

Household surveys: do competing standards serve country needs?

Discussion Paper No. 4, June 28, 2016, Joined-up Data Standards project

Beata Lisowska

Data Scientist, Development Initiatives

Contents

Executive summary	2
Introduction.....	3
The challenge facing users of data	5
Household surveys in Africa	6
Mapping between DHS and MICS surveys	8
Methodology.....	9
Household questionnaire	11
Men's questionnaire	12
Women's and children under-five questionnaire.....	13
Findings of mapping between DHS-VII and MICS5.....	14
How does a country decide which survey to conduct?	16
The cost	16
Survey design and sample size	16
Data needs: the case of Rwanda.....	17
Summary of findings	20
Conclusions	20
Recommendations.....	22
References	23

Executive summary

Household surveys are currently the most important data source for a range of key demographic and socioeconomic statistics in developing countries. They are the most effective method of filling the vacuums that exist because of a lack of credible data from more sustainable registry and administrative sources. Even as better systems are rolled out, surveys will continue to play an important quality-control role.

There are three major international household survey programmes in use around the world. These have become increasingly similar but each contains unique, useful modules. Our research finds that two-thirds of the questions in the two most widely used surveys are either identical or similar enough to be practically comparable.

This presents developing countries with a dilemma. Do they, at great expense, commission multiple surveys, or do they accept that they cannot afford to collect all the data they require?

There are two ways to solve the problem of competing standards: combine them into one, or establish functional links between them. We argue that the interests of developing countries would be best served by the integration of the three programmes, and that until this is possible it is critical that data from different surveys is capable of being joined up.

Over the past 18 months, the problems of duplication and the lack of interoperability and comparability – between surveys, datasets, countries and over time – have been recognised by the global statistical community in general, and by the three lead institutions in particular. The UN Statistical Commission has established a working group to coordinate efforts to tackle the problem and the three agencies – UNICEF, USAID and the World Bank – have signed a statement of collaboration and are working more closely together.

While welcoming these initiatives we urge all those involved to recognise that in the current political climate, with the adoption of the Sustainable Development Goals and the Transformative Agenda for Official Statistics, there is a real opportunity to fast-track the political and practical work required to pool all resources and expertise to ensure the most beneficial and cost-effective outcome for developing countries.

Introduction

Socioeconomic and demographic statistics can be produced through the collection of four main types of data sources:

- national population and household censuses (typically conducted every 10 years)
- registry data (such as Civil Registration which records births, deaths and marriages)
- administrative data (collected on a regular basis by government departments and agencies in the course of their duties)
- household surveys.

In developing countries, where vital registration and administrative systems are lacking and the information gaps are largest, household surveys currently play a critical role. Until better sources of data are available through credible, functioning and sustainable systems this will remain the case. Even in an ideal future data-ecosystem, surveys will remain key checks for quality assurance.

Household surveys have been an invaluable source of data for the Millennium Development Goals (MDGs) due to the flexibility of their framework. Surveys have included questions relating to goals such as eradicating poverty, reducing child and maternal mortality or achieving universal primary education, which in turn have helped to monitor progress towards these goals. The same flexibility may be needed to monitor progress towards the extended Sustainable Development Goals (SDGs). The SDGs include goals even more difficult to monitor, such as gender equality and inclusive communities.

There are three main household surveys used in developing countries across the world.

Acronym	Name	Agency
DHS	Demographic and Health Survey	United States Agency for International Development ¹ (USAID)
MICS	Multiple Indicator Cluster Survey	United Nations Children's Fund (UNICEF)
LSMS	Living Standards Measurement Study	World Bank (WB)

These three products compete with each other in a relatively open marketplace where developing countries choose the programme (or programmes) they wish to implement based on their data needs but are also strongly influenced by cost and available funding (from domestic or external resources). As many countries have employed more than one of these surveys, and as survey data is fed into globally comparable statistics, the need for consistent data baselines has forced these international programmes to harmonise both their questionnaires and their tools to allow easy comparison between their datasets. The recognition that this data has a critical role to play in the 2030 Agenda for Sustainable Development has led to further attention being focused on a more joined-up approach to data production.

- In January 2015 at the [Global Conference on a Transformative Agenda for Official Statistics](#) this need for continuous and standardised data was expressed in an agreement that an integrated household survey programme should be “mainstreamed to assist countries in streamlining the statistical production processes by facilitating cost efficiency, lowering response burden and ensuring the production of better-quality and consistent statistics”.

- In March 2015 the [UN Statistical Commission endorsed](#) the establishment of an Intersecretariat Working Group on Household Surveys under the aegis of the UN Statistics Division in order to “foster the coordination and harmonisation of household survey activities”.
- In May 2015, USAID, UNICEF and the World Bank [signed an agreement](#) to “increase the frequency, quality, and relevance of household survey data around the world by better serving countries in meeting their domestic and international data demands through improved comparability and integration across surveys, enhanced survey methods and techniques, and greater coordination on survey timing and scheduling”.

This paper explores the challenges facing these institutions in turning their visions and commitments into a working reality. It focuses on the two most widely conducted survey programmes: USAID’s DHS and UNICEF’s MICS.ⁱⁱ More specifically the analysis is based on the latest operational versions: DHS Version VII and MICS Version 5. We focus on the occurrence and content of surveys as these two elements are most relevant to an understanding of the duplication of effort. To simplify the geographic scope, the evidence presented in this paper focuses primarily on Africa.

We are grateful to members of staff at USAID, UNICEF, the UN Statistics Division, the World Bank, The Partnership in Statistics for Development in the 21st Century (PARIS21), the UK Department for International Development, the National Institute of Statistics of Rwanda and the Kenya Bureau of Statistics for answering specific questions. They have helped to inform our understanding of the issues involved in this study, but have not been party to our analysis. The opinions expressed here are entirely those of the [Joined-up Data Standards project](#) team.

The challenge facing users of data

The data from DHS and MICS household surveys has many uses. Beyond its national relevance it feeds into a number of global statistical databases; it is used for regional comparisons; and it will be required to calculate a number of SDG indicators. Table 1 shows the most recent data that is currently available from DHS and MICS surveys. Statisticians and data scientists needing to pull this data together for comparative analysis face two challenges. Firstly, it covers a ten-year time span. Secondly, the data is formatted in six different ways: not only is data coded differently between DHS and MICS, but data structures are substantially different between versions of the same programme.

Survey	Demographic and Health Survey			Multiple Indicator Cluster Survey (MICS)			total
Version	V	VI	VII	3	4	5	
2005	Moldova						1
2006	Azerbaijan India			Guinea-Bissau, Somalia, Syria, Trinidad and Tobago, Turkmenistan, Uzbekistan			8
2007				Vanuatu			1
2008	Albania, Bolivia, Sao Tome and Principe						3
2009	Guyana, Maldives	Lesotho, Timor-Leste					4
2010		Armenia, Burundi, Colombia			Bhutan, Central African Republic, Chad, Kazakhstan, South Sudan, Suriname, Swaziland		10
2011		Angola, Cameroon, Congo, Ethiopia, Mozambique			Afghanistan, Belize, Bosnia and Herzegovina, Costa Rica, Iraq, Jamaica, Laos, Macedonia, Mauritania		14
2012		Benin, Cote D'Ivoire, Comoros, Gabon, Guinea, Haiti, Honduras, Indonesia, Jordan, Mali, Niger, Pakistan, Peru, Tajikistan, Tanzania			Algeria, Barbados, Belarus, Moldova, Saint Lucia, Thailand, Tunisia, Ukraine		23
2013		DR Congo, Dominican Republic, Gambia, Liberia, Madagascar, Namibia, Nigeria, Philippines, Sierra Leone, Yemen, Zambia			Uruguay	Mongolia, Montenegro	14
2014		Egypt, Togo	Bangladesh, Burkina Faso, Cambodia, Ghana, Malawi, Senegal, Uganda			Kosovo, Kyrgyzstan, Nepal Palestine, Serbia, Sudan, Viet Nam, Zimbabwe	17
2015			Kenya, Rwanda				2
Total	8	38	9	7	25	10	97

Table 1. Most recent survey data currently available for 98 countries from DHS or MICS household surveysⁱⁱⁱ

Household surveys in Africa

Only three African countries currently have fully functioning Civil Registration and Vital Statistics (CRVS) systems. In many countries government-wide administrative data infrastructures lack financial, human and technical resources. Household surveys therefore play a critical role in providing data about the health and welfare of people.

