Technology and the data revolution: lessons from the MY World survey

Tanvi Bhatkal

Key messages

- Various recent initiatives have used information and communication technology to collect data more efficiently and from more people. MY World, an ongoing survey of people's priorities, has used offline, mobile phone and online surveys to collect over 7.5 million responses from 194 countries.
- Mobile phone and online surveys over-represent the youth and people with secondary or higher educational attainment. They under-represent people aged 40 and above.
- About 47% of total responses are from girls and women, but only 26% of mobile phone responses. Relatively fewer female responses, compared to men, are recorded in low and medium Human Development Index (HDI) countries, but the imbalance does not vary systematically with education level or age.
- MY World attempts to include as many people as possible and is not intended to be statistically representative. Careful study is needed to understand better who such crowd-sourced initiatives are reaching and who they are not, and what further efforts may be needed to tackle imbalances.

Acknowledgements

The author is grateful to Emma Samman for valuable guidance and comments on various drafts of this paper. Many thanks also to Claire Melamed, Elizabeth Stuart, and Johannes von Engelhardt for useful discussions and helpful comments to earlier versions of the paper. In addition, she thanks Frances Simpson Allen and Anand Kantaria for various inputs on the implementation of the MY World survey. Editorial support was provided by Kofo Sanusi and Ben Tritton. The usual disclaimers apply.

This working paper is an output of the following project: 'Building a post-2015 sustainable development agreement'[1]. Details of current funders of the project can be found online.

[1] http://www.odi.org/projects/2730-building-post-2015-sustainable-development-agreement

Table of contents

1 Introduction

Recent and reliable data are critical to monitor progress, understand people's priorities, and design policies to improve development outcomes. Disaggregated data can help to assess whether development is 'leaving no one behind', and to formulate policies targeted to the poorest. The importance of subjective data, which establishes people's priorities and their perceptions of progress, is also increasingly recognized (United Nations, 2014; Rodríguez Takeuchi et al, 2015). Equally, data can empower citizens and enable them to demand accountability from governments and other actors.

However, vast challenges exist. Official national statistics often exclude some of the poorest. Up to 350 million people may be omitted from household surveys and many censuses – a considerable share of whom are likely to be poor[2] (Carr-Hill, 2013). In addition, many statistics are estimated on the basis of sparse data and limited civil registration. For instance, of 181 countries for which maternal mortality figures are available, half rely on survey data in the absence of complete civil registration systems, and another 15% have no nationally representative data (Wilmoth et al, 2012).

Recognising these problems, data gaps in developing countries have emerged in discussion over the Sustainable Development Goals, which will supersede the Millennium Development Goals in 2015. In 2013, the UN Secretary General's High Level Panel of Eminent Persons on the Post-2015 Development Agenda called for a 'data revolution'[3] to improve the quality of statistics and information available to citizens (United Nations, 2013). Since then, this call has attracted considerable attention. More recently, the report of the UN's Expert Group on the Data Revolution (Independent Expert Advisory Group, 2014) stimulated increased debate over new ways of collecting and using data.

Optimism is growing that new technologies and their increasing penetration are increasing the volume and types of data exponentially. However, the use of technology poses constraints too – namely the risk of replicating inequalities in access to technology and under-representing those already undercounted (Stuart et al, 2015).

This paper seeks to evaluate the potential of new technologies to collect data on people's perceptions, and the associated biases. Focusing on MY World, a global survey of people's priorities developed by the United Nations, Overseas Development Institute and Ipsos MORI, it draws out some lessons for data collection using new technologies by comparing the socio-demographic characteristics of respondents based on mode of survey administration.

Section 2 outlines some of the ways in which technology has been used to collect data recently. Section 3 describes the MY World survey. Section 4 outlines the modes through which the MY World survey was conducted and explores some of differences in respondents' socio-demographic profiles based on survey mode. Section 5 concludes by situating the MY World experience alongside other evidence on access to, and cost of, technology.

[2] This includes the homeless, people living in institutions (including refugee camps), mobile populations, fragile and multiple occupancy households, slum populations, and those in areas posing security risks.

[3] A data revolution as 'an explosion in the volume of data, the speed with which data are produced, the number of producers of data, the dissemination of data, and the range of things on which there is data, coming from new technologies such as mobile phones and the "internet of things", and from other sources, such as qualitative data, citizen-generated data and perceptions data' (IEAG, 2014).

2 Leveraging technology for data collection

Technology offers the potential to reduce the time and cost of data collection and improve accuracy. This section outlines various ways in which technology has been used to collect data, to provide context for the analysis of MY World responses.

2.1 Electronic devices in household surveys

Census and household surveys are the main source of reliable and disaggregated data at the national level and will remain a cornerstone of data collection. However, the various stages of checking and data entry following paper-and-pencil surveys are time-consuming and expensive.

The adoption of electronic devices for data collection has helped bring down ongoing costs by eliminating data entry and checking operations. For instance, the use of mobile phones in a 2010 survey in Guatemala led to a cost reduction of over 70% per interview relative to paper-based surveys, which allowed the sample size to be increased from 200 to 700 (Schuster and Perez Brito, 2011). Mobile electronic equipment ranging from mobile phones and PDAs to laptops has become relatively cheap to procure.

Computer aided personal interviews (CAPI), popular in developed countries for allowing the immediate transfer of data from the field to central offices, is increasingly being used for census and household surveys in developing countries. In the 2010 census round, GPS/GIS technology was used by about three quarters of the 138 countries for which data are available (UN DESA, 2010).

CAPI improves data quality as devices can be programmed to conduct validation checks and prompt enumerators in cases of inconsistent answers. Moreover, they improve accuracy compared to paper-based surveys in reducing transcription errors and chances of losing surveys in transit (Patnaik et al, 2009; Trucano, 2014). They are also helpful in complex surveys where questions can automatically adapt to respondents' answers, where sections need to be skipped (Goldstein et al, 2012), or allow for language changes. As a result, interviews conducted using electronic devices take less time than their paper-based counterparts (Schuster and Perez Brito, 2011; Caeyers et al, 2011). They also increase accountability as built in mechanisms such as GPS stamps can verify that interviews have been conducted.

Electronic devices can also be used to collect new types of data using sensors, cameras and mapping software. Surveys can be enhanced through the collection of visual and media data such as photographs and videos, for instance to document the condition of assets and public amenities. Surveyors could also carry portable ECG monitors, blood pressure monitors or blood glucose meters that connect to smartphones to measure the health status of residents. Similarly, sensors can measure air and water quality around settlements.

Much of this is already being done in several developing countries by official statistical agencies and even by NGOs and the private sector (see Stuart et al, 2015). At the local level in several cases – from India to Kenya to Brazil – urban slum communities have come

together to map their settlements using cameras and mapping technologies (Lucci and Bhatkal, 2014).

However, field staff need to be well trained so that managing the device does not impede their interaction with respondents (Escobala and Benites, 2013). This increases training time and costs (Measure DHS/ ICF International, 2012), although training needs are lower as familiarity with mobile phones increases. Moreover, regular syncing and backup servers need to be in place so data are not lost in case of device theft or software issues. It is also essential that data are encrypted and password protected following collection, during transmission and storage, to address data security and privacy concerns.

2.2 Mobile phone surveys

The rapid growth of information and communication technologies (ICTs) over the past two decades – with the number of mobile phone subscriptions per 100 people globally increasing from 12 in 2000 to 93 in 2013 (Table 1) – presents an opportunity to improve the frequency of data collection and reduce costs. Although internet penetration rates remain relatively low in developing countries, mobile phone surveys have enabled high frequency data collection on a range of subjects. This is particularly useful in countries with weak statistical institutions and infrequent official data but high mobile phone penetration.

Instead of enumerators physically visiting households, respondents are contacted directly on their mobile phones to answer survey questions. Mobile phone surveys can be conducted in different ways – via short messaging service (SMS), interactive voice response (IVR), web-based surveys, and through computer assisted telephone interviewing (CATI) where enumerators at call centres interview people by phone and record their answers on a computer.

