Public Disclosure Authorized

Public Disclosure Authorized

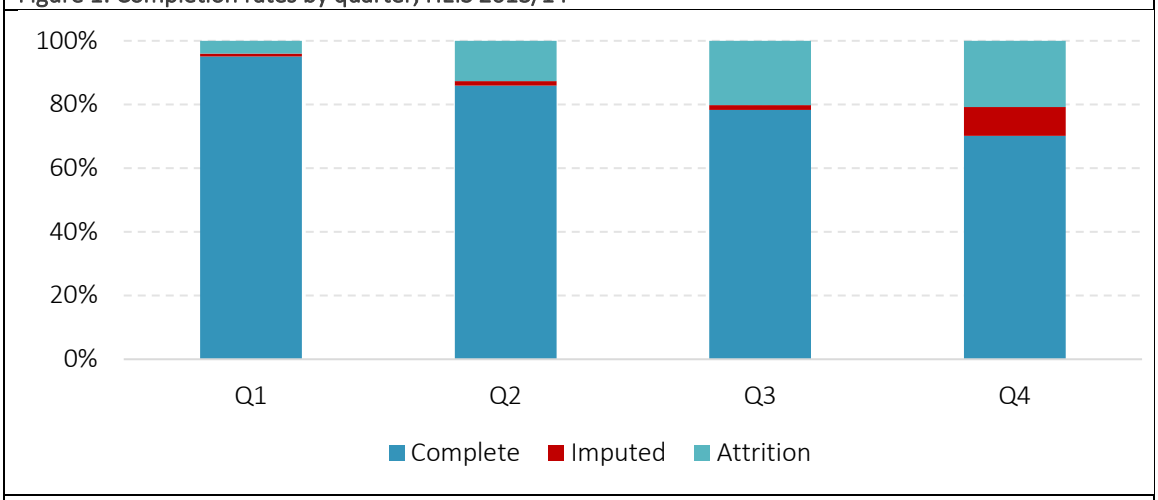
# JORDAN HOUSEHOLD EXPENDITURE AND INCOME SURVEY 2017-18
COMPLETION NOTE

## 1. INTRODUCTION

**The Household Expenditure and Income Survey (HEIS) is the central database that measures and tracks poverty and living standards in Jordan.** The HEIS is a household budget survey, which also collects data on household and individual demographics (age, gender, marital status, among others), education, employment status, assets, and income. The survey is used as the basis for estimating the prevalence of poverty in Jordan; it is a key data source for National Accounts and the CPI; and it is used to benchmark broader indicators of household welfare. Department of Statistics (DOS) in Jordan has implemented the HEIS since 1966, and the survey had been implemented on a bi-annual (every two years) basis for a decade up until 2013.

**The 2013/14 HEIS faced significant challenges, which negatively impacted data quality, and could not be used to produce credible poverty estimates**. 35% of sample households refused to participate in this survey at all; and over one-third of the households that did participate did not complete the survey (Figure 1). These non-response rates, which stemmed from a combination of methodological and implementation issues were significant enough to affect the credibility of the welfare and poverty estimates drawn from the data. Recognizing the serious data quality issues, the Minister of Planning and International Cooperation (MOPIC) in Jordan accepted the World Bank's technical recommendation to not release poverty estimates based on the 2013/14 HEIS.

Figure 1: Completion rates by quarter, HEIS 2013/14




*Source: World Bank staff estimates based on HEIS 2013/14*

**The 2016-18 HEIS TA program was initiated in response to a request from MOPIC to the World Bank to help the Department of Statistics (DOS) overhaul all aspects of the HEIS survey; and it was designed in the context of protracted refugee crisis following a large influx of Syrian refugees into Jordan over a 6-year period.** The need for this TA program was therefore urgent, not only to fill the gap in core

welfare data since 2010 but also because this 6-year data gap coincided with a large influx of Syrian refugees into Jordan. This protracted crisis defined an immediate need for data to enable the international donor community to design and implement programs that target Syrian refugees while at the same time benefitting host (Jordanian) populations. This need to inform policies for refugee and hosts defined an additional mandate: that of expanding the representation of the HEIS survey to include non-Jordanians

The objectives of this TA program were therefore two-fold: (i) To support a high-quality data collection effort for the 2017/18 HEIS to help fill knowledge gaps and inform policy and program design for all residents of Jordan, including Syrian Refugees; and (ii) To build sustainable capacity within DOS for continued updating of high quality micro data that help track and monitor welfare. To achieve these multiple objectives, the TA program was designed to: (i) support a redesign of the survey and sampling methodology, to address the weaknesses in the previous survey system; (ii) support a complete overhaul of survey implementation and data management systems; and (iii) expand survey representation to cover three types of populations (Jordanians, Syrians, and other non Jordanians).

The 2017/18 HEIS was launched in August 2017, after a complete re-design of all aspects of the survey, and with expanded representation. The HEIS in Jordan is now the first survey outside the developed world that collects core welfare data on all its residents. This improved design is central to being able to measure and track multiple dimensions of welfare, and to inform both government and donor-driven policy initiatives. In particular, the key strengths of this data will be its ability to: (i) support programming for refugees, and on resilience to refugee shocks for the host community; and (ii) inform policies on employment and economic inclusion through the labor and social protection modules.

A critical element of this TA program was the use of an embedded approach that involved an iterative model of learning by doing to build capacity; and a long-term investment in building client ownership for the significant changes proposed for this new survey. Key features that facilitated the implementation of these changes included careful field piloting of key changes in survey design; and the use of a system of performance-based awards to field enumerators to incentivize both field quality and the systematic use of data management systems.