The Multiple Indicator Cluster Survey (MICS) and the Demographic and Health Survey (DHS) are the two most widely used surveys. The DHS programme began in 1984 building on earlier work initiated in the 1970s through the [World Fertility Survey and Contraceptive Prevalence Surveys](#). MICS was created in 1994, with its main focus to monitor the goals of the [1990 World Summit for Children](#). Over the years both surveys have broadened their scope to gather comprehensive information [on socio-economic and health indicators](#).

Most African countries, as Figure 1 shows, have used both surveys over the years. Since 2009, 15 African countries have conducted both MICS and DHS surveys, and in addition a number of surveys have used a combination of DHS and MICS. (This involves, for example, a standard DHS survey that incorporates modules from the MICS questionnaire.) As the host country chooses which survey to conduct (Table 2), both the MICS and DHS programmes need to rely on one another's datasets to ensure a continuous data series for analysis of health trends and socioeconomic progress.

To understand this pattern of usage in Africa, it is relevant to explore the following questions in more detail.

- What is the difference between the DHS and MICS surveys?
- How does a country decide which survey to conduct?

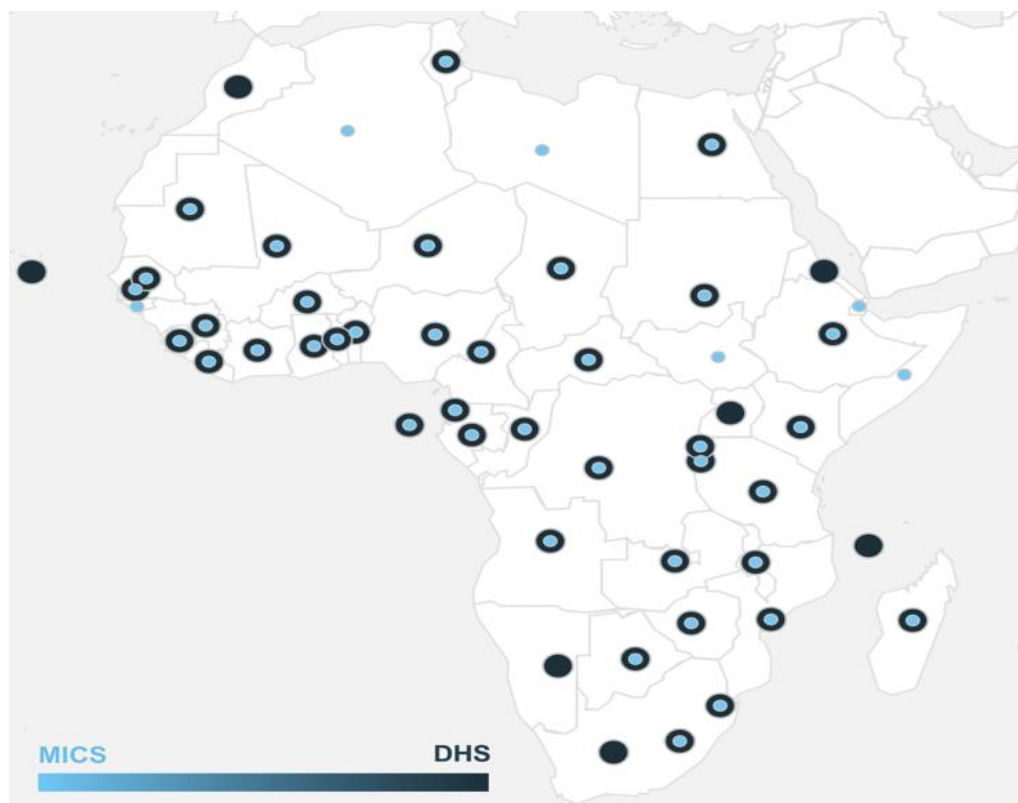


Figure 1. Use of MICS and DHS household surveys in Africa between 1985 and 2016^{iv}

Country	2009	2010	2011	2012	2013	2014	2015	2016
Algeria				MICS4				
Angola			DHS-VI				DHS-VII	
Benin			DHS-VI			MICS5		
Botswana								
Burkina Faso		DHS-VI				DHS-VII/LSMS		
Burundi		DHS-VI		DHS-VI				DHS-VII
Cameroon			DHS-VI			MICS5		DHS-VII
Cape Verde								
Central African Republic		MICS4						MICS6
Chad		MICS4				DHS-VII		
Cote d'Ivoire			DHS-VI					MICS5
Comoros				DHS-VI				
Congo (Brazzaville)			DHS-VI			MICS5		
Congo (DRC)		MICS4			DHS-VI			MICS6
Djibouti								
Egypt						DHS-VI	DHS-VII	
Equatorial Guinea			DHS-VI					MICS6
Eritrea								
Ethiopia			DHS-VI/LSMS		LSMS	DHS-VI		DHS-VII
Gabon				DHS-VI				
Gambia		MICS4			DHS-VI			
Ghana	LSMS		DHS-VI / MICS4			DHS-VII		MICS6
Guinea				DHS-VI				MICS5
Guinea-Bissau		MICS4				MICS5		
Kenya		DHS-VI				DHS-VII		
Lesotho	DHS-VI					DHS-VII		
Liberia			DHS-VI		DHS-VI			
Libya								
Madagascar			DHS-VI					
Malawi		DHS-VI/LSMS		DHS-VI	DHS-VI/MICS5/LSMS		DHS-VII	
Mali	MICS4	DHS-VI		DHS-VI		LSMS	MICS5	
Mauritania			MICS4				MICS5	
Mauritius								
Morocco								
Mozambique			DHS-VI					
Namibia	DHS-VI				DHS-VI			
Niger			LSMS	DHS-VI				
Nigeria		DHS-VI/LSMS	MICS4	LSMS	DHS-VI			MICS5
Rwanda		DHS-VI				DHS-VII		
Sao Tome and Principe						DHS-VII/MICS5		
Senegal		DHS-VI		DHS-VI		DHS-VII		
Seychelles								
Sierra Leone		MICS4			DHS-VI			MICS6
Somalia			MICS4*					
South Africa								DHS-VII
South Sudan		MICS4						
Sudan		MICS4						
Swaziland		MICS4				MICS5		
Tanzania		DHS-VI/LSMS	DHS-VI	LSMS			DHS-VII	
Togo		MICS4			DHS-VI			MICS6
Tunisia			MICS4					
Uganda	LSMS	LSMS	DHS-VI/LSMS					DHS-VII
Zambia					DHS-VI			
Zimbabwe		DHS-VI				MICS5	DHS-VII	

Table 2. Household surveys (programme and version) carried out in African countries, 2009–2016^v

Mapping between DHS and MICS surveys

DHS and MICS surveys are similarly structured. A survey is made up of a number of questionnaires, each of which is divided into modules that contain questions on a similar topic or theme (Figure 2, Table 3). Countries choose which modules to include to meet their particular needs. While the architecture is shared, the grouping of questions into modules is different.