Table 1: Penetration of information and communication technology (users per 100 people)

Region / income category	Mobile phone subscriptions[4]				Internet users[5]			
	1990	2000	2010	2013	1990	2000	2010	2013
High income	1	44	115	120	0	27	69	78
Middle income	0	5	74	92	0	2	23	33
Upper middle income	0	8	77	100	0	3	33	45
Lower middle income	0	1	71	85	0	1	13	21
Low income	0	0	33	53	0	0	4	7
Latin America & Caribbean	0	12	96	115	0	4	35	47
Europe & Central Asia	0	34	120	125	0	13	56	65
East Asia & Pacific	0	11	74	98	0	6	34	44
South Asia	0	0	60	71	0	0	7	14
Sub-Saharan Africa	0	2	45	66	0	1	10	17
Middle East & North Africa	0	5	93	110	0	2	27	39
World	**0**	**12**	**77**	**93**	**0**	**7**	**29**	**38**

Source: World Bank, World Development Indicators

These methods eliminate transport costs and time. The cost of phones, phone credit and, in the case of CATI, interviewers and an office, are the main expenses. Cost savings are considerable. For instance, the *Listening to LAC* project in Peru found that compared with $40 per interview for face-to-face surveys, mobile phone surveys cost $8 for SMS, $17 for IVR and $25 for voice calls (Ballivian and Azevedo, 2013). Similarly, the *Sauti za Wananchi* project in Tanzania recorded significant savings through mobile phone interviews. Face-to-face baseline interviews cost $50-100 per survey, while mobile interviews cost $4.10-7.30, although the latter were very short compared to the extensive baseline surveys (Croke et al, 2013).

Although the most expensive of the three types of mobile surveys, CATI is more appropriate in places with lower literacy levels and when populations speak different languages (Croke et al, 2012; Ballivian and Azevedo, 2013). It also has the highest retention rates in panel surveys (discussed in Section 5).

[4] Mobile phone subscriptions are estimated as the number of (pre-paid and post-paid) subscriptions to a public mobile telephone service using cellular technology.

[5] Internet users are defined as users with access to the worldwide network.

The feasibility of using these modes depends in part on the length of the survey. For instance, CATI surveys have been found to be effective only for surveys that last less than 30 minutes (Croke et al, 2012). SMS surveys can track changes with a high frequency but are only suited to very short polls with around 5-10 questions, and may require more extensive baseline surveys. These practical aspects need to be considered when choosing which technology to adopt.

2.3 Estimating welfare using 'big data'

The growing use of mobile phones and the internet has itself produced large volumes of data that can help estimate poverty and wellbeing.

Big data[6], which are digitally generated and passively collected during people's use of digital services, produces data that can be analysed in near real time. For instance, data exhaust – transactional data from the use of digital financial, communications or information services[7] – has been used to estimate the level and geographic dispersion of poverty and wellbeing in developing countries. Data from mobile phone call detail records[8] in particular can and have been used to estimate income and socioeconomic disparities, mobility patterns, and the early impacts of economic crises and livelihood programmes (Global Pulse, 2013). Similarly, physical sensors such as satellites and infrared imagery remotely capture data on human activity and are increasingly being used to estimate growth, poverty, urban development, housing deficiencies, and resource availability.

However, inequities in access to technology may compromise the reliability of big data in the development context (discussed in Section 5). Moreover, analysis of big data needs to be sensitive to the cultural and ethnographic context as this kind of data can lead to *apophenia* (the tendency to see patterns where none exist), a consequence of the massive quantity of data that may offer connections in all directions (Boyd and Crawford, 2011). For this reason, research using big data should be transparent about its methodology, analytical assumptions and underlying biases.

Perhaps more importantly, big data poses privacy challenges as users of digital services may be unaware that their data is being collected and used. A particular concern with mobile call detail records is the possibility to re-identify anonymised data (Narayan and Shmatikov, 2008). These privacy challenges are exacerbated in poor and fragile states where institutions may be weak or identification may pose security risks.

Overall, the rapid surge in the production and availability of big data in recent years offers to help measure development progress and identify shortfalls. However, its accompanying challenges need to be understood, data need to be better anonymised, and legal frameworks and institutional capacity need to tackle privacy issues.

[6] Big data broadly refers to the exponential growth in the availability of data (both structured and unstructured), with datasets too large and complex to process using traditional database management tools or applications.

[7] Data exhaust includes (anonymised) data on online purchases, money transfers, loan repayments, records of mobile phone usage and airtime purchases, and search queries on online search engines.

[8] Call detail records (CDRs) are generated for every phone call, text message or mobile airtime purchase and track the duration and approximate location of the caller and receiver of each transaction.

3 MY World

Numerous efforts over the past decade have deployed technology to measure development progress, and people's perceptions and priorities. International and local organisations have exploited its potential through a combination of modes, to offset the weaknesses of one with the strengths of another, and to reduce overall data collection costs (Roberts, 2007). For instance, face-to-face surveys employing rigorous sampling methodologies have been used as the baseline for subsequent mobile phone panel surveys. Others have attempted to broaden their reach by seeking responses through different modes simultaneously.

MY World has been one such initiative – it has utilised offline, mobile phone and online surveys method to encourage a maximum number of responses. Designed to be as open as possible, this 'mixed modes' method of data collection has helped reach many people– so far, over 7.5 million[9] people from 194 countries. This section describes the aims and scope of MY World, while the following sections report on who the survey has managed to reach through each mode of administration.

MY World is an ongoing global survey that aims to understand the development priorities of people from across the world. It was created to inform formulation of the Sustainable Development Goals and its findings were considered by the UN Secretary General's High Level Panel for Post-2015 when making recommendations for the new development framework (United Nations, 2013). The survey results were also a key input to the Secretary General's Synthesis Report on the Post-2015 Development Agenda (United Nations, 2014), which sought to guide the final round of negotiations for the Sustainable Development Goals.

The survey asks respondents one question[10]: which six out of 16 development priorities are most important to them and their families (see Box 1). The 16 options were selected based on the priorities poor people expressed in existing research and polling exercises, as well as on technical and political discussions about possible Sustainable Development Goals. Respondents also have the option of including one self-selected priority beyond the list.

[9] Data for analysis downloaded on 9 February 2015. MY World had 5,786,490 responses and further to data cleaning 5,655,335 responses were used in this analysis.

[10] The offline and online versions presented 16 options, requiring respondents to select six (or five and a self-selected sixth) priorities. The mobile phone surveys were adapted in line with the technology – respondents were asked the question four times, and required to select one out for four priorities in each instance. The responses to these four questions were then analysed to determine which six out of the 16 options they would have selected based on likelihood estimations.

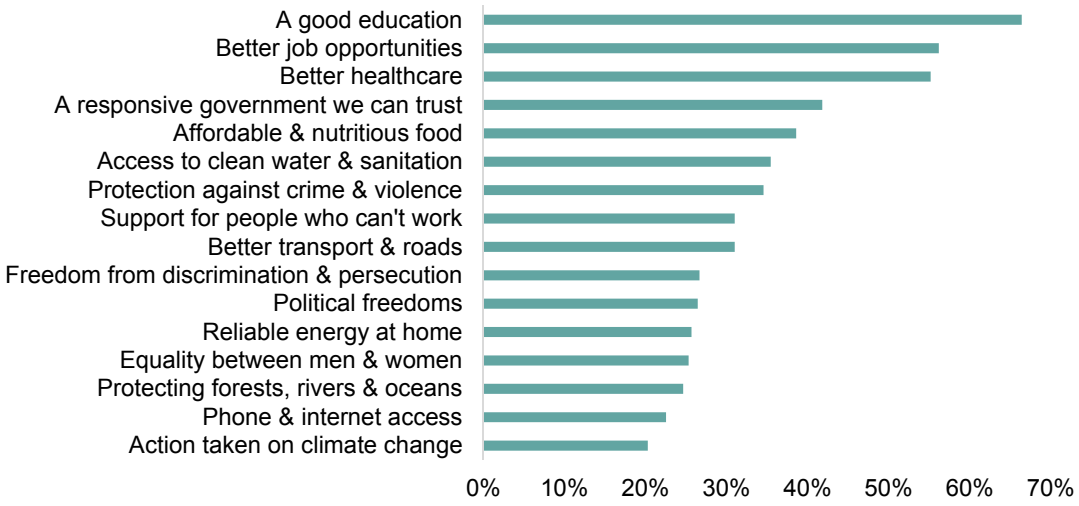
Box 1: The MY World survey

MY World asks people to identify six out of 16 possible issues they think would make the most difference to their lives and those of their families:

A good education	Support for people that cannot work
Better healthcare	Reliable energy at home
Better job opportunities	Freedom from discrimination and persecution
An honest and responsive government	Better transport and roads
Protection against crime and violence	Protecting forests, rivers and oceans
Affordable and nutritious food	Political freedoms
Access to clean water and sanitation	Phone and internet access
Equality between men and women	Action taken on climate change

Globally, the leading priorities are a good education, better job opportunities and better healthcare (Figure 1). Although differences exist across countries and based on age, gender and educational attainment (which stands as a proxy for income) – and these differences are important to note (see, for instance, Melamed and Ladd, 2013; Bergh et al, 2014; Bhatkal, 2014; Rodriguez Takeuchi, 2014) – the aggregate results indicate that the unfinished agenda of the Millennium Development Goals remains important. Equally, the future agenda should include issues relating to livelihoods and governance.

