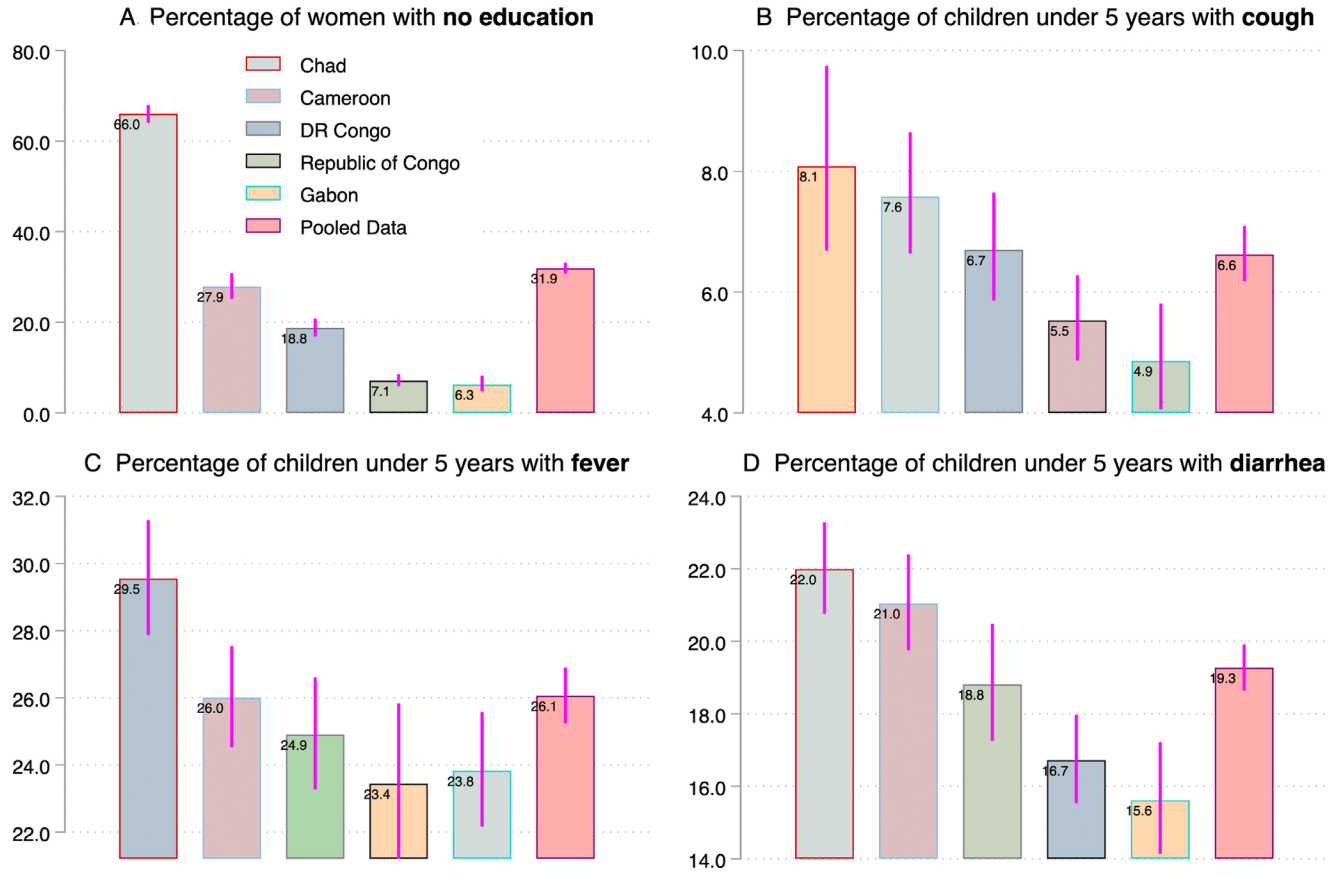


Figure 1 (A) Percentage of women of reproductive ages with no education in Central Africa. (B) Percentage of children under 5 years with cough. (C) Percentage of children under 5 years with fever. (D) Percentage of children under 5 years with diarrhoea. DR Congo, Democratic Republic of Congo.

constructed an indicator of *women without education* to demonstrate how countries in Central Africa might be different.