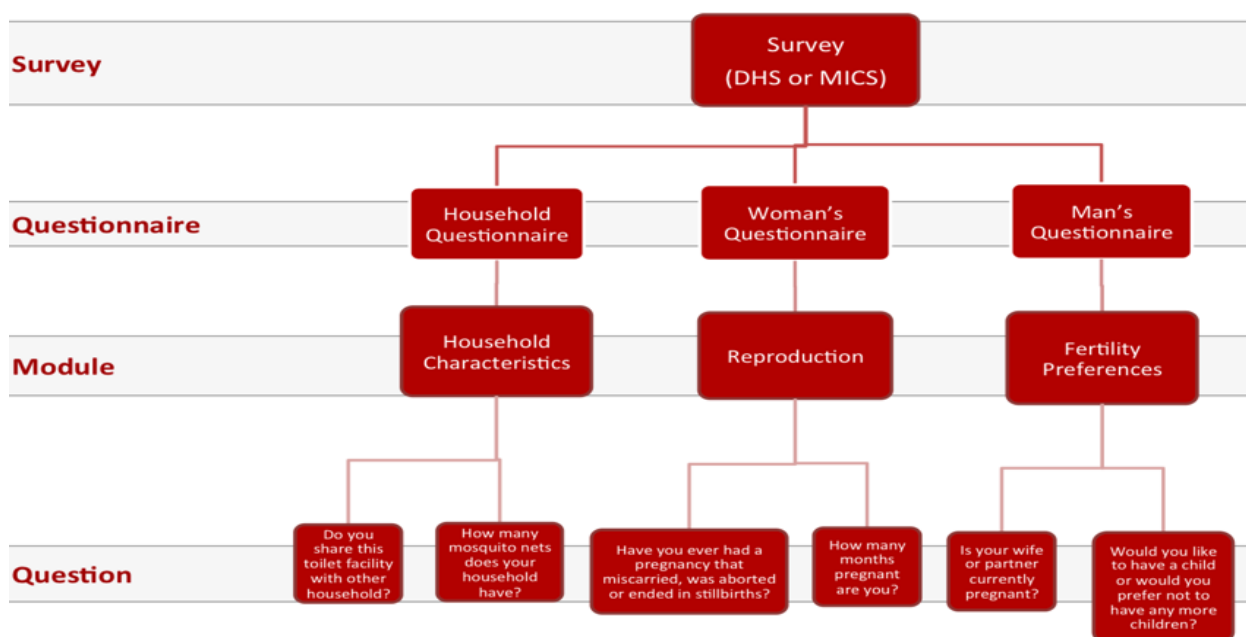


Figure 2. DHS and MICS survey architecture

		Modules	
		DHS-VII	MICS5
Questionnaires	Household	Household schedule, Household characteristics. Additional household characteristics, Mosquito nets	Household information panel, List of household members, Education, Child labour, Child discipline, Household characteristics, Insecticide treated nets, Indoor residual spraying, Water and sanitation, Handwashing, Salt iodisation
	Women	Respondent's background, Reproduction, Contraception, Pregnancy and postnatal care, Child immunization, Marriage and sexual activity, Fertility preference, Husband's background and woman's work, HIV/AIDS, Child health and nutrition, Other health issues	Woman's information panel, Woman's background, Access to mass media and use of information and communication technology, Fertility or fertility/birth history, Desire for last birth, Maternal and newborn health, Post-natal health checks, Illness symptoms, Contraception, Unmet need, Female genital mutilation/cutting, Attitudes toward domestic violence, Marriage/union, Sexual behaviour, HIV/AIDS, Maternal mortality, Tobacco and alcohol use, Life satisfaction
	Men	Respondent's background, Reproduction, Contraception, Marriage and sexual activity, Fertility preference, Employment and gender roles, HIV/AIDS, Other health issues	Man's information panel, Man's background, Access to mass media and use of information and communication technology, Fertility, Attitudes toward domestic violence, Marriage/union, Sexual behaviour, HIV/AIDS, Circumcision, Tobacco and alcohol use, Life satisfaction
	Children		Under five child information panel, Age, Birth registration, Early childhood development, Breastfeeding and dietary intake, Immunization, Care of illness, Anthropometry

Table 3. DHS and MICS modules

Methodology

We employ semantic mapping using the [Simple Knowledge Organization System](#) (SKOS) data model to classify and relate modules and questions across surveys. These mappings can be accessed through our [Online Thesaurus](#) and can be downloaded in the RDF/XML format. We focus on exact, close and broad/narrow matches.

Exact match

Exact matching describes two questions that are exact duplicates. For example, DHS-VII question “Does any member of this household have a bank account?” can be found in the exact same form in the MICS5 questionnaire. The machine-readable SKOS representation looks like this.

```
<rdf:RDF>
- <rdf:Description rdf:about="http://joinedupdata.org/Surveys/does-any-member-of-this-household-have-a-bank-account_hh_mics5">
  <skos:prefLabel xml:lang="en">
    does any member of this household have a bank account_hh_mics5
  </skos:prefLabel>
  <skos:broader rdf:resource="http://joinedupdata.org/Surveys/household-characteristics_hh_mics5"/>
  <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2016-05-09T11:31:16Z</dcterms:created>
  <skos:notation>HC15</skos:notation>
- <skos:altLabel xml:lang="en">
  Does any member of this household have a bank account?
</skos:altLabel>
<skos:narrower rdf:resource="http://joinedupdata.org/Surveys/yes_hh_mics5"/>
<skos:narrower rdf:resource="http://joinedupdata.org/Surveys/no_hh_mics5"/>
<skos:exactMatch rdf:resource="http://joinedupdata.org/Surveys/does-any-member-of-this-household-have-a-bank-account_hh_dhsq7"/>
</rdf:Description>
</rdf:RDF>
```

Close match

‘Close match’ indicates that the two matching questions differ only by slight rewording of a question. For example, DHS-VII: “Observe presence of soap, detergent, or other cleansing agent at the place for hand washing” has a close-match equivalent in the MICS5 question, “Do you have any soap or detergent or ash/mud/sand in your house for washing hands?”

```
<rdf:RDF>
- <rdf:Description rdf:about="http://joinedupdata.org/Surveys/observe-presence-of-soap-detergent-or-other-cleansing-agent-at-the-place-for-handwashing_hh_dhsq7">
  <skos:prefLabel xml:lang="en">
    observe presence of soap detergent or other cleansing agent at the place for handwashing_hh_dhsq7
  </skos:prefLabel>
  <skos:broader rdf:resource="http://joinedupdata.org/Surveys/additional-household-characteristics_hh_dhsq7"/>
  <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2016-05-09T09:58:51Z</dcterms:created>
  <skos:notation>I41</skos:notation>
<skos:closeMatch rdf:resource="http://joinedupdata.org/Surveys/do-you-have-any-soap-or-detergent-or-ash-or-mud-or-sand-in-your-house-for-washing-hands_hh_mics5"/>
  <skos:definition xml:lang="en">
    observe presence of soap, detergent, or other cleansing agent at the place for handwashing.
  </skos:definition>
  <skos:narrower rdf:resource="http://joinedupdata.org/Surveys/soap-or-detergent_hh_dhsq7"/>
  <skos:narrower rdf:resource="http://joinedupdata.org/Surveys/ash-mud-sand_hh_dhsq7"/>
  <skos:narrower rdf:resource="http://joinedupdata.org/Surveys/ash-mud-sand_hh_dhsq7"/>
  <skos:closeMatch rdf:resource="http://joinedupdata.org/Surveys/is-soap-detergent-or-ash-or-mud-or-sand-present-at-the-place-for-handwashing_hh_mics5"/>
</rdf:Description>
</rdf:RDF>
```