Figure 1: MY World global responses



Source: MY World 2015 (data accessed 9 February 2015)

The survey has sought to achieve a balance between scale and representativeness, by supplementing a global survey into which people self-select with representative household surveys in several low and middle income countries across Asia and Africa.[11]

Overall, however, the MY World survey is not intended to be representative statistically. Social and demographic data collected on respondents (age, educational attainment, gender,

[11] Representative surveys were conducted by partners IPSOS Mori in India (urban), Indonesia and Philippines in Asia and by UNDP in Ghana, Kenya, Liberia, Mozambique, Nigeria, Senegal and Uganda in Africa.

and country) give some insights into how well the sample from each country reflects its population and also how priorities vary across geographies and population sub-groups. Analysis of these data helps to illustrate the opportunities offered by and the constraints faced when using technology for data collection.

We use these data to explore 'who' responds to the survey based on survey mode. A separate concern involves 'how' people respond based on mode – mode effects may affect either the amount of effort people make in answering or their willingness to answer honestly (Roberts, 2007). While important, this issues goes beyond the scope of this paper.

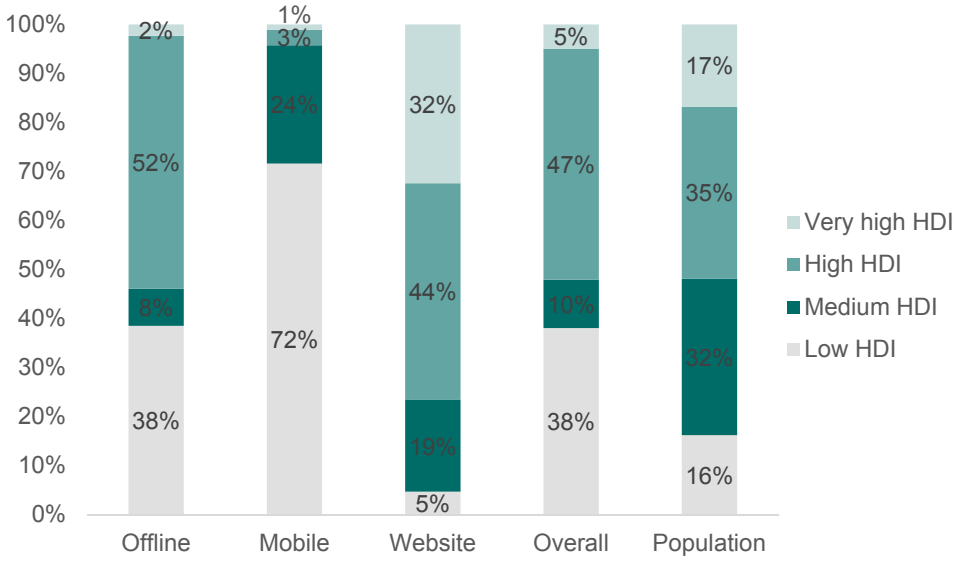
4 MY World results: who responds?

The MY World survey has been available online in the six UN languages since late 2012. A majority (83%) of MY World responses have been collected offline through paper-and-pencil based surveys conducted by grassroots organisations, faith based communities, youth groups, private sector bodies and NGO partners across the globe. It was also disseminated by mobile phone (using SMS and interactive voice response) starting January 2013 in collaboration with telecom partners in various countries. Although the share of mobile phone (8%) and online (9%) surveys is relatively low, together they account for nearly a million responses.

While MY World is a global survey, outreach has focused on developing countries. When disaggregating based on UNDP's Human Development Index[12] (HDI), low HDI countries are over-represented – they account for 38% of total responses but just 16% of the global population (Figure 2). This over-representation is particularly strong for mobile phone responses. Medium HDI countries are under-represented overall – accounting for 10% of responses but 32% of the world's people. High HDI countries are over-represented overall and particularly in online responses. Very high HDI countries are also over-represented among online responses, but under-represented overall.

[12] The Human Development Index (HDI) is a summary measure of average achievement in key dimensions of human development: a long and healthy life, being knowledgeable and have a decent standard of living.

Figure 2: Distribution of responses by source and HDI group



Low and medium HDI countries together constitute 96% of mobile phone responses. Although 175 countries are represented, a handful of countries account for the overwhelming majority. Sixteen countries – all low and middle HDI – account for about 93% of mobile phone responses.

Mobile phone survey partners initially reached out to their entire subscriber base in selected countries. The survey was often translated into regional languages to promote response. Subsequent efforts in some countries aimed to seek responses from certain sub-populations to address specific disparities (discussed in Section 4.2).

The mobile phone surveys in some countries were also accompanied by offline promotion. In Yemen, which accounts for 43% of mobile responses, an NGO partner promoted the survey through billboards, newspaper and radio advertisements, short films on television and social media, and fliers; the telecom partner (with over 1.5 million subscribers) sent messages requesting survey participation and followed up in cases of incomplete surveys. The campaigns in India and Pakistan included advertisements featuring Bollywood stars. However, in most countries, the survey was administered to 'captive' communities of mobile partners' subscriber bases without investment in promotion.

The online survey achieved responses from a wider range of countries than the mobile phone survey. Online responses were secured from 194 countries, with 22 countries[13] accounting for about three quarters of the responses. Low and medium HDI countries are underrepresented, and constitute only about a quarter of online responses.

4.1 Respondent characteristics by survey mode

While the MY World survey is a crowd-sourced polling exercise and is not designed to be representative of the global population, understanding the disparities in coverage sheds some light on whose voices are heard when using technology to gather data about people's

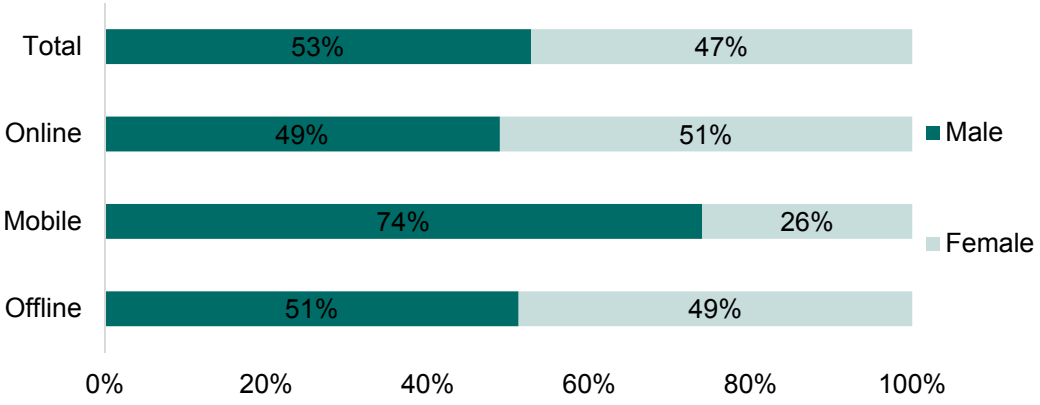
[13] India, Ukraine, United States of America, Mexico, Brazil, Thailand, United Kingdom, Sri Lanka, Spain, China, Australia, Canada, Argentina, Egypt, Belarus, Russian Federation, France, Colombia, Indonesia, Philippines, Germany and Bahrain constitute 75.4% of online responses.

opinions and priorities. This section explores the demographic characteristics of respondents to the MY World phone and online survey – namely, gender, age and educational attainment.

Where are the women? Gender differentials in MY World

At the aggregate level, about 47% of respondents to date are female. They comprise 49% of offline and 51% of online responses. However, a very stark gender gap emerges in the case of the mobile phone MY World survey: females constitute only 26% of mobile phone respondents (Figure 3).