On average, one-third of women of reproductive ages reported no education at the time of surveys. However, figure 1A shows that countries included in this study are very different when one looks into the percentage of women of reproductive ages with no education. Indeed, findings showed that percentage of women of reproductive ages with no education is much higher in Chad (66.0%) compared with other countries.

Looking into study outcome variables (cough, diarrhoea and fever) among children under 5 years also evidence heterogeneity across countries. For instance, one might expect that a country heavily disadvantaged by one disease would also be disadvantaged by others. Findings indicated strong variations depending upon the outcome of interest (see figure 1B–D).

Indeed, findings indicated that overall 6.6% (95% CI 6.2% to 7.1%) of children under 5 years were affected by cough in pooled data (figure 1B). However, this situation varied across countries from Gabon (8.1%; 95% CI 6.5% to 9.6%) to Republic of Congo (4.9%; 4.0% to 5.7%). A similar observation can be done for fever (figure 1C) and diarrhoea (figure 1D). Figure 1C shows that 26.1% (95% CI 25.2% to 26.9%) of children under 5 years had fever in the pooled dataset. This figure varied across countries from 29.5% (95% CI 27.8% to 31.2%) in the DRC to 23.4% (95% CI 21.1% to 25.8%) in Gabon. Finally, figure 1D shows that 19.3% (95% CI 18.6% to 19.9%) of children under 5 years suffered from diarrhoea in the 2 weeks preceding the survey. Likewise, there are noticeable variations across countries with Chad being the most affected country (22.0%; 95% CI 20.7% to 23.3%) compared with Cameroon (21.0%; 95% CI 19.7% to 22.4%), Republic of Congo (18.8%; 95% CI 17.2% to 20.4%), DRC (16.7%; 95% CI 15.5% to 17.9%) and Gabon (15.6%; 95% CI 14.1% to 17.2%). Addition information about the association between independent variables and child morbidity can be found in online supplemental table 1.

Multivariate results

Figure 2A–C summarises the net effects of maternal education measured in single years on child morbidity (cough, diarrhoea and fever, respectively) in the last weeks preceding the survey among children under 5 years in Central Africa. As defined in 'Statistical analysis', net effects denote the adjusted association between maternal education and each outcome after controlling for the

BMJ Public Health: first published as 10.1136/bmjph-2026-004892 on 27 October 2026. Downloaded from https://bmjpublichealth.bmj.com on 11 November 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

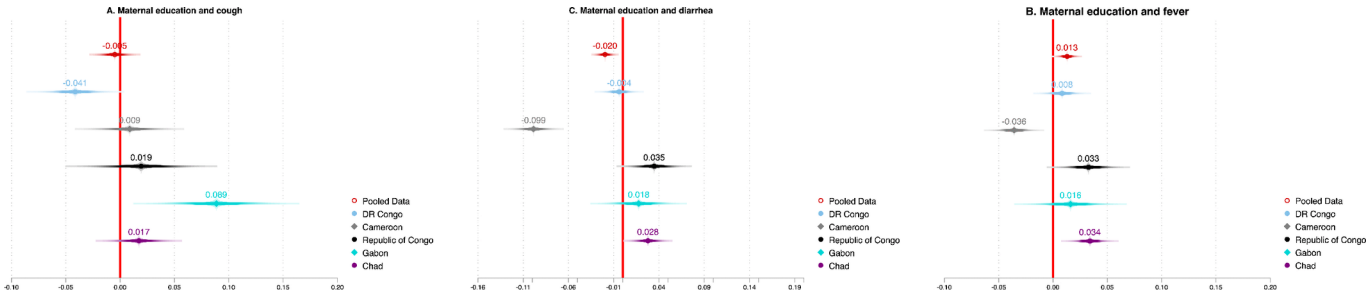


Figure 2 (A) Maternal education and cough in Central Africa. (B) Maternal education and fever in Central Africa. (C) Maternal education and diarrhoea in Central Africa. DR Congo, Democratic Republic of Congo.