Broad match and narrow match

The DHS-VII question, “How many sons are alive but do not live with you?” is a narrower match to a MICS5 question, “How many sons are alive but do not live with you and how many daughters are alive but do not live with you?” Conversely the same MICS5 question is a broad match to the DHS-VII question from the example above.

```
<?xml version="1.0" encoding="UTF-8"?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:skos="http://www.w3.org/2004/02/skos/core#">
  <rdf:Description rdf:about="http://joinedupdata.org/Surveys/how-many-sons-are-alive-but-do-not-live-with-you_women_dhsq7">
    <skos:prefLabel xml:lang="en">
      how many sons are alive but do not live with you_women_dhsq7
    </skos:prefLabel>
    <skos:broader rdf:resource="http://joinedupdata.org/Surveys/reproduction_women_dhsq7"/>
    <dcterms:created rdf:datatype="http://www.w3.org/2001/XMLSchema#dateTime">2016-05-09T09:59:10Z</dcterms:created>
    <skos:broadMatch rdf:resource="http://joinedupdata.org/Surveys/how-many-sons-are-alive-but-do-not-live-with-you-or-how-many-daughter-are-alive-but-do-not-live-with-you_women_mics5"/>
    <skos:altLabel xml:lang="en">How many sons are alive but do not live with you.</skos:altLabel>
    <skos:narrower rdf:resource="http://joinedupdata.org/Surveys/sons-elsewhere_women_dhsq7"/>
  </rdf:Description>
</rdf:RDF>
```

Presentation of mapping

For each of the three common questionnaires we present a Sankey diagram generated from the mappings recorded in our [online thesaurus](#). Sankey diagrams are a type of flow diagram, in which the width of the arrows is shown proportionally to the flow quantity. **The diagrams summarise all the comparable matches** (exact, close, narrow/broad) between questions grouped by modules, as well as unique questions that cannot be mapped

Household questionnaire

The mapping between the household questionnaires shows a high level of alignment between MICS5 and DHS-VII surveys (Table 4, Figure 3). The duplicated modules in this questionnaire include questions relating to education, handwashing, water and sanitation, household roster and the use of insecticide-treated mosquito nets. The higher number of MICS5 unique questions is attributed to two modules: child discipline and child labour. DHS-VII (purple, bottom-left) has only five questions that cannot be joined up to the MICS5 household questionnaire.

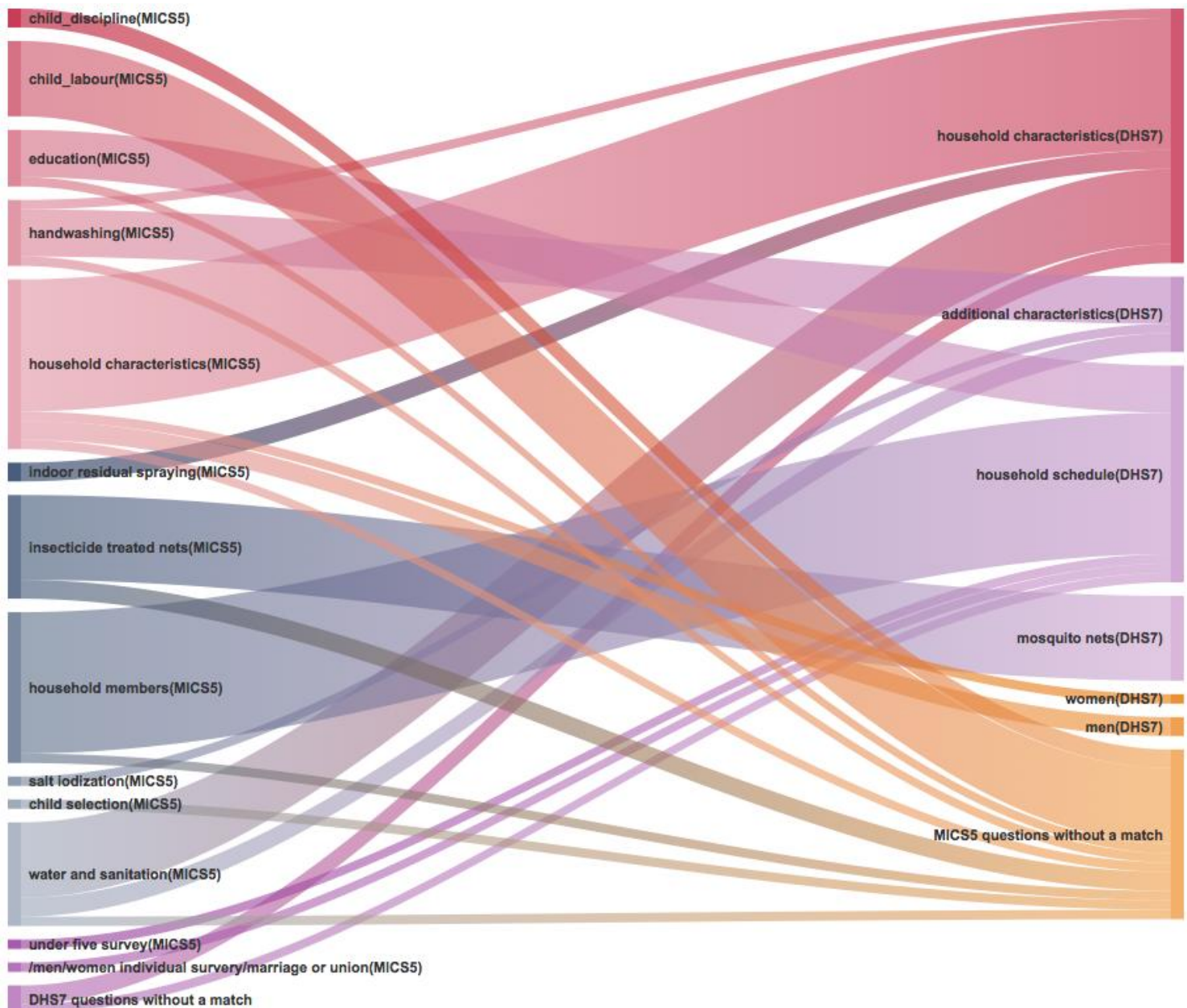


Figure 3. All mappings between DHS-VII and MICS5 household questionnaire

Notes: The questions unique to each of the surveys are grouped at the bottom of the diagram (purple for DHS and orange for MICS). The detailed mapping between questions can be accessed on the [online thesaurus](#).

Men's questionnaire

The mapping highlights a number of wide-spectrum questions being asked by the DHS-VII questionnaire (Table 4, Figure 4). These questions deal in detail with employment and gender roles that affect the financial situation of the surveyed household. The men are also questioned on their knowledge about their partners' fertile days, and fertility preference as well as children's health and antenatal care that a mother of a child received. Although these questions are also asked by the MICS5, they are posed in the women's questionnaire and not in the men's one. The unique module in MICS5 relates to subjective life satisfaction of the respondent and consumption of both alcohol and tobacco by the respondent. DHS-VII deals only with the consumption of tobacco.