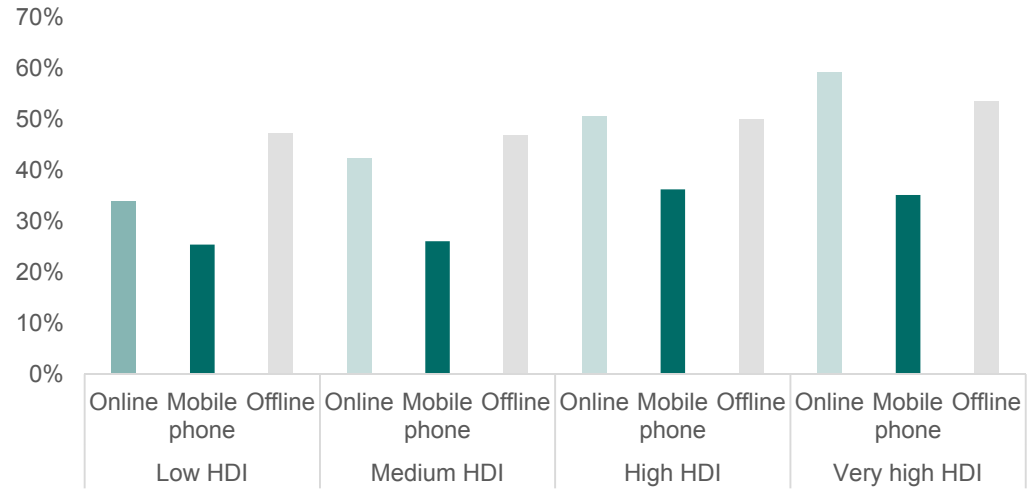
Figure 3: Gender distribution of MY World responses by mode



Source: MY World 2015 (data accessed 9 February 2015)

This gender disparity in mobile phone responses is most striking in low and medium HDI countries, but persists even in more developed countries (Figure 4). In the case of online responses, a male bias in low and medium HDI countries is reversed in very high HDI countries.

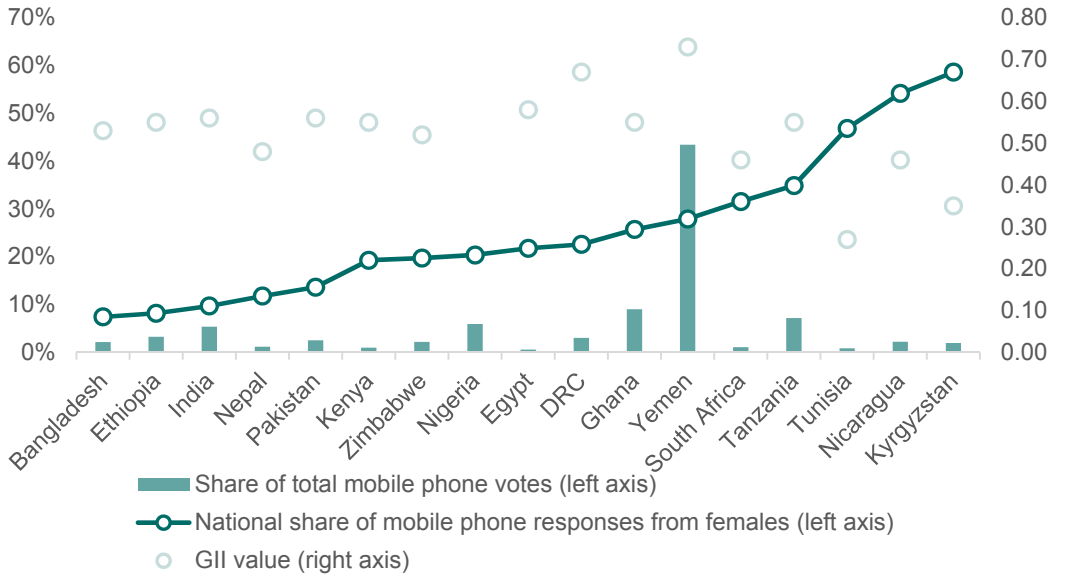
Figure 4: Share of female respondents by HDI and survey mode



Source: MY World 2015 (data accessed 9 February 2015)

A gender imbalance in mobile phone responses is evident in most of the 16 countries that constitute 93% of the mobile phone responses (Figure 5). This may be related to broader patterns of gender inequality in these countries. Of the 16 countries, 14 are in the bottom half and 10 are in the bottom quartile of countries according to UNDP's Gender Inequality Index (GII)[14], with high GII values relating to high levels of gender inequality (UNDP, 2014). Within this set of countries, however, GII performance and the share of female respondents are only weakly correlated (Spearman rank correlation $\rho = 0.38$, $P = 0.14$). This may relate to differential patterns of ownership and access – though we do not have the data that would enable a more thorough analysis of this hypothesis. We explore what is known about the gendered use of mobile phones in section 5.

Figure 5: Country context and share of mobile phone responses



Source: MY World 2015 (data accessed 9 February 2015); UNDP (2014)

4.2 Exploring age profiles

MY World mobile phone surveys reached out to all mobile phone users subscribed to partnering service providers, while the online survey was open to all internet users. As a result, biases in the age distribution of those with access to these technologies may be mirrored in the survey responses. In the case of the offline MY World surveys, a large share of the surveys were conducted by youth organisations and by grassroots NGOs in educational institutions and among young people, which may have contributed to a higher sampling of younger populations.

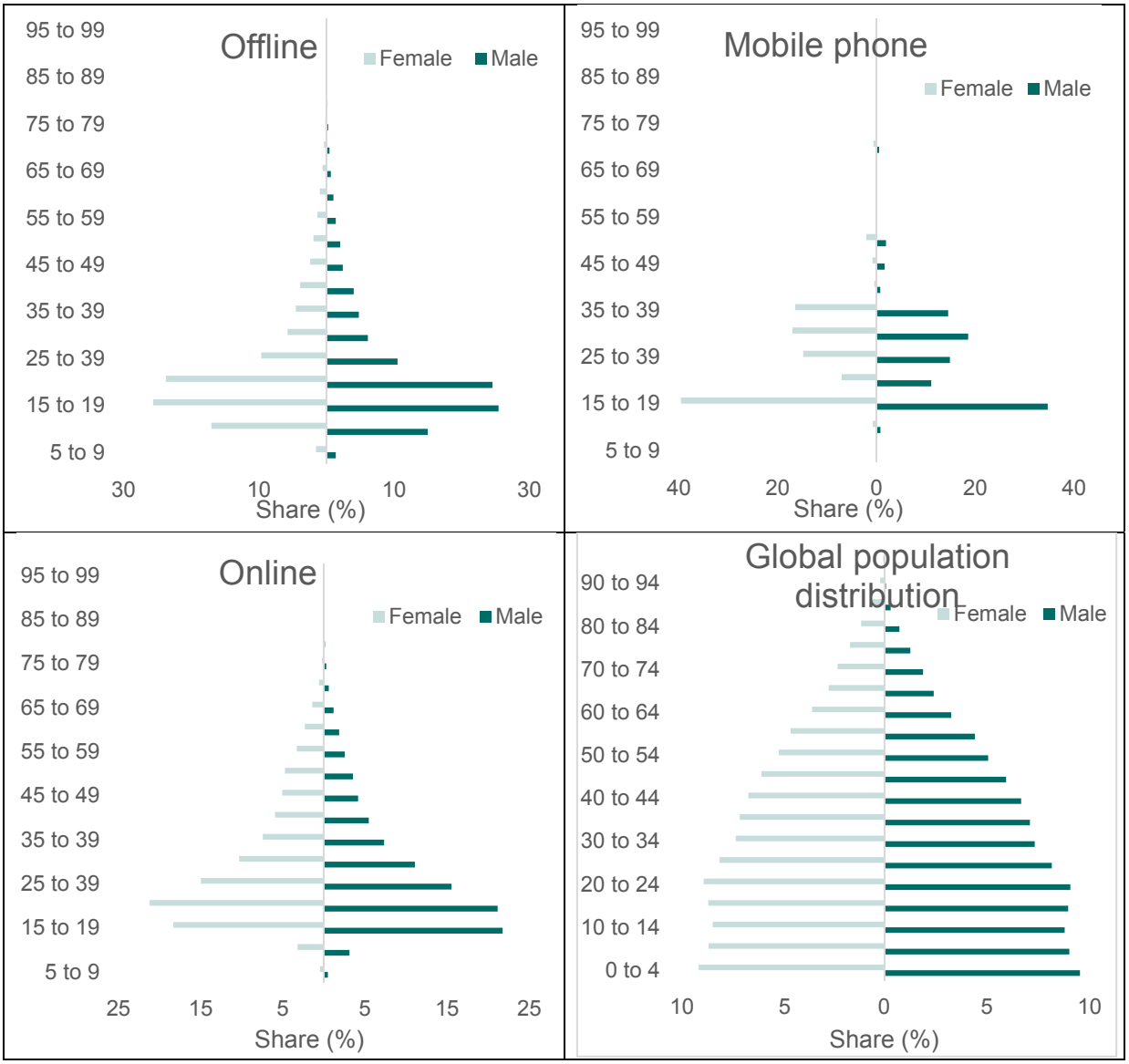
Overall, the median ages of those responding by mobile phone and online were 27 years and 26 years, respectively, while the median age of offline respondents was 21 years. The offline survey also had the lowest mean age (24 years), followed by mobile phone (26 years) and online (30 years).