This TA program achieved its multiple objectives. It set new standards on response rates in a middle-income country context and effected a significant improvement in data management systems to afford concurrent availability of data for analysis. Through the embedded approach to delivering the TA, this program delivered high quality data across the board whilst managing critical changes in sampling design and survey methodologies Moreover, it set the standard for what is possible to achieve in a middle-income country in terms of response rates (of 96% over the survey year); and defined a world-leading standard of timeliness in the availability of data for analysis. The time lag from the end of data collection to its availability for analysis was 4 weeks. Initial technical options for poverty estimates were presented within 2 weeks after the first survey quarter was completed; and in 6 weeks after the conclusion of the full survey. These outcomes are described in detail in section 3.

This completion note documents the **key elements of technical assistance** provided during the WB engagement with DOS, which included far reaching changes in sampling, questionnaire design, field implementation and data management; and it describes **key activities and innovations** that were implemented within the purview of the **embedded TA program** for the 2017/18 HEIS.

## 2. KEY ACTIVITIES: FROM DIAGNOSIS TO DESIGN

An in-depth diagnosis of the HEIS survey suggested that the entire survey system required a complete overhaul. As part of this:

i. The **sampling design** was revisited to be representative at the governorate level, and to deliver national estimates on a quarterly basis;
ii. The **questionnaire was redesigned** to reduce interview burden, in line with international best practice; and to be able to inform the evidence base for a wide range of policy relevant questions. This included a transition from diary to recall methods to measure consumption expenditures, and a move towards international best practice methods to measure labor and social protection indicators;
iii. An **integrated data and field management system**, which allowed for data quality to be monitored and improved on a concurrent basis, was designed and implemented.

For each of these TA elements, this section documents the challenges recognized by the DOS-WB team that needed to be addressed going forward; and links these key failures to technical solutions, along with the pros and cons of alternative options, where relevant.

### A. SAMPLE DESIGN

**A complete diagnostic of the sampling design for the 2013-14 HEIS pointed to multiple areas for improvement**. First, the sample design was complex, as it had been designed as a panel over the survey quarters (with data collected at 4 times during the year). To implement this panel, the data collection protocol involved up to 17 visits to each sample household over a survey year. The large non-response rate, which was over 65% when survey attrition was accounted for, could partially be attributed to this significant time burden on respondent households. Second, the sample frame for the 2013/14 survey was also outdated as it was based on the ten-year-old 2004 Population and Housing Census. It did not therefore capture the major changes the country experienced over the previous decade, including a large influx of over Syrian Refugees. The 2013/14 HEIS, which was implemented between July 2013 and June 2014, also coincided with the peak of a massive influx of Syrian refugees.

**Sampling designs also have important implications for field implementation, and for the timely availability of poverty estimates.** An overarching objective of the 2013/14 HEIS round was to produce poverty and other welfare indicators at the governorate level, in order to target policy efforts to address spatial disparities in living conditions. To do so, DOS decided to expand the sample size from 13,866 households in 2010 to 23,841 households in 2013/14. However, this intended sample size, could not be achieved in the field with the existing field and data management systems. Expanding sample size without commensurate changes in field

implementation further magnified the problem of data quality faced in the 2013/14 round of the HEIS. Last but not least, the sampling was designed to deliver nationally representative poverty estimates only at the end of the 12-month survey year: which meant that were a significant lag between the implementation of the survey and availability of estimates.

Moreover, the sampling design for the new HEIS was confronted with a new challenge: that of capturing major changes in the Kingdom induced by a large refugee influx in the recent past. The sampling design of the 2017/18 HEIS therefore had multiple goals: (i) Improving sample design from the point of view of reducing survey burden, and delivering more timely national estimates of poverty; (ii) Delivering credible poverty estimates at a sub-national level - given the difficulties faced with this in the past; and (ii) Expanding the representation of the survey to include all residents of Jordan, given the significance of the number of non-Jordanian nationals resident in the Kingdom.

The sampling design for the 2017/18 HEIS used the latest census (completed in 2015), to create an updated sampling frame. This new census presented two important opportunities for change: (i) it presented an opportunity to consider a systematic revision in the sampling design. In particular, concerns about a break in the data series are less relevant when the master frame is updated; and (ii) since this census covered all residents of Jordan, it presented the opportunity to expand the representation of the HEIS to include all residents of Jordan.

The sample for the 2017/18 HEIS was therefore to be representative spatially at the governorate level over a full year; and of all non-Jordanian residents of Jordan over the full survey year. In addition, the sample was also apportioned so as to be representative at a national level, for every survey quarter. Unlike previous rounds of the HEIS, which had a panel structure, the new sample design decreased the respondent burden by reducing the number of visits per sample household. In particular, households would need to be visited only at one time during the survey year (with an average of 4 visits per household). This sampling design yielded a total intended sample of 20,000 households, randomly selected in two stages. In the first stage 2,500 Primary Sampling Units (PSUs) were selected, using as a sample frame the list of blocks identified by the 2015 census. In the second stage, 8 households – 6 Jordanians and 2 Non-Jordanians – were selected in each PSU, using as a sample frame a full updated listing of all households in the PSU that was conducted as part of the survey effort.

Given the time lag between the census and HEIS, a full listing of selected EAs was required to update selection probabilities and stratify the sample across Jordanian/non-Jordanians. A key innovation was to increase the frequency of listing, in order to capture migrant populations who may be more mobile. In a typical survey, listing is done once before the survey, and this is usually enough to capture resident households. To allow for the prospect of mobility especially among refugee populations, the listing exercise for the 2017/18 was conducted on quarterly basis.