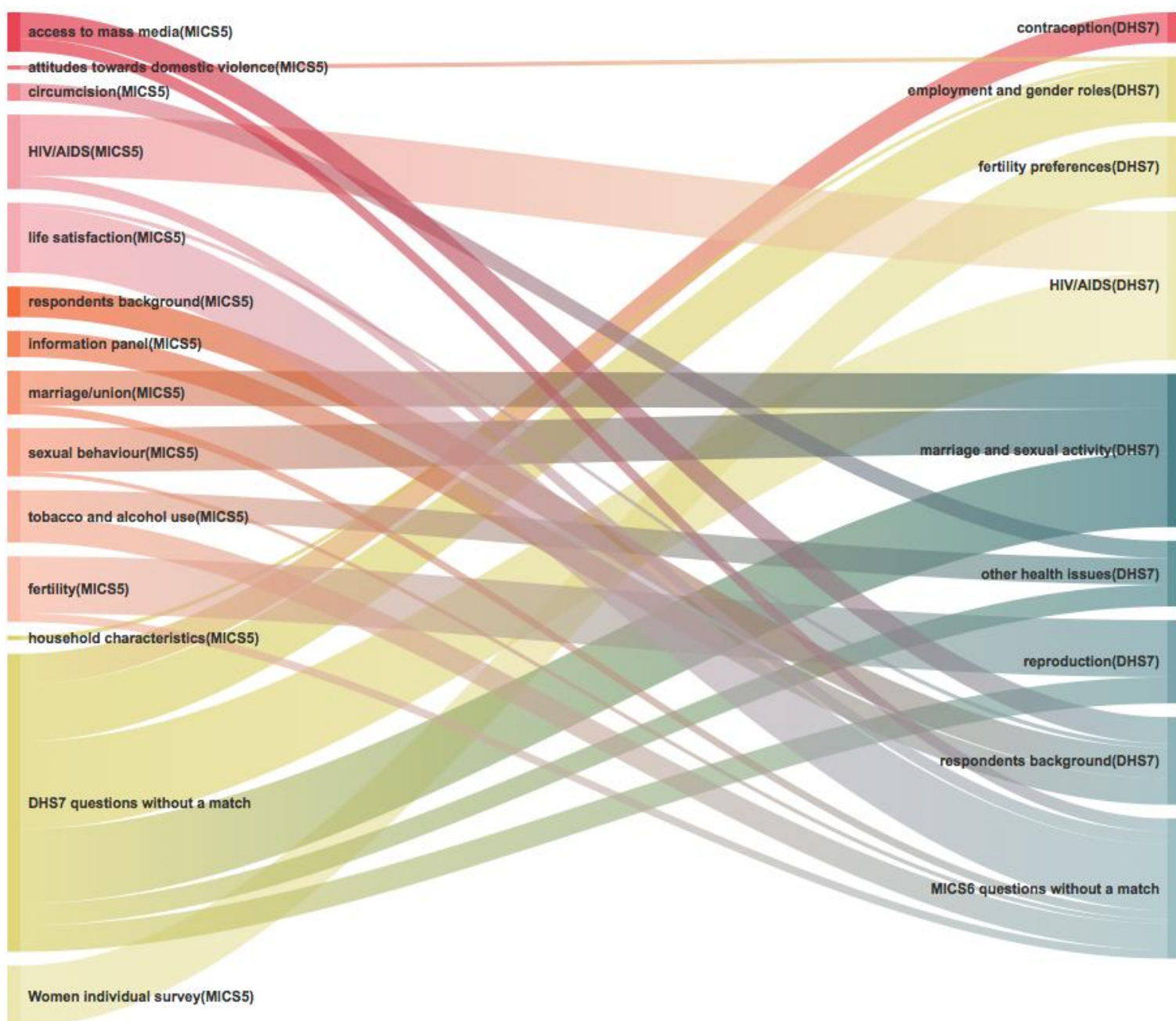


Figure 4. All mappings between DHS-VII and MICS5 men's questionnaire

Notes: The questions unique to each of the surveys are grouped at the bottom of the diagram (yellow for DHS and blue for MICS). The detailed mapping between questions can be accessed on the [online thesaurus](#).

Women’s and children under-five questionnaire

DHS asks women more detailed questions on gender roles, employment, and questions on contraception. MICS on the other hand enquires after subjective life satisfaction. DHS, as shown by Figure 5 and Table 4 includes the majority of the MICS questions but also expands on types of contraception, deals in more detail with family planning and goes beyond HIV/AIDS to cover other sexually transmitted infections (STIs).

MICS unique questions include modules on child development in the Under-Five Survey but, most importantly, enquire after every child living in a given household. DHS on the other hand focuses only on the children mothered by the respondent in the women’s questionnaire.

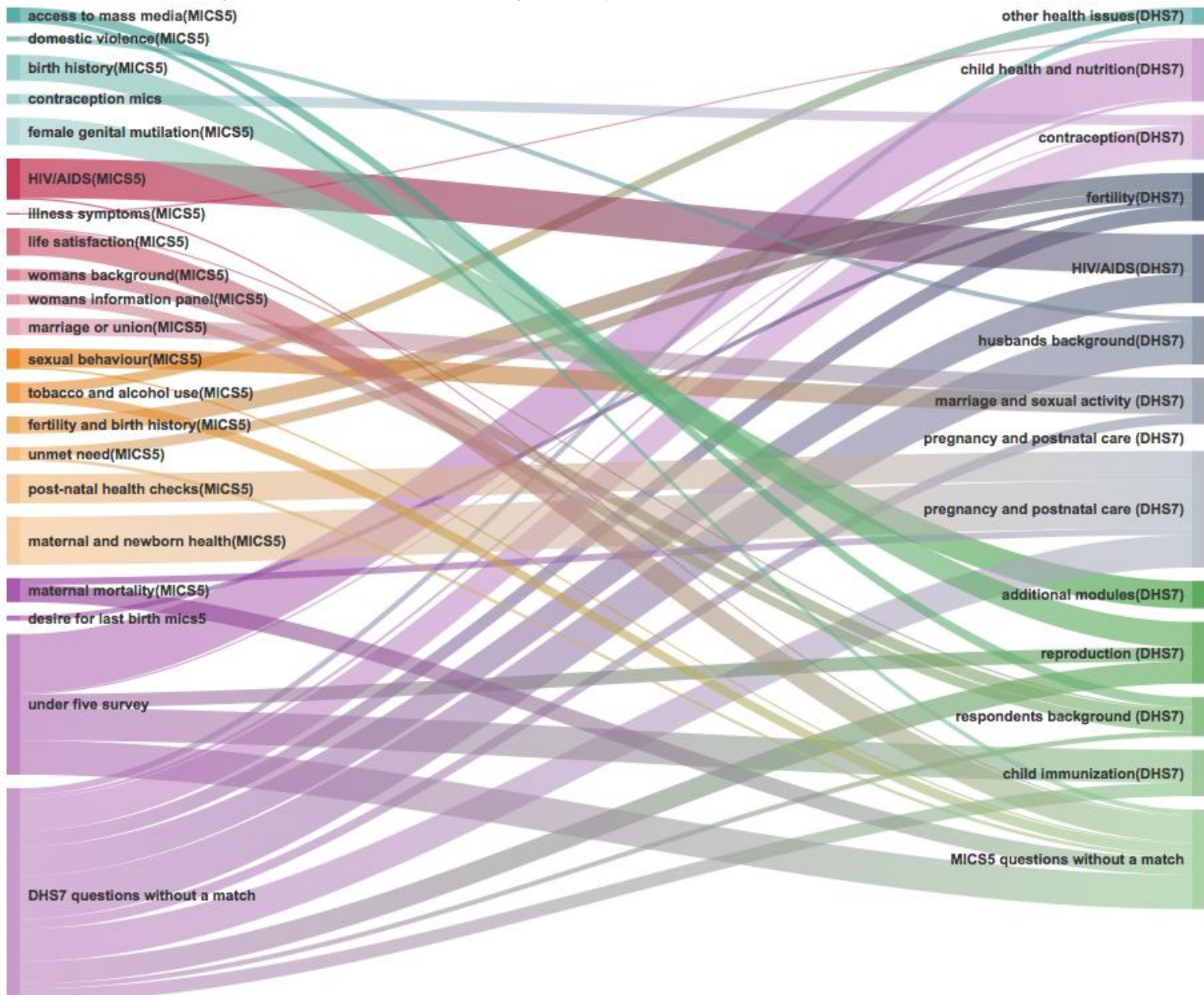


Figure 5. All mappings between DHS-VII and MICS5 for the women’s and under-fives questionnaire

Notes: The questions unique to each of the surveys are grouped at the bottom of the diagram (purple for DHS-VII and green for MICS5). The detailed mapping between questions can be accessed on the [online thesaurus](#).

Findings of mapping between DHS-VII and MICS5

On average, DHS-VII includes more questions in each questionnaire than MICS5 (Table 4, Figure 6). However, the majority of the core modules in these questionnaires are duplicated in both DHS-VII and MICS5 household surveys. There are questions that are unique to each survey but, as the questionnaires are designed in a modular way, questions from one survey can be incorporated as a foreign module in a different survey.

