Challenges and Solutions to Combining Data in IPUMS Global Health

March 19, 2025
Questions and Answers

The following are adaptations of questions received during the live webinar and their answers. If you have additional questions, please contact IPUMS User Support at ipums@umn.edu.

Data Availability and Access

1. **When will IPUMS Global Health be releasing more data?**

   IPUMS DHS is working on releasing standard variables from over 100 additional samples in late 2025 or early 2026.

   IPUMS PMA is preparing a data release of service delivery point data by summer 2025. IPUMS PMA will also release the last rounds of longitudinal household and female data and client exit interview data by fall 2025.

   IPUMS MICS is working on a data release of new samples and new variables that will be available by summer 2025. IPUMS MICS has submitted a proposal for another 5 years of funding, and if that is approved, that project will begin harmonizing and releasing additional data, including the latest MICS surveys.

2. **What is the current status of access to IPUMS DHS?**

   If you already have approved access to DHS data, you can log into IPUMS DHS and download the data you need, for an ongoing or for a new research project. This access is not affected by cuts to USAID and The DHS Program.

   The end of funding to The DHS Program affects: 1) the collection of new data (post-2023 surveys) and 2) the ability to extend access to additional datasets to approved DHS users and to give access to new users. We are cautiously optimistic that the second problem will be resolved eventually.

Data Discovery and Management

3. **How can we identify parallel variables in IPUMS DHS and IPUMS MICS?**

   You can use the IPUMS Discovery search tool at
   https://variable-search.ipums.org/#/
   Specify the data collections (and, if relevant, countries and years) important to you, and enter search terms.

   Alternatively, you could visit the separate IPUMS Global Health data collection websites and use the search tool there, or take notes on variable names from the drop-down topical menus for each unit of analysis.

   Pending renewal of IPUMS MICS funding, we plan to offer variable name crosswalks between IPUMS DHS and IPUMS MICS.

4. **I'm interested in studying child discipline and child labor across as many surveys as possible. I see only IPUMS MICS has older children as a unit of analysis. Can I only use IPUMS MICS for this?**

   Although IPUMS DHS does not have a separate unit of analysis for children over age 4, it does collect information about child discipline and child labor through the household member questionnaire. Additionally, child labor information can sometimes be found on the household member and household units of analysis in IPUMS MICS, depending on the sample.

   In general, search for variables on a topic, and then see which unit(s) of analysis you need to analyze accordingly.

5. **Does IPUMS include the functionality to automatically merge data for different units of analysis, for example children aged 5-17 with household datasets in MICS?**

   Please consult the user note about linking MICS data across units of analysis, at https://mics.ipums.org/mics/linking.shtml. Pending renewal of IPUMS MICS funding, we plan to develop and share code to simplify linking data across units of analysis.

6. **What should I do if I want to get population-level information (e.g., combine men and women) information?**

   The easiest way to proceed is to use household members' data, since that covers males and females of all ages. That would give you basic demographic information on the population, for example.

   For more detailed information on specific topics (for example, on attitudes toward domestic violence or toward family planning), you will need to download and analyze data separately for each unit of analysis (e.g., men and women, children 0-4 and children 5-17).

Interoperability

7. Why do variables about the same thing in the three global health data collections have different names?

Each IPUMS Global Health data collection was developed independently, under separate funding mechanisms and through distinct partnerships with organizations that collected the original survey data. Our data collection partners wanted to maintain the distinct identity of their data, which often results in unique variable names for each data collection. Additionally, there are some subtle differences in similar variables across data collections, and using the same variable names in every case might have obscured those differences.

8. When we combine multiple datasets from separate surveys, what adjustments are necessary for the sample weights?

We are developing guidance on this topic. The appropriate adjustment will depend on the samples you are combining and your analysis (e.g., time trends versus pooling). For the example used in the webinar, which presented family planning use prevalence by country, age group, and mobile phone ownership using data from DHS, MICS, and PMA, we separately analyzed the datasets, using their respective weights, and then combined final results in our graphs.

9. Where can we find more detailed information about combining data across IPUMS Global Health datasets?

See the user note at https://www.idhsdata.org/idhs/interop.shtml

Other

10. Where can I pose a question about a Global Health variable and/or sample that is unique to my research project?

Send an email describing the issue to IPUMS User Support at ipums@umn.edu .