The 2,500 PSUs were allocated into the 12 governorates in order to bring about accurate estimates both for each governorate and for the Kingdom as a whole, as shown in the first columns of Table 1 below. The last columns give maximum margins of errors that this sample is expected to deliver in each domain, for the estimation of any prevalence (such as the poverty rate). A more detailed account of the sampling methodology is described in Annex 1.

| Table 1: Allocation of the 2017 HEIS sample by governorate and type of population, and maximum margins of error expected for the estimation of prevalences | | | | | | | |
|---|---|---|---|---|---|---|---|
| | | Sample size (households) | | | Margins of error | | |
| Governorate | | Jordanians | Non-Jordanians | Total | Jordanians | Non-Jordanians | Total |
| 11 | Amman | 4239 | 1379 | 5618 | 1.8% | 2.8% | 1.5% |
| 12 | Balqa | 937 | 279 | 1216 | 3.9% | 6.2% | 3.4% |
| 13 | Zarqa | 1539 | 478 | 2017 | 3.1% | 4.7% | 2.6% |
| 14 | Madaba | 845 | 219 | 1064 | 4.1% | 6.9% | 3.7% |
| 21 | Irbid | 1940 | 573 | 2513 | 2.7% | 4.3% | 2.3% |
| 22 | Mafraq | 899 | 261 | 1160 | 4.0% | 6.4% | 3.4% |
| 23 | Jerash | 842 | 238 | 1080 | 4.1% | 6.7% | 3.5% |
| 24 | Ajloun | 864 | 200 | 1064 | 4.1% | 7.3% | 3.8% |
| 31 | Karak | 890 | 222 | 1112 | 4.0% | 6.9% | 3.6% |
| 32 | Tafilah | 887 | 153 | 1040 | 4.0% | 8.3% | 3.9% |
| 33 | Maan | 880 | 176 | 1056 | 4.0% | 7.7% | 3.7% |
| 34 | Aqaba | 815 | 245 | 1060 | 4.2% | 6.6% | 3.6% |
| Total | | 15577 | 4423 | 20000 | 1.1% | 1.8% | 0.9% |

## B. QUESTIONNAIRE DESIGN

**The HEIS was traditionally designed to be primarily a household budget survey** with its main objectives centered on the calculation of poverty indicators, generating weights for constructing the Consumer Price Index, and measuring consumption and income of households for National Accounts. [1] The primary modules implemented by this survey focused on consumption expenditures and incomes; and brief modules were used to collect data on household demographics, assets, and labor. Consumption expenditures in this survey were collected using a 7-day diary for food consumption expenditures; and a 30-day recall for 364 non- food items.

**The diary method of collecting consumption expenditures, which as used by Jordan until 2013, faced significant implementation challenges and enumerator effort.** Survey practices around the world collect food consumption (and expenditures) either through the use of "diary" method or "recall". The literature says that there is no clear evidence to what extent these methods are substitutes or if there are biases introduced by switching from one to the other. [2] While there still continues to be a debate around which one is a preferred method, the food consumption expenditure module in the 2013/14 HEIS was based on a 7 day diary, and it faced serious implementation challenges. Many households were unwilling to fill in diaries, and this trend was increasing over time. In response to this problem, the 2013/14 HEIS used an intensive field operation where field

[1] Department of Statistics (2012), Household Expenditure and Income Survey 2010, The Hashemite Kingdom of Jordan.

[2] For details on this debate, please refer to Beegle, K. et al. (2012)

interviewers assisted households in completing the diary. In effect, this involved once in two-day visits over a 7-day period to the household. On each visit, field interviewers used (2 day) recall data to fill in the diaries. This practice effectively blurred the line between food consumption data collected by diary and by recall interview.

**A transition from diary to recall methods was adopted in response to the increasing difficulty of collecting accurate diaries and to address concerns about respondent burden. Key elements of this transition included: (i) analysis of past diary data to inform the list of food items to be included in recall modules; and (ii) field piloting to inform the design of the recall module.** First, food expenditure data from the 2010 HEIS diary was analyzed to identify items, which were rarely reported and had insignificant budget shares. This analysis was used to reduce items in the 2017/18 HEIS recall list. Similarly, through identification of similar items with respect to expenditure patterns and calories, some consolidation of items was undertaken. This process required a careful balance between consolidation into manageable recall list and comprehensive coverage; as well as between grouping similar items (in terms of expenditures and calories) while allowing for heterogeneous and differentiated food consumption to be captured. As a result of this analysis, 233 items that were reported in food diaries in 2010 were consolidated into 191 items used in the recall module 2017. The design of the non-food module was changed from one-month recall to multi-period recall. These new recall modules were intensively tested in the field before being rolled out as part of the survey. This field test or pilot involved two stages: a pre-test where the survey was administered to a small set of respondents to decide on the design of the module between two widely used alternatives; and a full scale pilot of the chosen module (from interview to data entry and checks) which was supported by the TA program. The findings of these tests were incorporated into the final recall module implemented in the HEIS 2017/18. Further details that relate to this analysis are presented in Annex 2.

**The critical need to get improve data quality and get credible poverty estimates necessitated a fundamental change in the data collection methodology for consumption expenditure data, and the start of a new poverty series in Jordan.** Consequently, a decision was made to sacrifice comparability for improvement in survey methodologies and data quality, and to use this an opportunity to start a new poverty series The team also discussed the possibility of simultaneously using dairies in a subsample to preserve comparability, but in the end this was seen as too complex to feasibly implement within the current organizational and governance systems at DOS.