other covariates in the multivariable model, shown separately for the pooled sample and for each country. Details on multivariate models between independent variables and child morbidity are given in online supplemental table 2. Based on previous research, one might expect that maternal education is negatively associated with child morbidity outcomes, including cough, diarrhoea and fever.

Maternal education and cough

From figure 2A, findings indicate that overall (pooled data) maternal education is negatively associated with cough even though the relationship did not reach statistical significance. Findings also indicate variations across countries. For instance, while findings went in the expected direction in the DRC, unexpected relationships were observed in the rest of the countries. More important, maternal education significantly increased the likelihood of cough in Gabon.

Maternal education and fever

Findings show that maternal education was positively associated with the occurrence of fever in the last couple of weeks preceding the survey in the pooled data (see figure 2B). Again, the net effects of maternal education varied across countries and going even in the unexpected direction. Indeed, maternal education was negatively and significantly associated with the occurrence of fever in Cameroon. In the rest of countries included in the study, maternal education increased the risk of fever among children under 5 years in the last 2 weeks preceding the survey. For instance, maternal education was positively and significantly associated with fever in Chad.

Maternal education and diarrhoea

From figure 2C, findings indicate that maternal education is negatively and significantly associated with diarrhoea in the pooled dataset. However, this association varies across countries. Indeed, maternal education was negatively and significantly associated with the occurrence of diarrhoea in Cameroon while the negative association was marginal in the DRC. In contrast, maternal education was positively associated with the occurrence of diarrhoea in the Republic of Congo, Gabon and Chad, even though the association reached statistical significance only in Chad.

DISCUSSION

Pooling datasets has become an important analytic strategy in many studies using DHS in SSA.[5] Increasing sample size is one of the main justifications for pooling datasets.[10 11] The paper argues that this main justification does not hold when using DHS datasets which are cross-sectional nationally representative datasets with large sample sizes. Specifically, the paper used DHS datasets in Central Africa, a region with the worst development indicators (eg, Corruption Index; Democratic accountability; Law and order; for more details, see ref[19]) and maternal education as the key independent variable to show that findings from pooled DHS datasets are questionable, especially from a policy perspective.

Previous research showed that pooling datasets has several advantages. These include, among others, (1) increased sample size and statistical power.[20 21] Pooling increases the number of observations, which improves the precision of estimates and allows detection of smaller effects; (2) greater generalisability.[22] Indeed, combining data from different populations, regions or time periods provides a broader picture and improves external validity; and (3) the ability to study rare events.[23] However, pooling datasets has also several disadvantages, including heterogeneity across datasets,[24] loss of contextual specificity,[25] statistical complexities[24] and potential for bias.[24] In the next subsection, the paper focuses on two important pitfalls of pooling datasets which have major policy implications: heterogeneity across datasets and loss of contextual specificity.

MAIN FINDINGS

Heterogeneity across countries

Previous research showed that heterogeneity across datasets or countries might result from differences in definitions, measurement tools, survey design or population characteristics may introduce bias or limit comparability. However, this is not the case in DHS, which use standardised questionnaires. The paper demonstrated that countries included in the study were heterogeneous in terms of key independent variable (maternal education measured in single completed years) and child health outcomes (child morbidity measured by cough, diarrhoea

BMJ Public Health: first published as 10.1136/bmjph-2026-004892 on 27 October 2026. Downloaded from https://bmjpublichealth.bmj.com on 11 November 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

and fever). For the key independent variable, the paper used the percentage of women with no education to demonstrate the significant variations across countries. For instance, findings showed that this percentage of women of reproductive ages with *no education* was as low as 6% (95% CI 4.6% to 7.9%) in Gabon to a majority of 66% (96% CI 64.0% to 67.9%) in Chad. Previous studies provide valuable insights pooling together DHS datasets.[5 7] However, efforts to account for heterogeneity across countries to nuance the conclusions of the studies were missing.