There are clear differences in the age distribution of respondents based on survey mode and gender (Figure 6). Noticeably, all three modes of survey administration under-represent older people, and over-represent youth compared to the global population. This is

[14] The Gender Inequality Index is a composite measure reflecting inequality in achievement between women and men in three dimensions: reproductive health, empowerment and the labour market.

particularly the case in the mobile phone surveys which underrepresent both males and females over 40 years old.

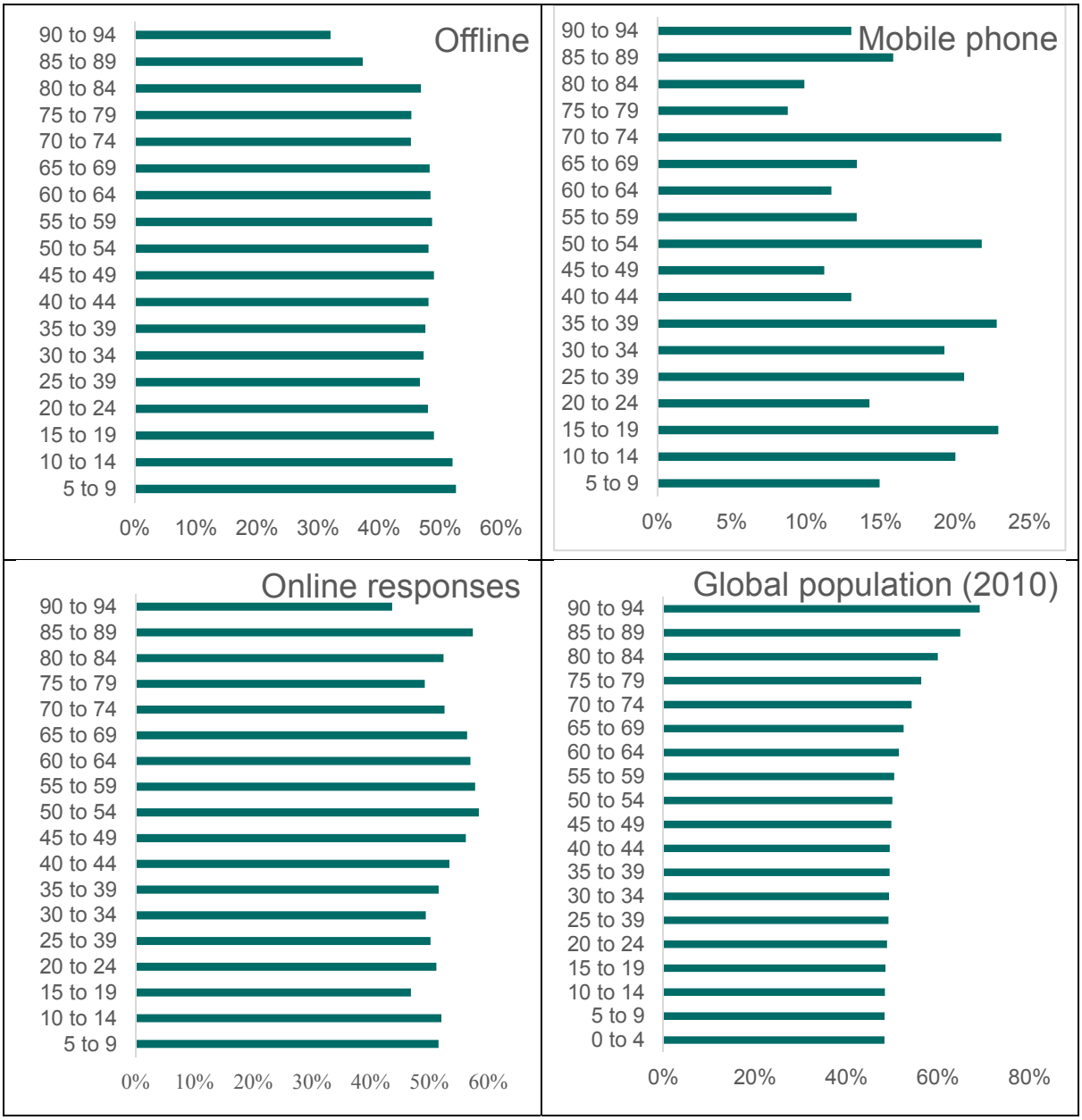
Figure 6: Age distribution of MY World respondents



Source: MY World 2015 (data accessed 9 February 2015); authors elaboration based on UN DESA (2014)

Looking more closely at the share of female respondents at different age levels, no systematic pattern emerges (Figure 7). The share of female mobile phone responses is below 30% for all age groups.

Figure 7: Share of female respondents by age and mode



Source: MY World 2015 (data accessed 9 February 2015); UN DESA (2014)

4.3 Education levels

At the global level, respondents with low education attainment (up to a complete primary education) constitute over a third of the offline sample. This may be since the offline surveys conducted by grassroots organisations sought to be inclusive and made a deliberate attempt to reach out to disadvantaged groups.

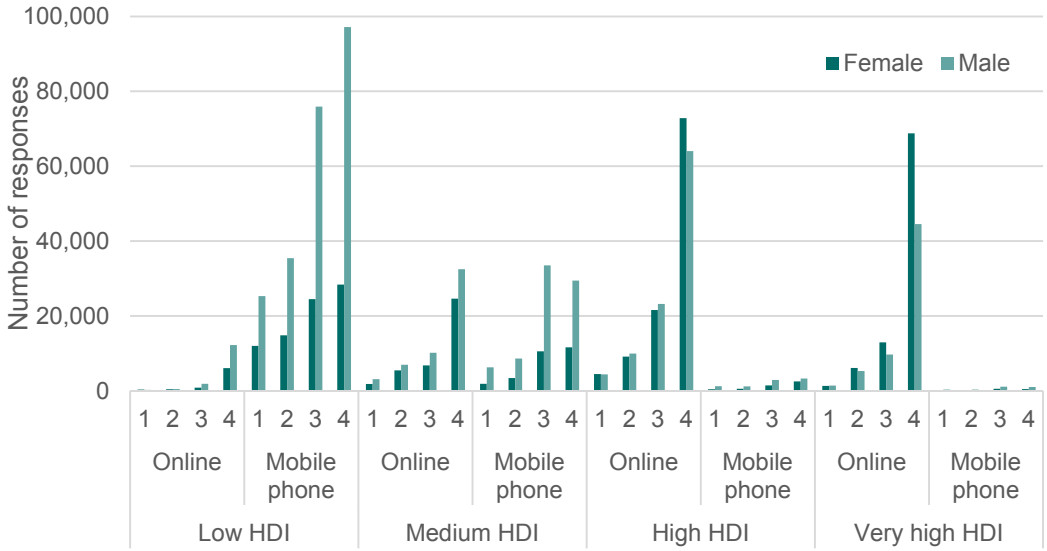
On the other hand, both the mobile phone and online surveys over-represent those with higher educational attainment.[15] While only 63.6% of the global population (aged 25 years

[15] It is difficult to quantify the extent of this bias in the absence of global data on age-specific educational attainment in the context of large increases in school completion rates over the past several decades. For example, primary school completion rate in low and middle income countries increased from 78.3% in 1990 to 91.1% in 2012 (World Bank, WDI).

and above) have at least some secondary education, this share is 74.3% for mobile phone responses and 87.4% for online responses (UNESCO Institute for Statistics, 2013). This trend is evident in countries at all HDI levels – although higher education levels in more developed countries[16] may imply higher imbalance in less developed countries.

This trend is visible among both male and female respondents (Figure 8). About 71% of female mobile phone respondents have at least some secondary education. In comparison, this share is 54% globally among women aged 25 years and above, while the net secondary enrolment ratio for girls was 61% in 2011 (UNESCO Institute for Statistics, 2013). Among males, about 76% of survey respondents have at least some secondary education. This compared with 64% of the global male population 25+ and an average global net secondary school enrolment rate for males of 64% (ibid.). The over-representation of those with more education is stronger still among online respondents – about 86% of males and 88% of females have at least some secondary education.