**Given the risks to implementation capacity that could be induced by moving to a full multi topic LSMS style survey, this TA program selectively focused on modifying key modules like the design of the labor and social protection modules - which are of central policy relevance in Jordan.** In addition to the changes in the consumption module described above, the program introduced significant changes to the labor and social protection modules. The labor module, which was answered by direct respondents for the first time, was implemented with an improved design that explicitly accounted for key response patterns that may be more applicable to non-Jordanians. The social protection module was expanded significantly, to include programs targeted to non-Jordanians

(UNHCR, WFP etc); as well as to better measure access to formal and informal social protection mechanisms and transfers for Jordanians. The design of this module was finalized with advice and inputs from the Ministry of Social Development.

**Finally, the 2017 HEIS also included questions that allow for estimation of the flow of services from durable goods for the first time.** In the past, the HEIS collected information on purchases of durable assets during last year. While this approach is used in some countries due to its simplicity this does not follow best practice methods. In particular, it can be misleading because durables are consumed over a longer time range, and the purchase price corresponds to the value of good for its entire life (Amendola and Vecchi 2014). Instead of focusing on purchases, survey data should be able to measure the use value of a durable for a given reference period – which is typically a year. In order to properly estimate a consumption flow from durables, the 2017/18 HEIS collected detailed information on the stock of durables, including the year of purchase, and purchase and current values.

## C. CHOICE OF DATA MANAGEMENT SYSTEM: CAFÉ VS CAPI

**The benefits of integrating computer-based quality controls to fieldwork are well established** as a key to the success of complex surveys such as the HEIS. With the rapid progress in technology and its increasing affordability, computer assisted paperless interviewing (CAPI) is an increasingly popular choice for survey interviews. In the case of Jordan, the pros and cons of CAPI versus paper based but Computer assisted Field Entry (CAFÉ) techniques were considered. In particular, this choice was considered with the key end term objective of data quality.

**Data quality depends on the interface of field management systems with technology - which in turn rely heavily on human capital and management capacity.** Most computer-assisted survey applications in developing countries have typically been in simpler surveys (e.g.: market research, opinion polls), or in surveys that are smaller in scale or scope (e.g.: labor force surveys, census). However, household surveys like the HEIS are complex multi-topic surveys with very detailed food and non-food consumption and expenditure modules. These complex surveys typically need to be administered in several visits, and they often collect data from different respondents in each household. In such surveys, global experience suggests that the success of paperless CAPI will depend substantially on the effort spent on programming, piloting and testing the application; and on the attention paid to the underlying data management and transfer systems. This would therefore have required considerable resources and time to refine the application *before* taking it to the field.

**The significant revisions in the design and field implementation, and the risks associated with a rapid transition to paperless CAPI within the timeframe allowed for implementation of the survey implied that CAPI could not be adopted without risks to quality of data.** While CAPI was considered, a quick move to CAPI posed a critical challenge - it would not allow adequate time for planning, development and beta testing; and this would also pose a risk to the entire field operation. A

technical memo- that detailed the pros and cons of adopting CAPI versus CAFÉ - was presented to help inform this choice for DOS and MoPIC (see Annex 3).

A decision was made to gradually **introduce the efficiencies associated with a decentralized data entry system, by integrating computer-based quality controls to fieldwork through CAFÉ** . This approach would retain paper surveys as backup to the electronic data; and the quality of the data reported on paper questionnaires would be checked while the enumerators are still in the field so that inconsistencies can be solved by direct verification with the households. In this integrated approach, fieldwork was integrated with a concurrent data entry and data quality management system that was customized for the Jordanian context (described in the next section).

## D. FIELD AND DATA MANAGEMENT SYSTEMS FOR CONCURRENT QUALITY CONTROLS

**Implementing a system of concurrent data quality controls relies on the interface between the field and data management systems.** An integrated approach was used to monitor field progress and data quality concurrently through a series of high frequency reports on critical indicators, using these reports to identify bottlenecks and poor performance and implement corrective measures, and incentivizing good performance through a monthly system of performance based awards for the best field teams. As a natural consequence, learning and improvements in field implementation and supervision were heavily concentrated in the pilot phase and in the first quarter, and followed by a stable period of steadily improving markers of quality. This concurrent monitoring sharply reduced the time lag from data collection to availability for analysis, as well as for analysis itself.

**A key element of this TA was focused on designing an interface that was contextualized to the Jordanian context.** In particular, data entry was decentralized to 12 governorate field offices rather than completely decentralized to the field enumerators. This was a workable system as the short distances within governorates made this feasible. At the same time, the time spent by a team in the field was expanded to allow the concurrent checks to be completed before the field team completed an enumeration area. This expansion was in addition to extension of fieldwork to Saturdays and evening hours so as to able to reach working households as well as direct respondents for labor.

**A customized monitoring system using concurrent data quality checks was built to track key indicators of data quality, and effort by enumerator, team and governorate.** These indicators were tracked by *survey period* (every 10 days) for the entire duration of the survey. These included tracking indicators to monitor overall enumerator and team effort, for instance, through tracking the average number of visits per completed household (Table 2). Deviations for a team were tracked relative to the performance of other teams during the same period, and across multiple periods. Other indicators were tracked to monitor whether enumerators exerted adequate effort in key modules. Recording a greater number of respondents or individuals in some modules triggered follow-up questions, which could create incentives for enumerators to record fewer individuals. For instance, if enumerators recorded fewer children currently attending school, they would have to ask fewer follow up questions. Table 3 shows an example of tracking this indicator.

Another key area of focus was the labor module, which was redesigned to secure non-proxy responses to the extent possible. Again, enumerators may have incentives to record more proxy respondents to try and finish up the questionnaire faster. In Table 4 below for instance, team 9 was found to be recording a lower than average number of personal interviews than other teams in the province in the earlier periods, but subsequently increased their effort due to the timely feedback provided from the monitoring system.