The paper substantively demonstrated that maternal education was not uniformly associated with lower child morbidity across all countries. In some settings, higher maternal education was associated with an increased likelihood of reported fever, cough or diarrhoea. Although these findings may appear counterintuitive, several explanations are plausible. First, educated mothers may be more likely to recognise symptoms and accurately report childhood illnesses, whereas less educated mothers may underreport symptoms due to limited health knowledge or reduced awareness of disease manifestations. Second, higher levels of education are often associated with greater healthcare utilisation and interaction with health services, which may increase disease detection and reporting. Third, differences in epidemiological contexts, environmental exposures and health system performance across countries may modify the relationship between maternal education and child health outcomes. Finally, residual confounding from unmeasured factors—such as household health-seeking behaviour, quality of care, maternal autonomy or local disease ecology—may also contribute to these unexpected associations. These findings reinforce the importance of interpreting pooled estimates within their specific country contexts and caution against assuming that the effects of maternal education on child health are homogeneous across settings.

Loss of contextual specificity and policy implications

While pooling datasets is justifiable in certain circumstances, especially when sample sizes are small, scholars should not elude the fact that contexts equally matter when analysing social issues, which strongly impact national and local policymaking to improve people's wellbeing. Therefore, one of the drawbacks when pooling datasets is that findings mask country-specific contexts which are of paramount importance to devise more effective policies at national and sub-national levels. For instance, the paper substantiated that the effect of maternal education on diarrhoea and cough is negative and significant in the pooled dataset. However, this effect varied at country level, being *positive* in the Republic of Congo, Gabon and Chad, *negative* in Cameroon and almost *nil* in the DRC. Programmatically, it means that using findings from pooled datasets can be misleading when addressing local issues like child morbidity since the contexts do matter.

Study strengths and limitations

The paper used nationally representative data which yielded robust estimates of the associations between maternal education and child morbidity in the context of pooling datasets. The strengths include, among others, larger sample sizes and statistical power; cross-country comparisons; generalisability and methodological innovation which enables multilevel and hierarchical modelling since individuals are nested in households, clusters and countries. Some limitations include heterogeneity across countries (cultural contexts can make direct comparisons difficult); survey weights and representativeness since weights are designed for each survey but pooled datasets. Rescaling and/or harmonisation is necessary for pooled analysis. Policymakers should also avoid over-generalisation of findings from pooled datasets because there might be important variations across countries as shown in the paper.

Finally, the three morbidity indicators (fever, diarrhoea and cough) were analysed separately because DHS records them as distinct binary outcomes and because each condition has potentially different determinants and policy implications. Plausibly, analysing the co-occurrence of these symptoms may provide additional information. However, modelling joint morbidity would require either multivariate response models, latent-variable approaches or constructing composite morbidity indices. Such analyses are beyond the scope of the present methodological paper.

CONCLUSION

In this study, maternal education was associated with child morbidity in pooled analyses, but country-level findings revealed notable variations. These results underscore the importance of balancing pooled evidence with country-specific insights to inform context-specific child health interventions. Although pooling datasets enhances statistical power and allows the detection of regional patterns, context-specific evidence remains essential for designing effective child health strategies in Central Africa.

Pooling datasets, such as the DHS, offer clear advantages, including larger sample sizes, improved statistical power and the ability to analyse rare outcomes and cross-country or temporal patterns. These strengths make pooled analyses a valuable approach for informing regional and global health and development policies. However, pooling also presents important challenges: differences in survey timing, sampling designs and variable coding can complicate comparability, while contextual heterogeneity may obscure country-specific realities. For robust and policy-relevant findings, pooled analyses should therefore be accompanied by rigorous harmonisation procedures, appropriate weighting strategies and sensitivity analyses that account for

BMJ Public Health: first published as 10.1136/bmjph-2026-004892 on 27 October 2026. Downloaded from https://bmjpublichealth.bmj.com on 11 November 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.

both methodological and contextual differences. As such, findings from pooled datasets might not fully serve effective local policies, especially because these findings mask local specificity. Policymakers should use findings from pooled datasets with caution.