For example, the DHS survey conducted in Ghana (2011) contained MICS modules on child development, labour and discipline. Conversely, the unique feature of the DHS Program – biomarker surveying – can be incorporated in MICS surveys, as was the case in Sao Tome and Principe [2014 MICS survey](#).

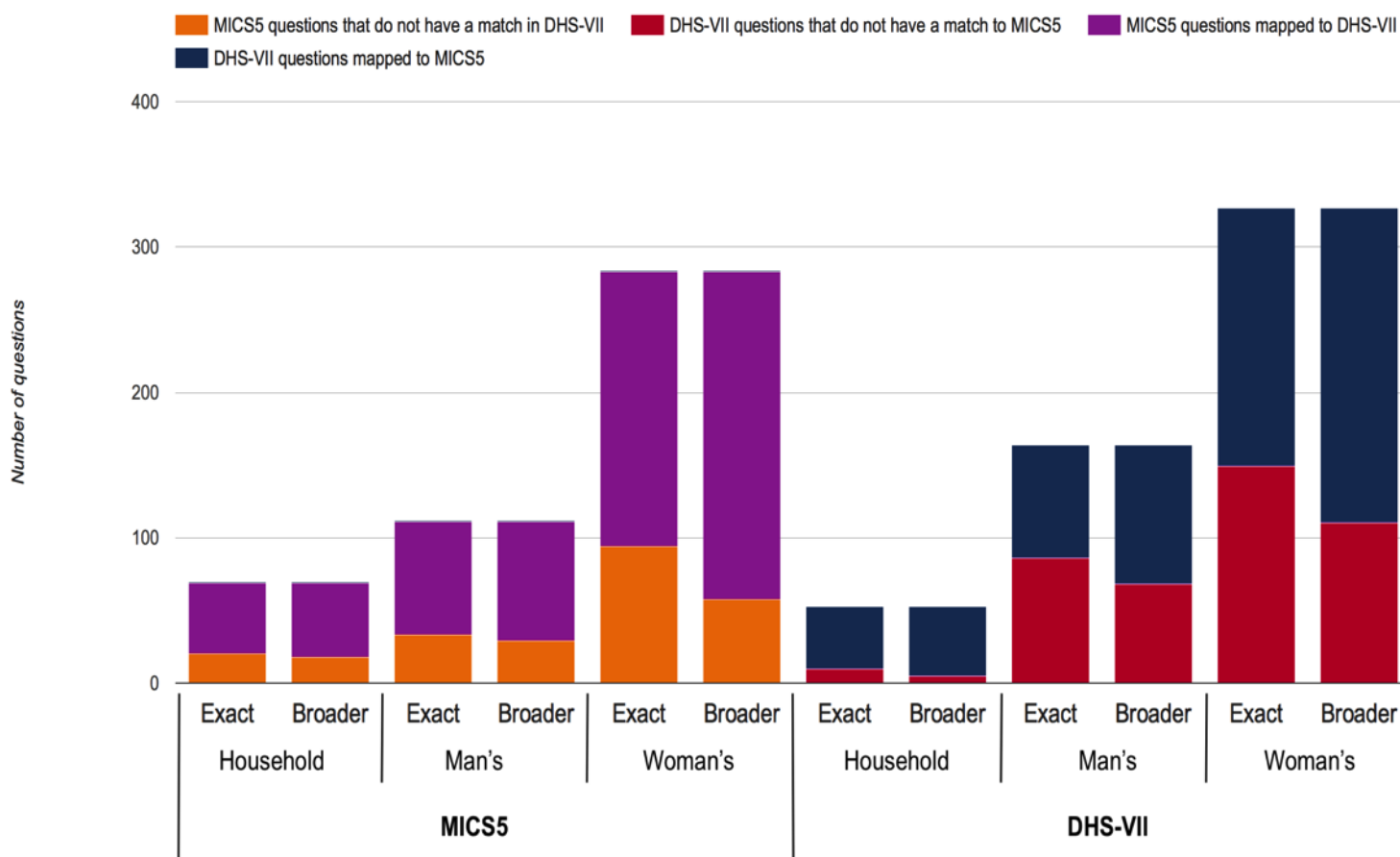


Figure 6. Summary of findings from the semantic mapping between DHS and MICS questionnaires^{vi}

Notes: The red (DHS) and orange (MICS) segments show the number of individual questions that are unique to the survey programmes. The purple (MICS) and blue (DHS) columns show the questions shared by both MICS and DHS.

Questions	Questionnaire						Total (Common Questionnaires)	
	Women		Men		Household		MICS	DHS
	MICS	DHS	MICS	DHS	MICS	DHS		
Total number of questions in questionnaire	283	327	111	164	69	53	463	544
Exact or close matches with counterpart	189	178	78	78	49	43	316	299
All matches with counterpart	225	217	82	96	51	48	358	361
Percentage of all matches with counterpart	80%	66%	74%	59%	74%	91%	77%	66%

Table 4. Number of DHS and MICS question matches across questionnaires^{vii}

Table 4 contains the most startling finding of this study: 77% of all MICS questions can be found in DHS, and 66% of all DHS questions can be found in MICS. However, since each survey follows its own coding standard for variables, it does not follow that duplicated questions can be easily matched. In terms of data analysis, this involves a laborious and confusing manual mapping exercise that requires an extensive knowledge of both systems and therefore makes it inaccessible for data-users.

How does a country decide which survey to conduct?

As shown in Table 2, African countries' choice of household surveys is varied. Since 2009, 20 have chosen DHS, 11 MICS and 15 both. It has not been easy getting answers to our questions concerning the thinking behind these choices. Costs, donor relations (including donor pressures), sample sizes and data needs are all factors.

The cost

The agency delivering a survey programme provides technical assistance, questionnaires, guidance and help to secure adequate funds to cover the cost of the survey.⁸ USAID contributes approximately two-thirds towards the funding of DHS surveys. Between 1984 and 2007 USAID invested \$380 million in the DHS programme and, according to the Agency, “each dollar leveraged approximately US\$0.33 in donor and host country contribution”.⁹ UNICEF, with MICS, prefers to contribute in the form of top-up funding, leaving the cost to be sorted out by the host country and its donors.

The cost of the survey depends on many factors, such as: the sample size of the surveyed population; the size of the survey itself (number of modules); and the level of technical assistance which is dependent on the statistical capacity of a given country. In a study prepared for the PARIS21 Task Force on Improved Statistical Support for Monitoring Development Goals,¹⁰ the authors estimated that the average cost per household per survey was \$185 for DHS and \$67 for MICS. The higher price of DHS surveys is largely down to the inclusion of its biomarker module, which tests respondents' blood samples. Table 5 shows the historical data on sample size for household surveys from 15 African countries that carried out both MICS and DHS. While cost is known to be a key factor in survey choice, the data available provides no discernible pattern.

[The World Bank study](#) that led to the formation of the new Intersecretariat Working Group on Household Surveys found that “significant cost-savings can also be achieved through enhanced coordination among donors and development partners. Duplicate and conflicting data collection activities abound, resulting in wasted funds and placing a heavy burden on national statistics offices and respondents.”

Survey design and sample size

The cost of surveys is directly related to the survey design, and in particular the sample size. This is determined by two dimensions: the number of sub-national 'domains' required for disaggregated reporting; and a calculation (Figure 7) that works out the number of households required in each domain sample in order to reach the target population.

What appears at first to be counter-intuitive to non-statisticians is that the sample size required for any country is not related to the total population of the country but rather to the number of domains (sub-national divisions) chosen. Table 5 lists the number of statistical domains used in our sample of 15 countries and compares this with the number of first- and second-level areas used to administer and govern the country. Surveys in Ghana, for example, are designed around its 10 administrative

regions. A survey that provided the same quality of data disaggregated for each of the country's 275 districts would require a 27-fold total sample size, and cost.