Table 2: Average number of visits per completed household, periods 11-29, in governorate X

| Team | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 3.0 | 3.1 | 3.5 | 3.5 | 3.4 | 3.5 | 3.4 | 3.8 | 3.5 | 3.6 | 3.5 | 3.0 | 3.5 | 3.9 | 3.5 | 3.5 | 4.0 | 3.5 | 3.5 |
| 2 | 3.3 | 3.6 | 3.5 | 3.6 | 3.5 | 3.4 | 3.2 | 4.2 | 3.3 | 3.6 | 3.8 | 3.3 | 3.5 | 4.0 | 3.5 | 3.8 | 3.9 | 3.6 | 3.7 |
| 3 | 2.7 | 3.4 | 3.5 | 3.3 | 3.5 | 3.5 | 3.5 | 3.4 | 3.8 | 3.6 | 3.6 | 3.4 | 3.6 | 4.0 | 3.6 | 3.7 | 4.0 | 3.6 | 3.5 |
| 4 | 2.6 | 3.3 | 3.5 | 3.5 | 3.5 | 3.4 | 3.4 | 4.0 | 3.3 | 3.5 | 3.3 | 3.0 | 3.5 | 4.0 | 3.5 | 3.4 | 4.0 | 3.4 | 3.4 |
| 5 | 3.6 | 3.6 | 4.1 | 3.6 | 3.8 | 3.5 | 3.7 | 4.3 | 3.4 | 3.9 | 3.8 | 3.1 | 3.8 | 4.1 | 3.5 | 3.6 | 4.0 | 3.6 | 3.8 |
| 6 | 4.0 | 3.5 | 3.6 | 3.5 | 3.5 | 3.9 | 3.4 | 4.1 | 3.5 | 3.8 | 3.5 | 3.0 | 3.5 | 4.0 | 3.6 | 3.5 | 4.0 | 3.6 | 3.5 |
| 7 | 3.9 | 3.5 | 3.4 | 3.8 | 3.4 | 3.6 | 3.6 | 4.1 | 3.6 | 4.0 | 3.7 | 3.2 | 3.6 | 4.2 | 3.7 | 3.6 | 4.1 | 3.8 | 3.8 |
| 8 | 4.0 | 3.3 | 3.6 | 3.6 | 3.5 | 3.5 | 3.5 | 4.2 | 3.5 | 3.8 | 3.7 | 3.2 | 3.6 | 4.3 | 3.5 | 3.6 | 4.0 | 3.6 | 3.5 |
| 9 | 3.6 | 3.6 | 3.3 | 3.6 | 3.6 | 3.5 | 3.5 | 3.9 | 3.5 | 3.4 | 3.5 | 3.1 | 3.6 | 3.9 | 3.7 | 3.5 | 4.0 | 3.4 | 3.5 |

Table 3: Percentage of persons aged 6-17 currently attending school, periods 14-26, governorate Y

| Team | P14 | P15 | P16 | P17 | P18 | P19 | P20 | P21 | P22 | P23 | P24 | P25 | P26 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1.00 | 1.00 | 0.90 | 1.00 | 1.00 | 0.90 | 0.80 | 0.90 | 1.00 | 0.90 | 1.00 | 0.90 | 0.90 |
| 2 | 1.00 | 0.90 | 0.80 | 0.80 | 1.00 | 0.90 | 0.80 | 0.90 | 1.00 | 0.90 | 1.00 | 1.00 | 0.80 |
| 3 | 0.90 | 1.00 | 0.90 | 0.80 | 0.80 | 0.90 | 1.00 | 0.90 | 0.90 | 1.00 | 1.00 | 0.90 | 1.00 |
| 4 | 1.00 | 0.90 | 0.90 | 1.00 | 0.90 | 0.90 | 1.00 | 1.00 | 0.90 | 0.90 | 1.00 | 0.90 | 1.00 |
| 5 | 1.00 | 0.90 | 0.80 | 1.00 | 0.90 | 0.90 | 0.80 | 1.00 | 1.00 | 0.90 | 0.90 | 0.80 | 1.00 |
| 6 | 1.00 | 1.00 | 0.90 | 1.00 | 0.60 | 1.00 | 0.80 | 0.80 | 1.00 | 0.90 | 0.90 | 1.00 | 1.00 |
| 7 | 1.00 | 1.00 | 1.00 | 0.90 | 0.90 | 0.90 | 1.00 | 1.00 | 0.90 | 0.90 | 1.00 | 1.00 | 1.00 |
| 8 | 0.90 | 1.00 | 1.00 | 0.90 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.80 | 1.00 | 1.00 | 1.00 |
| 9 | 0.90 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.80 | 1.00 | 1.00 | 0.90 | 0.90 | 0.90 | 1.00 |
| 10 | 1.00 | 0.70 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

Table 4: Average number of personal interviews in labor for persons 15 yrs (NO PROXY), periods 14-26, governorate Z

| Team | P14 | P15 | P16 | P17 | P18 | P19 | P20 | P21 | P22 | P23 | P24 | P25 | P26 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0.92 | 0.97 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.91 | 1.00 | 1.00 | 1.00 | 1.00 |
| 2 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 3 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.95 | 1.00 | 1.00 | 0.92 | 1.00 | 1.00 | 1.00 |
| 4 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 5 | 1.00 | 0.98 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 6 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 7 | 1.00 | 1.00 | 1.00 | 0.95 | 0.92 | 0.97 | 0.91 | 0.88 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |
| 8 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.89 | 1.00 | 1.00 | 1.00 | 1.00 | 0.95 | 1.00 | 1.00 |
| 9 | 1.00 | 0.88 | 0.82 | 0.84 | 0.77 | 0.82 | 0.95 | 0.97 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 |

The steady improvements in enumerator effort and the effective feedback from data quality checks to improved fieldwork and supervision are evidenced in improvements in response rates across survey periods, especially in highly urbanized environments such as Amman governorate, and the southern governorates of Maan and Aqaba, which had faced much higher rates of non-response in past survey rounds.