Author affiliations
[1]School of Population and Development Sciences, University of Kinshasa, Kinshasa, Congo (the Democratic Republic of the)
[2]University of Zambia, Lusaka, Zambia
[3]Université du Québec, Rimouski, Quebec, Canada
[4]Dar es Salaam University College of Education, Universiry of Dar es Salaam, Dar es Salaam, Tanzania, Dar es Salaam, Tanzania, United Republic of
[5]Arba Minch University, Arba Minch, Ethiopia
[6]University of Technology and Applied Sciences, Navrongo, Ghana

Acknowledgements The DHS programme is acknowledged for permitting us to use available datasets for Central Africa, including Cameroon, Chad, Democratic Republic of the Congo, Gabon and the Republic of Congo.

Contributors ZTD contributed to the conceptualisation of research questions, drafted the analysis plan and wrote the initial draft of the study and responding to reviewers' comments. ETN reviewed the first draft of the manuscript and provided inputs on revised manuscript. PK reviewed the first draft of the manuscript and provided inputs on revised manuscript. JK reviewed the first draft of the manuscript and provided inputs on revised manuscript. NBS reviewed the first draft of the manuscript and provided inputs on revised manuscript. ZTD is the guarantor.

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.

Ethics approval The DHS programme obtained ethical approval from the relevant national ethics committees and ICF Institutional Review Board before survey implementation in each participating country. Permission to use the datasets was obtained through the DHS programme. The datasets are anonymised and publicly available for research purposes; therefore, no additional ethical approval was required for this secondary analysis. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. All replication files can be found here https://osf.io/96j74/files/osfstorage.

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: https://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Zacharie Tsala Dimbuene https://orcid.org/0000-0002-7977-1689
Philippe Kabore https://orcid.org/0000-0003-2228-2080
Janet Kalinga https://orcid.org/0009-0006-7704-3684
Negussie Boti Sidamo https://orcid.org/0000-0001-8949-573X
Francisca Achaliwie https://orcid.org/0000-0002-8209-2353