Figure 8: Education attainment of MY World respondents



Source: MY World 2015 (data accessed 9 February 2015)

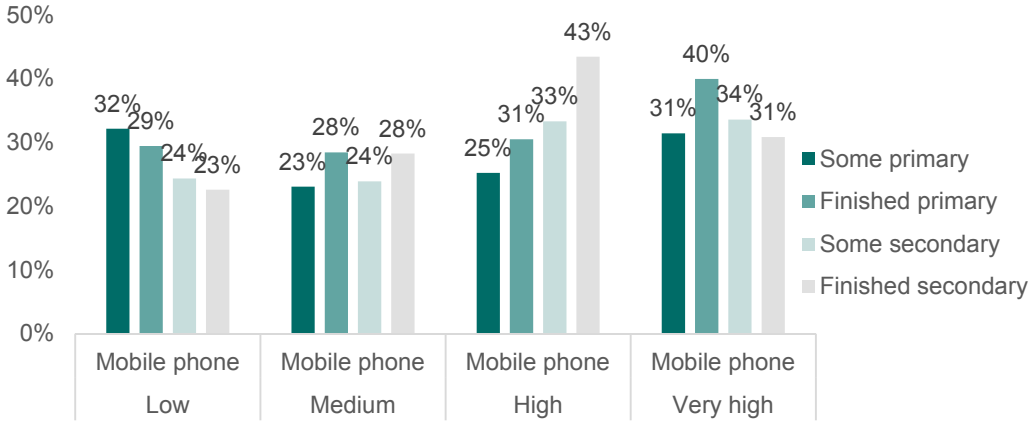
Note: For the education levels, 1 refers to some primary; 2 refers to completed primary; 3 refers to some secondary; and 4 refers to completed secondary education.

Female mobile phone respondents are systematically under-represented across education levels in all country groups,[17] particularly in low and medium HDI countries (Figure 9). In fact, in low HDI countries, the gender balance is better among less educated groups: 32% of responses are from females with only some primary education while only 23% of responses are from females with at least a complete secondary education.

[16] About 87% of people aged 25 years and above have at least some secondary education in very high HDI countries, compared to 65% in high HDI countries, 48% in medium HDI countries and 22% in low HDI countries (UNESCO Institute for Statistics, 2013).

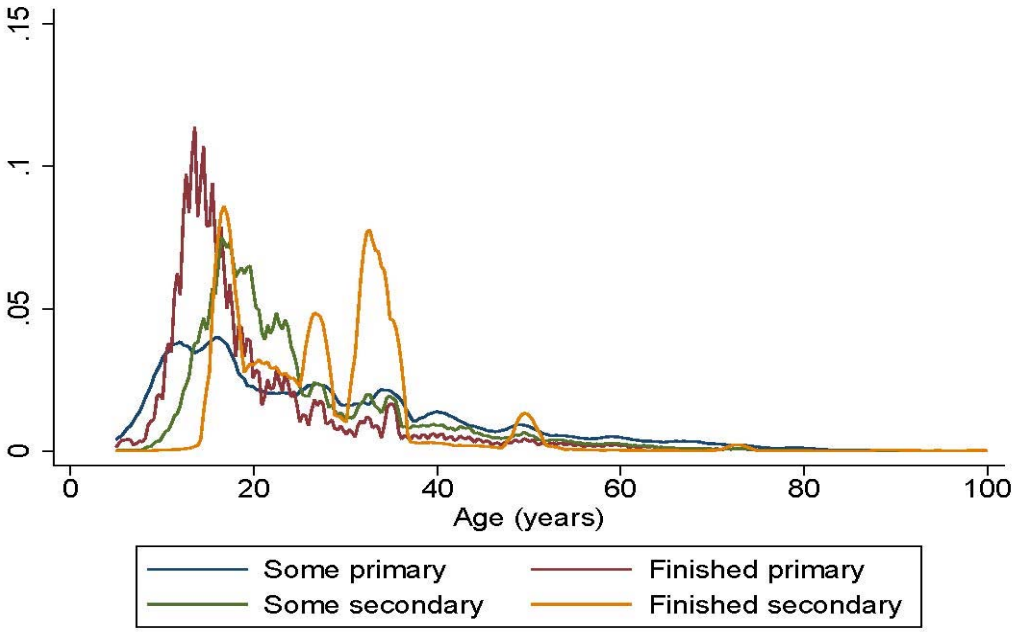
[17] However, high HDI countries account for only 13,795 responses (3.2% of total mobile phone responses) and very high HDI countries constitute only 4,739 responses (1.1% of total mobile phone responses).

Figure 9: Share of female respondents in mobile phone responses

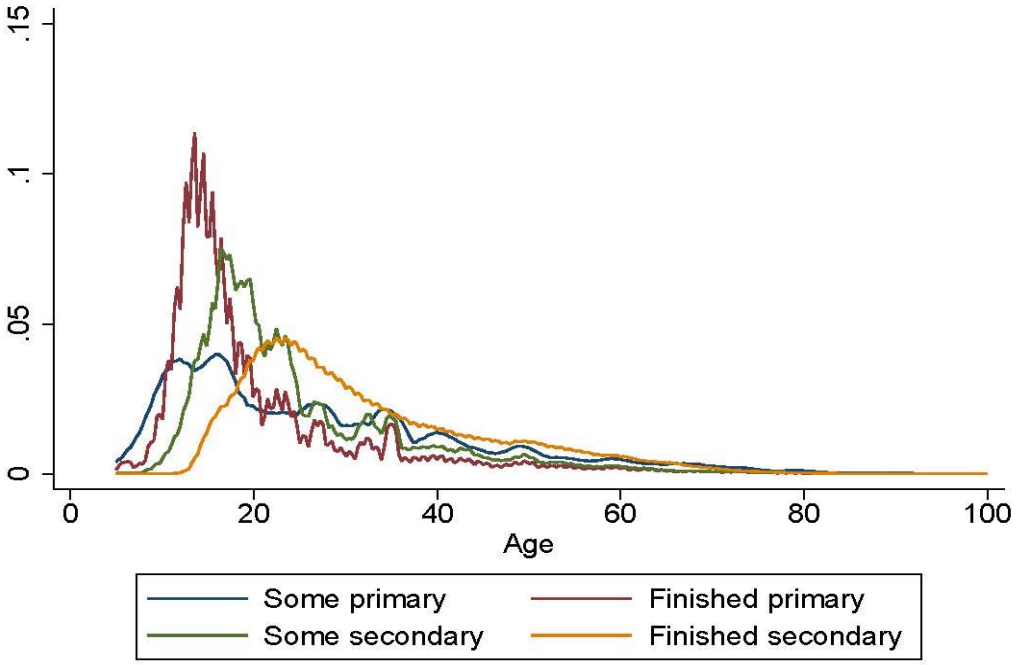


Exploring the age profiles relative to education reveals that the overall distribution of responses by age is similar across education levels for the mobile phone and online surveys. However, in both cases, while there are fewer respondents aged 30 years and above among groups with up to some secondary education, older respondents have at least a completed secondary education (Figure 10).

Figure 10: Age distribution of respondents by education level
(a) mobile phone



(b) online



The MY World survey, apart from highlighting the people's development priorities, sheds light on some challenges that new technology present for collecting data. Older populations and those with lower education participated less in the mobile phone and online surveys. Among online responses, the share of female responses is higher in more developed countries as measured by HDI. Strikingly, females constitute only about a quarter of responses to the mobile phone survey, and this imbalance holds across all age groups and education levels. Although the gender imbalance is starker in low and medium HDI countries, it holds even in more developed countries.

4.4 Response rates – what factors seem to encourage participation?

Given the different response rates for the mobile phone and online MY World surveys, we explore some factors that may increase participation. The mobile phone surveys have been conducted through telecom partners, many of which operated in multiple countries. However, data inadequacies preclude analysis of how response rates vary based on socio-demographic characteristics. However, this section brings together the limited data available in order to understand factors that affect response rates.

One of the partners, biNu, whose mobile applications enable low end mobile phone users to access internet and smartphone applications, has a subscriber base of over 2 million people across Asia, Africa and Latin America, and accounts for about a quarter of mobile phone responses. Initially, biNu reached out to all its users for participation. By October 2013, biNu had a response rate of about 35%, with about 69,190 responses received. However, males accounted for 83% of SMS responses (UNDP, n.d.). Due to the data limitations, it is difficult to disentangle whether the existing gender disparity is due to differential mobile phone ownership (discussed in Section 5) or differential response rates.