Table 5: Response rates by governorate, Periods 1-4 vs 11-14

| Governorate | Total Rate completion - P1 | Total Rate completion - P2 | Total Rate completion - P3 | Total Rate completion - P4 | Total Rate completion - P11 | Total Rate completion - P12 | Total Rate completion - P13 | Total Rate completion- P14 |
|---|---|---|---|---|---|---|---|---|
| Amman | 86% | 92% | 86% | 88% | 85% | 99% | 99% | 98% |
| Balqa | 94% | 100% | 100% | 97% | 100% | 100% | 100% | 100% |
| Zarqa | 88% | 100% | 100% | 95% | 100% | 98% | 98% | 100% |
| Madaba | 96% | 100% | 100% | 100% | 100% | 100% | 100% | 100% |
| Irbid | 86% | 99% | 100% | 99% | 100% | 97% | 97% | 97% |
| Mafraq | 100% | 100% | 100% | 100% | 94% | 100% | 100% | 100% |
| Jerash | 100% | 96% | 94% | 97% | 100% | 100% | 100% | 100% |
| Ajloun | 97% | 91% | 100% | 91% | 97% | 88% | 88% | 97% |
| Karak | 75% | 100% | 100% | 97% | 100% | 97% | 97% | 96% |
| Tafilah | 97% | 83% | 94% | 96% | 96% | 100% | 100% | 100% |
| Maan | 63% | 67% | 94% | 78% | 88% | 97% | 97% | 88% |
| Aqaba | 59% | 50% | 81% | 81% | 75% | 78% | 78% | 100% |
| Total | 87% | 92% | 94% | 92% | 93% | 97% | 97% | 98% |

## 3. KEY OUTCOMES

This TA program successfully supported a transition from diary to recall methods, alongside important changes in sampling design, and field and data management systems. In addition to the overall outcome of producing high-quality data across the board, and setting new standards on the concurrent availability of data for analysis, two other major outcome are highlighted below.

**First, it set a new standard in terms of survey response rates that may be attainable in the context of a middle-income country and for the MENA region.** The TA resulted in record response rates for surveys of this nature, well above the norm for middle income countries and countries in the MNA region. Response rates grew steadily over survey quarters, from 93 percent in quarter 1 to 97 percent in quarter 4 (Figure 2). These high response rates were achieved across all governorates in Jordan, including in highly urbanized settings such as Amman, and governorates in the South (such as Maan and Aqaba) where response rates have traditionally been far lower (Figure 3). For the whole survey, a response rate of over 95% was reached, with a total sample of 19,216 households.

**Another major achievement was the success of the non-proxy labor module.** On average, 90 percent of respondents were direct participants in the labor market, self-reporting on a range of labor market outcomes. This is a huge improvement in the quality of labor market statistics in contexts

such as Jordan, where a significant proportion of employment is informal, and is likely to be mis- or under-reported when labor modules are answered by proxy. The HEIS 2017/18 was also characterized by no item non-response, due to improvements in survey design, enhanced field protocols and supervisions, and a revamped data entry program.

Figure 2: Quarterly response rates, HEIS 2017/18

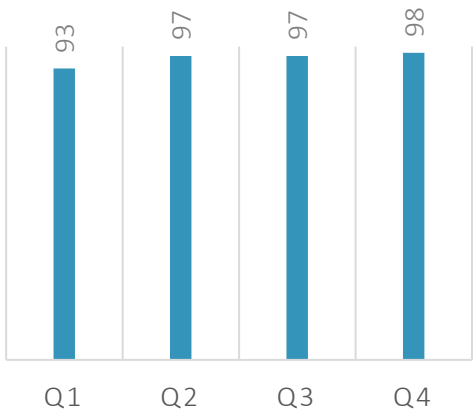



Figure 3: Governorate response rates, HEIS 2017/18

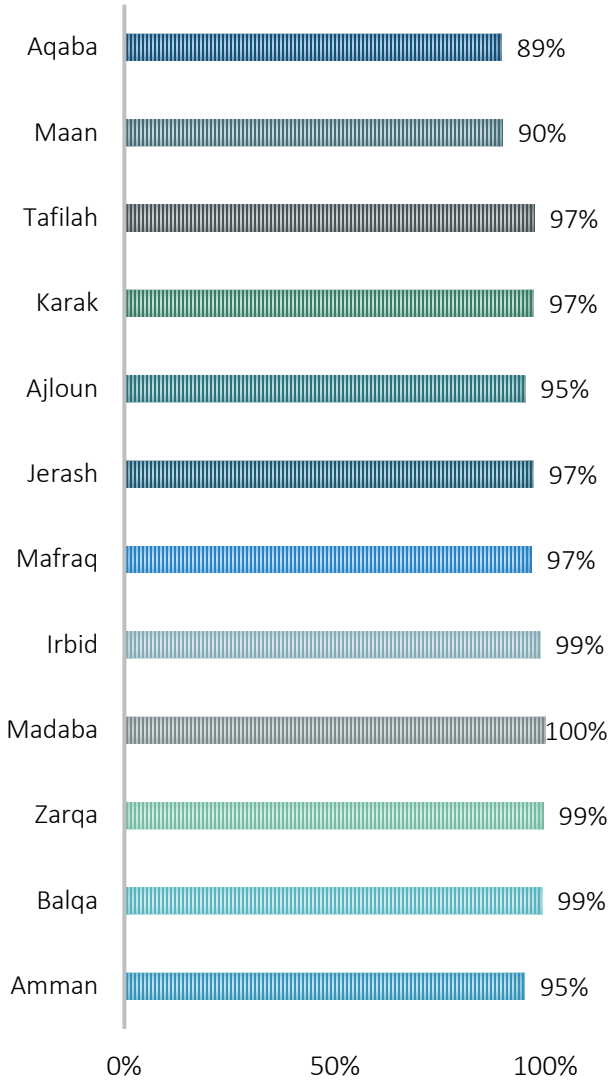
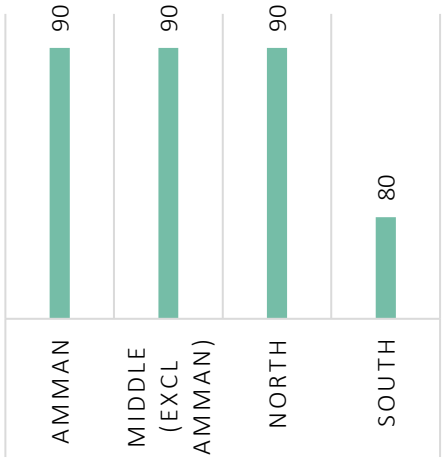