SAMPLE SIZE CALCULATION <i>for one domain</i>			
INPUT VALUES		OUTPUT VALUES	
Parameter	Value	Estimate	Value
Predicted value of indicator (in target/base population)	<i>r</i> 0.2	Predicted <i>r</i>	0.2
Design effect	<i>deff</i> 1.5	Confidence limits (at 95% confidence)	
Relative margin of error at 95% confidence	<i>RME</i> 0.12	Upper	0.224
Proportion of target/base population in total population	<i>pb</i> 0.15	Lower	0.176
Average household size	<i>AveSize</i> 5	Number of households (Sample size): <i>n</i>	2469
Response rate	<i>RR</i> 0.9	Standard error (se)	0.012
ADDITIONAL INPUTS		ADDITIONAL OUTPUTS	
Cluster size (Number of households per cluster)	20	Number of clusters	123
Typical target/base populations in total population		Expected numbers of completed observations	
Proportions of:		Effective number of households	2222
Women age 15-49 years	0.24	Number of household members	11111
Children age 0-4 years	0.15	Number of women age 15-49 years	2667
Children age 12-23 months	0.025	Number of children age 0-4 years	1667
Men age 15-49 years	0.27	Number of children age 12-23 months	278
Women age 15-49 with a live birth in last 2 years	0.25	Number of men age 15-49 years	3000
		Number of last live births in last 2 years	667

This template can be used to examine sample sizes for several candidate indicators. Changing the values as necessary in your country will automatically generate a new sample size (cell F10).
 Indicators with comparatively low values of *r* (cell C6) should be used only. Do not use indicators with *r* values that are higher than 0.4 or lower than 0.1.
 Do not change the cell contents that are red.

Confidence limits (95% confidence):
 Upper: $r * (1 + RME)$
 Lower: $r * (1 - RME)$

Sample size:
 $4 * r * (1-r) * deff$

$$n = \frac{4 * r * (1-r) * deff}{(RME * r)^2 * pb * AveSize * RR}$$

Standard error (se):
 $(r * RME) / 2$

Figure 7. Template used by DHS and MICS¹¹ to calculate required number of households per domain

Data needs: the case of Rwanda

As both MICS and DHS programmes recognise, the choice of survey should be based on data needs and data gaps for a given country.¹² Once the choice is made, the detailed contents of the survey's questionnaire are decided by the country government and in-country steering committee. As our research suggests, and both of the programmes emphasise, notwithstanding huge overlaps, the surveys differ in a number of particulars. For example:

- MICS includes questions on all the children within one household regardless of whether their mother lives in a surveyed household while DHS collects comprehensive data on children whose mother in the household
- DHS collects biomarker data.

In 2000 Rwanda carried out both MICS and DHS surveys. At that time the Statistics Department of Rwanda within the Ministry of Finance was still developing and had limited statistical capacity.¹³ However, after comparing the results from both surveys, the Statistics Department decided¹⁴ that its aspiration for its data users was to have one data source with harmonised variables, to avoid duplicating efforts, wasting resources or producing conflicting statistics that confuse data users.¹⁵ The National Institute of Statistics Rwanda (NISR) noticed the similarities between the surveys and decided, instead of conducting both MICS and DHS, to include the modules unique to MICS into the standard DHS survey. As a result, in 2005, the standard DHS model questionnaires were modified to include questions on orphans and vulnerable children (OVC).¹⁶

Since 2010 the standard DHS survey has been modified to include MICS modules on early childhood development, education and child labour. These modules were carried through to the DHS standard survey from the MICS survey. The NISR

combined modules from both survey programmes in the name of a DHS survey. This is an example of success in merging the two surveys for the benefit of harmonised data on population and health.

Country	Number of statistical domains	Sub-national administrative areas		Sample size															
		First level	Second level	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
Benin	12	12	77		5769					17511					17422			14077	
Cameroon	12	10	58	4435				10462		9667					14214			10213	
Cote d'Ivoire	10	14	31	7331						7600					9686				4368
Congo (Brazzaville)	12	12	93						5879				7096		11632			12811	
Congo (DRC)	11	26	192		8704						8886			11490			18171		
Equatorial Guinea	7	7	32	4284											4223				
Ghana	10	10	275				6251			6302		11778	5009		12150			11835	
Guinea	7	8	33						6282								7109		
Malawi	3	3	28	14213				13664		31200				24825		3404	28479	3405	
Mali	8	9	49		12331					12998			N/A			10105		3804	11830
Nigeria	6	37	774				7225				27750	34070		5000	29077	5000	38522		
Sao Tome and Principe	4	2	7	3254						5600		3536						3492	
Sierra Leone	4	4	13	4207					8000			7284		11923			12629		
Togo	6	5	31	4584						6600				6975			9549		
Zimbabwe	10	10	59						9285				12500	9756				17047	

Table 5. Sample size and number of statistical domains in different surveys, with number of country-specific sub-national administrative areas, 2000–2015¹⁷

Note: Red denotes DHS programme surveys, blue UNICEF/MICS surveys and grey World Bank LSMS survey

Summary of findings

In many countries the international household survey programmes are the most important source of socioeconomic, demographic and health data in the absence of national registry and administrative systems. In Africa the two most common household surveys conducted are DHS and MICS.

Historically, each had a purposeful origin: DHS originally monitored world fertility and contraception prevalence while MICS captured the status of children. As the surveys evolved and worked together on harmonisation of common practices and associated tools, they began to mimic one another.

As our efforts to map MICS and DHS suggest, this co-evolution led to the merging of the core questions on essential development indicators resulting in the duplication of questions across these two survey programmes. The duplicated questions, however, remain coded differently and as a result merging the datasets is labour intensive and costly for countries conducting the survey, and confusing for the data users worldwide. Countries such as Rwanda and Malawi have instituted their own integration of DHS and MICS surveys to meet their particular needs and promote the continuity of the data.

There is political consensus within the statistical community, as evidenced by the proceedings of the last two sessions of the UN Statistical Commission, that the data from household surveys should be comparable between datasets, between countries and over time, and compliant with international standards. Host countries should not have to merge and pick and choose questions from different sources; it is the role of the international household survey programmes to anticipate such needs and provide solutions.

Conclusions

Integration of household surveys is no longer optional; it is now necessary. We have highlighted the extent of the duplication between the DHS and MICS, the costs incurred by both international donors and the host country of the surveys, and the urgent need for comprehensive and quality data on population and health in developing countries. Semantic mapping of the international household surveys, as demonstrated in this paper, can offer a temporary solution to this problem but does not resolve the need for an integrated survey and governance structure.

The Joined-up Data Standards translator currently allows for a machine-readable translation between MICS5 and the DHS7. With assistance from the survey-setters, more key household survey data, including other MICS/DHS versions and related LSMS modules, could be mapped and made more useful. The harmonisation of the core questions across household surveys and clearly defined subject-specific modules in reducing the cost of the surveys per country, along with standardisation of the coding system between the surveys, is a crucial first step to creating a transparent and easy-to-use household survey framework.

The [‘Report of the World Bank](#) on improving household surveys in the post-2015 development era: issues and recommendations for a shared agenda’, presented at the 46th session of the UN Statistical Commission recognises that “large disparities remain across countries and surveys” and that “International databases of key socioeconomic indicators derived from household surveys demonstrate persistent and significant gaps and weaknesses”.¹⁸ The report also highlights how this can

contribute to the “poor coordination of international support”, “unpredictable funding” and a “lack of globally accepted methodological standards and methods for measurement of the key socioeconomic indicators”.