Given this gender imbalance, the company made deliberate efforts to seek female respondents more actively. An additional request was sent to registered[18] female users. In order to reach unregistered female users, the survey was promoted through mobile

[18] Users can be registered or unregistered, but need to be registered in order to receive credits. When registering users need to provide their gender, and this allowed targeting women users.

applications known to be widely used by women.[19] It also tested small incentives for survey completion including through benefits on applications, such as getting a free book. In addition, different languages were used to encourage wider participation.

In the case of online responses, people that visited the survey website through a social media referral were significantly more likely to respond. In addition, higher response rates have been observed when the survey was publicised widely in popular media and newspapers.

In contrast, people that arrived on the MY World website through an advertisement or email link were less likely to respond to the survey online. The experience of Anchor Free, a software that provides a virtual private network for secure browsing and access to blocked websites with 20 million active users across 190 countries, provides an interesting case study. A partner of the MY World survey, Anchor Free promoted it by presenting advertisements to all their subscribers during the 'Global Week of Action' in May 2014. This directed high traffic to the website, but recorded low 'conversions' or response rates. One possible cause is that people were arriving to the survey website without any context or perhaps by accident, and were therefore less amenable to respond.

Overall, the response rate during the Global Week of Action campaign peaked at 40%. However, the experience with the online MY World survey illustrates the importance of building awareness around a survey in order to increase response rates. Engaging with people online in a meaningful way is also likely to help attract respondents who will recommend the survey to others within their social circle, creating a cycle which would increase responses and keep traffic high.

5 Technology and the data revolution

The analysis conducted in this paper has pointed to some of the challenges encountered when using communication technology – both mobile phones and the internet – to collect data. In MY World, while the mobile phone survey was conducted largely in developing countries, the online survey had higher uptake in more developed countries.

Respondents to the mobile phone and online surveys over-represented people with more education. In addition, both the mobile phone and online surveys over-represented youth and under-represented older people. A stark imbalance concerns the gender of mobile phone respondents, only 26% of whom were female. This gender imbalance was highest in low and medium HDI countries but did not vary systematically by age or education level. This gender bias did not extend to the online surveys.

5.1 Cost savings and response rates associated with new technologies

The findings of the MY World survey mirror the experience of other such initiatives in recent years, and highlight some of the strengths and limitations of using technology for data collection in terms of who is reached.

Different initiatives have recorded widely varying rates. The *Listening to Dar* project in Dar es Salaam, Tanzania employed a baseline personal interview followed by CATI every

[19] For instance, the survey targeted the Mills and Boon reading app which has an overwhelmingly female audience.

two weeks; the survey had a large initial burst of attrition (attributed to low personal ownership of mobile phones), but recorded about 66% of responses over several subsequent rounds, which later increased to 75% once oversight was tightened when project management was transferred to the World Bank (Croke et al, 2012). This survey informed the national *Sauti za Wananchi* (Voices of Citizens) where a randomly selected sample was given mobile phones and solar chargers and contacted for regular CATI surveys. Response rates exceeded 90% in the first two rounds (Twaweza, 2013).

The World Bank's *Listening to LAC* initiative in Peru conducted similar benchmark surveys followed by regular mobile phone surveys through different modes, and also suffered high initial attrition. However, the attrition rate for CATI at 50% was significantly lower than for SMS (70%) and IVR (80%) surveys (Gallup, 2012). In comparison, dropout rates in traditional longitudinal surveys have ranged from less than 3% per year (e.g. Kagera Health and Development Survey in Tanzania 1991-2004) to over 30% per year (e.g. Kenyan Ideational Change Survey 1994-96) (Outes-Leon and Dercon, 2008).

5.2 Undercounting the marginalised?

MY World is distinct from many other surveys in aiming to maximize responses rather than ensure representativeness. It follows that the profile of respondents varies from surveys where respondents are rigorously sampled. The MY World experience therefore illustrates some systematic biases in technology use.

Overall, mobile phone surveys are considered reliable for research if technology penetration rates are at or above 80% (Croke et al, 2012)[20]. Although mobile subscription rates in some developing regions are still below this threshold, this method is viable for many developing countries, particularly in urban areas where mobile phone use is higher. For example, a survey of 23 sub-Saharan African countries found 80% of urban residents had mobile phones relative to only 63% those in rural areas in 2013 (Gallup, 2014) – and mobile phone penetration has continued to grow since then.

Further, household mobile phone ownership – which is usually higher than individual level ownership - is often 'good enough' for certain survey purposes. For example, in 2011, while only 63% of Tanzanian adults owned mobile phones, about 80% of households had at least one mobile phone either shared among the family or used by the household head and borrowed by others (Twaweza, 2013).

Within countries, while access to mobile phones has increased significantly over the past two of decades, considerable disparities remain. For instance, a 2010 survey in 17 sub-Saharan African countries found the average mobile phone owner is more likely to be male (62% of males compared to 52% of females owned mobile phones), educated (75% of those with at least nine years of formal education owned phones versus 44% of those with up to eight years), wealthy and urban (Gallup, 2011). Households with higher incomes have greater mobile phone ownership, with 80% of the richest quintile of households owning mobile phones compared with 55% of those in the bottom quintile in a survey across 23 countries In sub-Saharan Africa (Gallup, 2014).

Gender disparities are pronounced. A woman is 21% less likely to own a mobile phone than a man; this difference is even higher in Africa (23%), the Middle East (24%), and South Asia (37%) (GSM Association, 2010). Women are more likely to be borrowers of phones, from friends or household members, and access a phone used collectively by the household. A recent survey found 29% of women (compared with 6% of men) in India, 10% of women

[20] For instance, political polling in the US typically uses landline samples only – although only 80% of households own a landline - yet the point estimates of these surveys are considered reliable when corrected by re-weighting.

(compared with 1% of men) in Jordan, and 19% of women (and 8% of men) in Niger to be mobile phone borrowers (GSM Association, 2015).

However, the gender imbalance in MY World responses is more pronounced than these disparities; gender gaps in mobile phone ownership appear to be exacerbated by other factors that condition usage, such as varying usage patterns. For instance, in India 55% of women that owned mobile phones reported never sending an SMS compared with 33% of men; such gender disparities exist, though they are lower, in many other countries (GSM Association, 2015).

Social structures and power imbalances may also explain the observed gender gap. For instance, in the Dar es Salaam mobile phone survey, a few instances of domestic violence were linked to female respondents participating without their husband's approval and some women dropped out under instruction from their partners; the survey subsequently required participation consent from the household head (Twaweza, n.d.). In addition, findings from a recent survey by the World Bank and Center for Global Development in Afghanistan, Ethiopia, Mozambique and Zimbabwe identified that using a male voice for surveys may contribute to low response among women (Leo and Morello, 2015).

In the case of online surveys, the composition of MY World respondents reflects reported differences in internet access. For instance, people between 18 and 34 years of age were 15 percentage points more likely to report accessing the internet that their older counterparts in 29 out of 32 developing countries covered in recent research (Pew Research Centre, 2015). The survey also found that people with secondary education or higher and those with English language ability were more likely to use the internet than their less educated counterparts.

5.3 Glass half empty or half full?

The MY World survey attempts to capture a snapshot of the development priorities of people from across the world to inform the Sustainable Development Goals and is not intended to be statistically representative. However, a range of methods can help correct for selection bias. For instance, Zagheni and Weber (2012) inferred migration rates using geo-coded email data by making adjustments to correct for selection bias[21] caused by varying internet penetration (higher among more educated, richer, and urban populations). Similarly, non-representative polls are commonly used in election forecasting. For example, Wang et al (2014) analysed data from an opt-in poll from a gaming platform prior to the United States presidential election in 2012 and generated forecasts similar to those of leading poll analysts using statistical adjustments. In the case of the MY World survey, the additional social and demographic information collected can be used to reweight or disaggregate population subgroups to paint a picture of how people's priorities may vary.

Given weak statistical systems and the time and resources needed for household surveys, the use of technology holds promise for data collection in many contexts. In the foreseeable future, representative household surveys will remain fundamental to tracking development progress. Technology will not replace traditional surveys, but rather supplement them, offering cost effective and continuous data collection. As with anything, a greater dependence on technology comes with trade-offs, and it is important to understand these.

In the context of ongoing discussion about the 'data revolution', various innovative approaches have been utilised to collect data in a quick and cost-effective way. These approaches include mobile phone and internet surveys. MY World, despite the imbalances studied in this paper, exploited the opportunity presented by communication technology to employ mixed-mode data collection to understand what is important to people. Careful

[21] The paper estimated parameters at the value that maximised the likelihood of the actual figures for countries where data were available, which could then be applied to other countries (Zagheni and Weber, 2012).

study of its biases add to our understanding of who these types of surveys are reaching and who they are not. With the continual expansion of new technologies, understanding and addressing these imbalances will be crucial.