Figure 4: % Labor market respondents non-proxy



# 4. KEY ELEMENTS OF TA THAT RESULTED IN THESE OUTCOMES

In conclusion, this note describes two key elements that were important to the success of this TA: (i) an embedded program of Technical Assistance that supported a process of learning-by doing to build client buy-in and capacity; and (ii) the design of a performance-based system of awards that was designed to activate the regular use for data monitoring tools by the client teams.

**To deliver on timely, high quality information while strengthening long term capacity in DOS this TA program invested heavily in providing hands-on support in the early phase field implementation and data management.** This included designing feedback loops between the two, to inform course-corrections in real time; and implementing a system of performance-based awards for field enumerators, which was used to both incentivize data quality and activate the feedback loop to work on a regular and systematic basis. Technical assistance in the first survey quarter therefore involved an equal combination of hard investments in transferring technical best practices, and soft investments by embedding a resident field coordinator alongside significant field presence from the task team. This hands-on assistance was gradually reduced over the third survey quarter, with DOS taking over full implementation in the final survey quarter. Importantly, the quality and frequency of monitoring remained the same thorough the survey year.

**This embedded model was directly linked to the success of multiple preparatory activities, and the success of this TA in building sustainable capacity to manage a high-quality data collection effort.** In particular, this approach supported a two way-learning process: for the Bank team on constraints faced by the client team, and for the clients on pros and cons of different technical options. This process created a partnership, where best practices were adapted to fit with implementation context; quality standards where both enforced and incentivized; and the clients assumed complete ownership of these changes. Examples of this include the successful implementation of two important pilot activities for the new consumption and labor modules; customization of principles that underlie best CAFÉ practices to set up a field and data management system for the Jordanian context. Throughout the preparatory phase, key decisions were made using a consensus building approach where the Bank team provided the technical rationale for the preferred options highlighting the pros and cons, which the Dos team could deliberate and ultimately make the decisions.

**While data quality reports are generated in a wide variety of contexts, it is often the case that they are not used systematically to monitor data quality**; and it is even less often the case that client teams take ownership of a systematic process monitoring process through these reports. In the case of Jordan one of main objectives of the TA program was to activate the use of the monitoring system (systematically) by the client team. To strengthen the incentives to actively use these data quality reports by the survey management team, and for field teams and supervisors to respond to feedback and deliver quality data, a system of performance-based awards was designed. The awards were designed based on the data quality report for quarter 1, using key metrics of effort and data quality, and aggregated to create scores for each field enumerator and team.

**The first awards were given for quarter 1, and thereafter, to the best field team in each of the 4 regions each month.** Awards were given per region to recognize heterogeneity in field conditions.

Within governorates, teams were allocated randomly to clusters, so that observed differences over time could be attributed to team performance rather than particularly challenging or easy field conditions. In addition, every quarter, governorate supervisors were awarded for good performance, and the best field team over each quarter was further recognized with a cash award.

**Finally, key changes in the governance and organizational structure of the HEIS survey team - that took a careful and continuous process of negotiation - were critical in being able to translate this combination of technical inputs into high quality data survey.** In particular, the programs' success in breaking through the status quo practices at DOS where each department functioned in separate silos (Survey, Information Technology, Human Resources, etc.) to create a cross-departmental team that was collectively responsible for the different inter-linked components of survey design and implementation played a critical role in translating the TA investments into results on the ground.

**The comprehensive TA approach from design to implementation yielded high quality data and delivered it for analysis within 4 weeks of completion of the survey.** The ultimate goal of data is to provide timely evidence to inform policy and strategy. This TA by virtue of its design innovation and data management structure was able to respond to key requests made by the Government for insights from the data during intervening survey quarters. Moreover, after the survey implementation was completed in early August, key options for welfare measurement methodology along with the results for each proposed method through analysis of the HEIS 2017/18 survey round was presented to senior Government officials by mid-September 2018. The TA has thus set a record for the *least* time taken between data collection and analysis: across the developing world the time lag between collection and analysis takes anywhere between 6 months in the best case-scenarios and up-to 2 or 3 years in many country contexts.

## ACKNOWLEDGEMENTS

This TA program was implemented by the World Bank's Poverty and Equity team in partnership with the Jordanian Department of Statistics.