REFERENCES

1 Adhikari K, Patten SB, Patel AB, *et al*. Data harmonization and data pooling from cohort studies: a practical approach for data management. *Int J Popul Data Sci* 2021;6:1680.
2 Burstein R, Henry NJ, Collison ML, *et al*. Mapping 123 million neonatal, infant and child deaths between 2000 and 2017. *Nature* 2019;574:353–8.
3 Martín AIC, Bil K, Salumu P, *et al*. Mortality Rates above Emergency Threshold in Population Affected by Conflict in North Kivu, Democratic Republic of Congo, July 2012–April 2013. *PLoS Negl Trop Dis* 2012.
4 Ezbakhe F, Pérez Foguet A. Child mortality levels and trends: A new compositional approach. *DemRes* 2020;43:1263–96.
5 Adedokun ST, Yaya S. Childhood morbidity and its determinants: evidence from 31 countries in sub-Saharan Africa. *BMJ Glob Health* 2020;5:e003109.
6 Mayen MSA, Nisha SN, Afrin S, *et al*. Evaluating the current methodological practices and issues in existing literature in pooling complex surveys: a systematic review. *BMC Med Res Methodol* 2024;24:279.
7 Tessema ZT, Teshale AB, Tesema GA, *et al*. RETRACTED: Pooled prevalence and determinants of modern contraceptive utilization in East Africa: A Multi-country Analysis of recent Demographic and Health Surveys. *PLoS ONE* 2021;16:e0247992.
8 Monden CWS, Smits J. Mortality among twins and singletons in sub-Saharan Africa between 1995 and 2014: a pooled analysis of data from 90 Demographic and Health Surveys in 30 countries. *Lancet Glob Health* 2017;5:e673–9.
9 Akombi BJ, Ghimire PR, Agho KE, *et al*. Stillbirth in the African Great Lakes region: A pooled analysis of Demographic and Health Surveys. *PLoS ONE* 2018;13:e0202603.
10 Wendt M. Considerations before pooling data from two different cycles of the general social survey. Ottawa, 2007.17. Available: https://www23.statcan.gc.ca/imdb-bmdi/document/ 8011_D1_T9_V1-eng.pdf
11 Backenroth D, Royce T, Pinheiro J, *et al*. Considerations for pooling real-world data as a comparator cohort to a single arm trial: a simulation study on assessment of heterogeneity. *BMC Med Res Methodol* 2023;23:193.
12 Zhou HH, Singh V, Johnson SC, *et al*. The alzheimer's disease neuroimaging initiative. sta- tistical tests and identifiability conditions for pooling and analyzing multisite datasets. *Proceed- Ings of the National Academy of Sciences* 2018;1481–6.
13 Lensen S. When to pool data in a meta-analysis (and when not to)? *Fertil Steril* 2023;119:902–3.
14 Miratrix LW, Weiss MJ, Henderson B. An Applied Researcher's Guide to Estimating Effects from Multisite Individually Randomized Trials: Estimands, Estimators, and Estimates. *J Res Educ Eff* 2021;14:270–308.
15 Rezaeizadeh G, Mansournia MA, Keshtkar A, *et al*. Maternal education and its influence on child growth and nutritional status during the first two years of life: a systematic review and meta-analysis. *EClinicalMedicine* 2024;71:102574.
16 Afrifa-Anane GF, Kyei-Arthur F, Agyekum MW, *et al*. Factors associated with comorbidity of diarrhoea and acute respiratory infections among children under five years in Ghana. *PLoS ONE* 2022;17:e0271685.
17 Amugsi DA, Aborigo RA, Oduro AR, *et al*. Socio-demographic and environmental determinants of infectious disease morbidity in children under 5 years in Ghana. *Glob Health Action* 2015;8:29349.
18 Aheto JMK, Duah HO, Agbadi P, *et al*. A predictive model, and predictors of under-five child malaria prevalence in Ghana: How do LASSO, Ridge and Elastic net regression approaches compare? *Prev Med Rep* 2021;23:101475.
19 Pongou R, Kuate Defo B, Tsala Dimbuene Z. Excess Male Infant Mortality: The Gene-Institution Interactions. *Am Econ Rev* 2017;107:541–5.
20 Hussong AM, Curran PJ, Bauer DJ. Integrative data analysis in clinical psychology research. *Annu Rev Clin Psychol* 2013;9:61–89.
21 Curran PJ, Hussong AM. Integrative data analysis: the simultaneous analysis of multiple data sets. *Psychol Methods* 2009;14:81–100.
22 Little TD, Rhemtulla M. Planned Missing Data Designs for Developmental Researchers. *Child Dev Perspect* 2013;7:199–204.
23 Riley RD, Lambert PC, Abo-Zaid G. Meta-analysis of individual participant data: rationale, conduct, and reporting. *BMJ* 2010;340:bmj.c221.
24 Stewart LA, Clarke M, Rovers M, *et al*. Preferred Reporting Items for Systematic Review and Meta-Analyses of individual participant data: the PRISMA-IPD Statement. *JAMA* 2015;313:1657–65.
25 Smith GD, Ebrahim S. Data dredging, bias, or confounding. BMJ 2002;325:1437–8.

BMJ Public Health: first published as 10.1136/bmjph-2026-004892 on 27 October 2026. Downloaded from https://bmjpublichealth.bmj.com on 11 November 2026 by guest. Protected by copyright, including for uses related to text and data mining, AI training, and similar technologies.