References

Ballivian, A and J P Azevedo (2013). 'Listening to LAC: Using Mobile Phones for High Frequency Data Collection'. Poverty, Equity and Gender Team of World Bank Latin America and Caribbean Region. Washington DC: World Bank.

Bergh, Gina, Alina Rocha Menocal and Laura Rodriguez Takeuchi (2014). 'What's behind the demand for governance? An assessment of people's views'. London: Overseas Development Institute.

Bhatkal, Tanvi (2014). 'What do women want? Gender, perceptions data and development priorities'. London: Overseas Development Institute.

Boyd, Danah and Kate Crawford (2011). 'Six provocations for big data'. A Decade in Internet Time: Symposium on the Dynamics of the Internet and Society, September 2011.

Caeyers, Bet, Neil Chalmers and Joachim De Weerdt (2011). "Improving Consumption Measurement and other Survey Data through CAPI: Evidence from a Randomized Experiment". Journal of Development Economics 98(1) pp 19-33.

Carr-Hill, Roy (2013). 'Missing Millions and Measuring Development Progress' World Development 46 pp 30-44.

Croke, K, A Dabalen, G Demombybes, M Giugale, and J Hoogeveen (2012) 'Collecting High Frequency Panel Data in Africa Using Mobile Phone Interviews'. Policy Research Working Paper 6097. Washington DC: World Bank.

Croke, K, A Dabalen, G Demombynes, M Giugale, and J Hoogeveen (2013) 'Collecting High-Frequency Data Using Mobile Phones: Do Timely Data Lead to Accountability?' Washington DC: World Bank.

Escobala, Javier and Sara Benites (2013). 'PDAs in socio-economic surveys: instrument bias, surveyor bias or both?' International Journal of Social Research Methodology 16(1) pp 47-63.

Gallup (2011). 'Mobile Phone Access Varies Widely in sub-Saharan Africa'. Available at http://www.gallup.com/poll/149519/mobile-phone-access-varies-widely-sub-saharan-africa.aspx#1

Gallup (2012). 'The World Bank Listening to LAC (L2L) Project: Final Report' October 2012.

Gallup (2014). 'Africa Continues Going Mobile'. Available at http://www.gallup.com/poll/168797/africa-continues-going-mobile.aspx

Global Pulse (2013). United Nations Global Pulse Information Sheet – September 2013. Available at http://www.un.org/millenniumgoals/pdf/GP%20Backgrounder-General2013_Sept2013.pdf

Goldstein, Markus, Raka Banerjee and Talip Kilic (2012). 'Paper versus plastic part 1: The revolution is in progress'. World Bank blog, 24 July 2012.

GSM Association (2010). 'Women and mobile: a global opportunity'. GSMA Development Fund, Cherie Blair Foundation for Women and Vital Wave Consulting.

GSM Association (2015). 'Bridging the gender gap: mobile access and usage in low- and middle-income countries'. GSMA Connected Women Global Development Alliance.

Independent Expert Advisory Group (2014). 'A world that counts: mobilising the data revolution for sustainable development'. November 2014.

Leo, Ben and Robert Morello (2015). 'Do mobile phone surveys work in poor countries?' Blog available at http://www.cgdev.org/blog/do-mobile-phone-surveys-work-poor-countries. Center for Global Development.

Lucci, Paula and Tanvi Bhatkal (2014). 'Monitoring progress on urban poverty: are current data and indicators fit for purpose?' Working paper 405. London: Overseas Development Institute.

Measure DHS/ ICF International (2012). 'Survey Organisation Manual for Demographic and Health Surveys'. MEASURE DHS. Calverton. Maryland: ICF International.

Melamed, Claire and Paul Ladd (2013). 'Listening to 1 million voices' Available at http://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/8580.pdf

Narayan, A and V Shmatikov (2008) 'Robust De-anonymization of Large Sparse Datasets'. SP 2008. IEEE Symposium on Security and Privacy. Oakland: IEEE, 2008.

Outes-Leon, Ingo and Stefan Dercon (2008). 'Survey attrition and attrition bias in Young Lives' Young Lives Technical Note No. 5. Oxford: Young Lives

Patnaik, S, E Brunskill, and W Thies (2009). 'Evaluating the accuracy of data collection on mobile phones: A study of forms, SMS, and voice'. Information and Communication Technologies and Development (ICTD), 2009 International Conference on IEEE, 2009.

Pew Research Centre (2015). 'Internet Seen as Positive Influence on Education but Negative Influence on Morality in Emerging and Developing Nations'. Available at http://www.pewglobal.org/files/2015/03/Pew-Research-Center-Technology-Report-FINAL-March-19-20151.pdf

Roberts, Caroline (2007). 'Mixed modes of data collection in surveys: a methodological review' National Centre for Research Methods discussion paper. Available at http://eprints.ncrm.ac.uk/418/

Rodríguez Takeuchi, Laura, Sébastien Hine with Cirenia Chávez (2015). 'Asking people what they think: using perceptions data to monitor the post-2015 agenda'. Working paper 413. London: Overseas Development Institute.

Rodriguez Takeuchi, Laura (2014). 'More and better jobs: does everyone want the same?' London: Overseas Development Institute.

Schuster, C and C Perez Brito (2011). 'Cutting costs, boosting quality and collecting data real-time – Lessons from a Cell Phone-Based Beneficiary Survey to Strengthen Guatemala's Conditional Cash Transfer Program'. World Bank LAC en breve No. 166. February 2011.

Stuart, Elizabeth, Emma Samman, William Avis, and Tom Berliner (2015). 'The data revolution: finding the mission millions' London: Overseas Development Institute.

Trucano, M. (2014) 'Using mobile phones in data collection: Opportunities, issues and challenges'. World Bank Blogs (18 April, 2014). Washington DC: World Bank. Available at http://blogs.worldbank.org/edutech/

Twaweza (2013). ‘Sauti za Wananchi: collecting national data using mobile phones’. Brief no. 1. September 2013. Dar es Salaam: Twaweza.

Twaweza, n.d. ‘Sauti za Wananchi: collecting national data using mobile phones’. Dar es Salaam: Twaweza.

UN DESA (2010). ‘Report on the results of a survey on census methods used by countries in the 2010 census round’. Working paper UNSD/DSSB/1.

UNDP (n.d.). ‘MY World SMS Outreach Report’

United Nations (2013). ‘A new global partnership: eradicate poverty and transform economies through sustainable development’. The Report of the High Level Panel of Eminent Persons on the Post-2015 Development Agenda. New York: UN.

United Nations (2014). ‘The road to dignity by 2030: ending poverty, transforming all lives and protecting the planet’. Synthesis Report of the Secretary General on the Post-2015 Agenda. 4 December 2014. New York: UN.

Wang, Wei, David Rothschild, Sharad Goel, and Andrew Gelman (2014). ‘Forecasting elections with non-representative polls’. International Journal of Forecasting.

Wilmoth, John; Nobuko Mizoguchi; Mikkel Oestergaard; Lale Say; Colin Mathers; Sarah Zureick-Brown; Mie Inoue; and Doris Chou (2012). ‘A New Method for Deriving Global Estimates of Maternal Mortality’. Statistics, Politics, and Policy 3(2). DOI: 10.1515/2151-7509.1038

World Bank, World Development Indicators.

Zagheni, Emily and Ingmar Weber (2012). ‘You are where you email: using email data to estimate international migration rates’. In Association for Computing Machinery eds. Proceedings of ACM WebSci 2012 (22-24 June 2012) Evanston. New York: ACM.

UNDP (2014). ‘Sustaining Human Progress: reducing vulnerabilities and building resilience’ Human Development Report 2014. New York: UNDP.

UN DESA (2013). World Population Prospects: The 2012 Revision. New York.

UN DESA (2014). World Population Prospects, the 2013 Revision. New York.

ODI is the UK's leading independent think tank on international development and humanitarian issues.

Our mission is to inspire and inform policy and practice which lead to the reduction of poverty, the alleviation of suffering and the achievement of sustainable livelihoods.

We do this by locking together high-quality applied research, practical policy advice and policy-focused dissemination and debate.

We work with partners in the public and private sectors, in both developing and developed countries.

ISSN: 2052-7209

Overseas Development Institute
203 Blackfriars Road
London SE1 8NJ
Tel +44 (0)20 7922 0300
Fax +44 (0)20 7922 0399