The World Bank made a number of concrete recommendations to the UN Statistical Commission for creating an institutional framework to coordinate efforts on the harmonisation of standards among development partners, implementing a common international code of practice on household surveys and developing a coordinated programme of research into improved standards, methods and practices in household surveys. This has resulted in the formation of the Intersecretariat Working Group on Household Surveys (IWGHS) which aims to promote coordination and cooperation in the planning, funding and implementation of household surveys (Figure 8).¹⁹

A first priority for the Working Group will be to make progress on an international code of practice for household surveys and a task force will be established to develop and pilot the standards, based on the World Bank’s proposals. However, further task forces will be required to examine the additional priorities listed in the World Bank’s report.²⁰

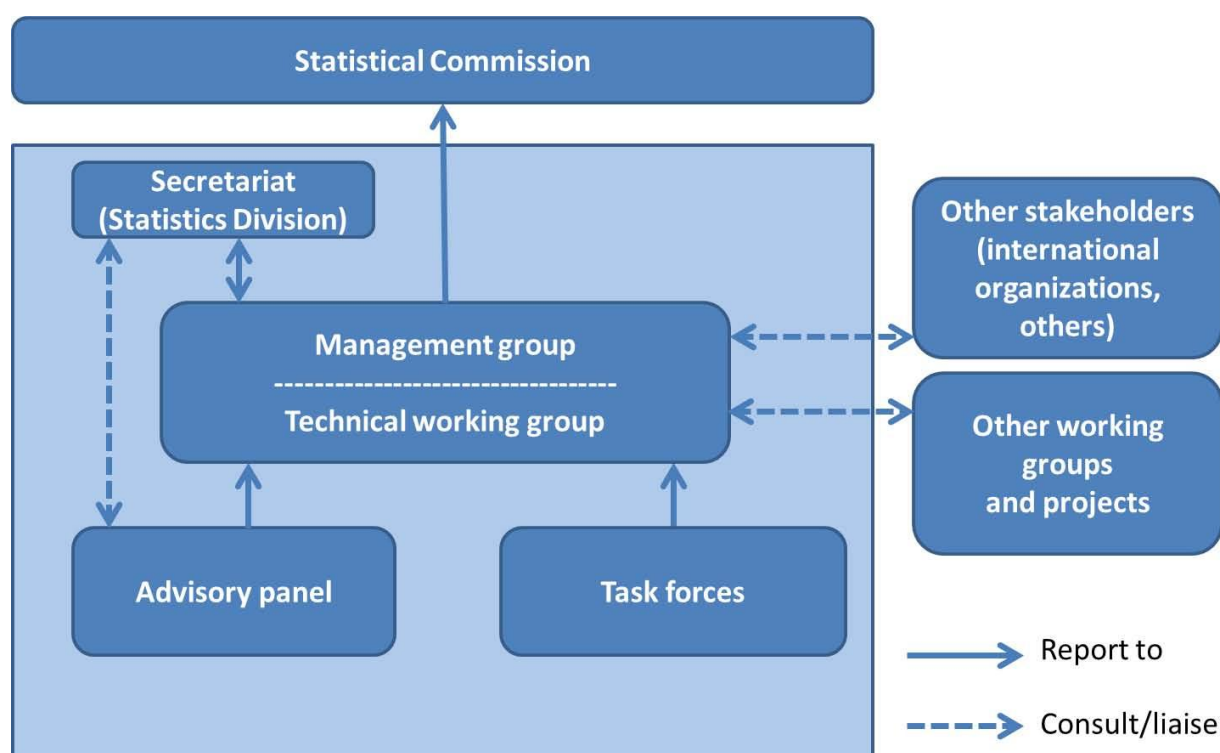


Figure 8. Structure of the new Intersecretariat Working Group on Household Surveys²¹

In addition to the IWGHS, UNICEF, USAID and the World Bank announced the establishment of a collaborative group to “1) share information on the scheduling of surveys at the country level, 2) foster further harmonisation of survey tools, and 3) work together on new methodological advances in household surveys”²². They are also working together in a working group that is part of the new WHO-initiated [Health Data Collaborative](#).

While we welcome these first steps, it will be critical, as our research indicates, for development partners to maintain their ambition and commitment to making

demonstrable progress in this area in order to harness the momentum generated in this field since March 2015.

Recommendations

1. A Single taxonomy

The IWGHS should begin work immediately facilitating the creation of a curated, publicly accessible taxonomy which contains a normalised superset of all questions (and structured answers) in all household survey questionnaires.

2. Official cross-mapping

The providers of survey data should begin work immediately on a cross-mapping service that enables users of survey data to seamlessly join up data between all survey datasets, between countries and across time.

3. An integrated governance structure

The IWGHS should play a decisive role in leading the development of a truly integrated approach. To achieve this, it needs to proactively ensure that all survey providers and representatives of national statistics offices are included in its work and structures.

4. An integrated survey

The IWGHS should commit itself to the realisation of a single, unified and standardised household survey where the resources and expertise of all institutions are pooled to ensure the most beneficial and cost-effective outcome for developing countries.

References

- ⁱ DHS is funded by USAID and implemented by ICF International, a private company.
- ⁱⁱ The World Bank's LSMS programme also includes questionnaires covering the same terrain, but its overall scope is broader.
- ⁱⁱⁱ From research by data scientists at Development Initiatives working on global poverty data
- ^{iv} Collated from DHS, MICS and International Household Survey Network (IHSN) websites
- ^v Collated from DHS, MICS and IHSN websites
- ^{vi} Analysis based on author's semantic mappings in the online thesaurus
- ^{vii} Author's analysis
- ⁸ <http://www.uis.unesco.org/Library/Documents/hhsguide04-en.pdf>
- ⁹ <http://www.who.int/bulletin/volumes/90/8/11-095513/en/>
- ¹⁰ <http://www.paris21.org/sites/default/files/monitoringMDG-household-full.pdf>
- ¹¹ <http://mics.unicef.org/files?job=W1siZiIsIjIwMTUvMDEvMTQvMDYvMDIvMzAvNDA0L0VvZ2xpc2hfTUIDU19TYW1wbGVfU2I6ZV9DYWxjdWxhdGlvbI9UZW1wbGF0ZV8yMDEzMDQyMS54bHN4Il1d&sha=c2d7fcd734ce4e31>
- ¹² <http://mics.unicef.org/workshops#survey-design-workshops>
- ¹³ <http://www.statistics.gov.rw/publication/strategic-plan-development-national-statistical-system-2003-2007>
- ¹⁴ <http://www.statistics.gov.rw/publication/first-national-strategy-development-statistics-2009-2014> and <http://www.statistics.gov.rw/publication/second-national-strategy-development-statistics-2014-2018>
- ¹⁵ This information was provided by Dominique Habimana, Director, Statistical Methods, Research and Publication, National Institute of Statistics of Rwanda.
- ¹⁶ <http://dhsprogram.com/pubs/pdf/FR183/FR183.pdf>
- ¹⁷ Statistical domains and sample sizes from DHS and MICS websites. Administrative subdivisions from Wikipedia
- ¹⁸ "despite decades of technical and financial assistance, large disparities remain both across countries and surveys, with many countries still unable to sustain a long-term programme of quality surveys that are comparable over time and compliant with international standards. International databases of key socioeconomic indicators derived from household surveys demonstrate persistent and significant gaps and weaknesses" (<http://unstats.un.org/unsd/statcom/doc15/2015-10-HouseholdSurveys-E.pdf>).
- ¹⁹ Resolution of 46th session of the UN Statistical Commission, March 2015 (<http://unstats.un.org/unsd/statcom/46th-session/documents/statcom-2015-46th-report-E.pdf>).
- ²⁰ Report of the IWGHS to the 47th session of the UN Statistical Commission, March 2016 (<http://unstats.un.org/unsd/statcom/47th-session/documents/2016-21-ISWG-on-household-surveys-E.pdf>).
- ²¹ <http://unstats.un.org/unsd/statcom/47th-session/documents/2016-21-ISWG-on-household-surveys-E.pdf>
- ²² Establishment of a Collaborative Group among the DHS, MICS and LSMS, May 2015 (http://siteresources.worldbank.org/INTLSMS/Resources/DHS_MICS_LSMS_CG_Announcement.pdf).