From DOS, the team that worked on the HEIS 2017 was ably led by Fatmeh Al Awamreh, Sana Al Momani, and Rania Abu Dhaim; with overall guidance from Qasem Al Zoubi (Director General, DOS). The openness of the team that led this survey to the proposed changes, as well as their dedication to implementing these changes over the entire survey year was instrumental in the ability of this TA program to deliver its far-reaching results.

From the World Bank, this Technical Assistance Program was led by a task team comprising of Nethra Palaniswamy (Economist) and Tara Vishwanath (Lead Economist). A team of technical specialists provided sustained support: Valerie Evans (Survey Specialist); Beatriz Godoy (Survey Specialist); Andrea Germiniasi (Poverty Measurement); and Juan Munoz (Sampling Expert). Meghan Lucas was based in Amman as a field coordinator for the task team for 6 months before the survey, and during the first two survey quarters.

# Annex 1: SAMPLING DESIGN AND SAMPLING WEIGHTS

This annex describes the sampling design of the Jordan Household Expenditure and Income Survey (HEIS) that the Jordan Department of Statistics (DoS) fielded for 12 months, from mid-2017 to mid-2018.

## *Total sample size and sampling stages*

The total nominal sample size was 20,000 households, selected in two stages. In the first stage, 2,500 Census Blocks (CBs) were selected, using as a sample frame the list of CBs developed by the 2014 Census. In the second stage, 8 households were selected in each CB using an updated list of households in the CB as a sample frame. Both stages we random selections.

## *Stratification*

The first sampling stage used Jordan's 12 governorates as explicit sampling strata. The sample of 2,500 CBs was allocated into governorates as shown in Figure 1 below, with the intention of delivering estimates of comparable quality for each governorate, as well as estimates of good quality for the Kingdom as a whole.

Annex Figure 1: Number of Census Blocks allocated to the 2017 HEIS, by governorate

| Governorate | Number of CBs |
|---|---|
| Amman | 702 |
| Balqa | 152 |
| Zarqa | 252 |
| Madaba | 133 |
| Irbid | 314 |
| Mafraq | 145 |
| Jarash | 135 |
| Ajloun | 133 |
| Karak | 139 |
| Tafeila | 130 |
| Maan | 132 |
| Aqaba | 133 |
| Total | 2,500 |

Within each governorate, the sample of CBs was selected with probability proportional to size (PPS), using the total number of households reported by the 2015 Census as a measure of size, and with implicit stratification by district and sub-district.

The second stage used the nationality of the household head, as reported by the listing operation, as strata, allocating the 8 households sampled in each CB into three categories:

- Jordanian or unknown: 4 households
- Syrian: 2 households
- Other nationalities: 2 households

with explicit rules for the rare situations in which the total number of households listed in any of the categories was less than the nominal sample sizes. The listing operation was supported by an electronic application that automatically recorded the coordinates of each dwelling, to let DOS assure its completeness; and transferred the data to headquarters, to have the sample of 8 households selected. Within each nationality category, the target sample was selected with equal probabilities, and a few additional households were designated as a reserve, to be used if nonresponse.

### *Allocation of the sample to fieldworkers and in time – teams, periods and trios*

The survey was conducted by fieldwork teams, each composed of three interviewers and headed by one supervisor; and based in a specific governorate. The 12 months of fieldwork were divided into four quarters and further into 9 periods of 10 days. Each interviewer was responsible for visiting one CB and interviewing the eight households selected there in each period. The sample was allocated into teams and periods was done as follows:

- First, the 2,500 selected CBs were randomly divided into four groups of 625 CBs, one for each quarter.
- Second, the 625 CBs of each quarter were grouped into *trios*, each composed of three neighboring CBs in the same governorate and expected to be visited by one team in one period.
- Third, the trios were randomly allocated into periods and, in the governorates with more than one team, also randomly allocated into teams.

### *Selection probabilities and sampling weights*

In the absence of nonresponse, the probability $p_{gij}$ of selecting a household of nationality $j$ in CB $gi$ of governorate $g$ is given by

$$p_{gij} = \frac{k_g n_{gi}}{N_g} \times \frac{m_{gij}}{n'_{gij}} \qquad [1]$$

where

$k_g$ is the number of CBs allocated to the governorate (as per Figure 1),

$n_{gi}$ is the number of households reported by the 2015 Census in the CB

$N_g$ is the total number of households reported by the 2015 Census in the governorate,

$m_{gij}$ is the number of households of that nationality interviewed in the CB, and

$n'_{gij}$ is the number of households of that nationality listed in the CB.

The two factors on the right-hand side of Formula [1] respectively represent the probability of selecting the CB, and the probability of selecting the household in the CB.

To obtain unbiased estimates from the sample, the answers of that household should be affected by a sampling weight $w_{gij}$, equal to the inverse of its selection probability:

$$w_{gij} = \frac{N_g}{k_g n_{gi}} \times \frac{n'_{gij}}{m_{gij}} \qquad [2]$$

*Adjustment for nonresponse due to listing errors*

The household listing operation should have in principle listed only occupied dwellings but, as is often the case, some unoccupied dwellings, or even some structures that were not dwellings, were occasionally reported as such. When a household from the target sample was replaced by a reserve household for that reason, the number $n'_{gij}$ of households of nationality $j$ listed in the CB was replaced by an estimation of the correct number $n''_{gij}$, computed as $n''_{gij} = n'_{gij} \frac{(m_{gi} - r_{gi})}{m_{gi}}$, where $m_{gi}$ is the number of households of all nationalities that the survey tried to contact in the CB, and $r_{gi}$ is the number of households of all nationalities replaced for that reason in the CB.

*Collapsing Census Blocks*

To reduce the variability of weights due to the small number of households of certain nationalities observed in some CBs, neighboring CBs were collapsed into groups of approximately twenty CBs; the second factor of Formula [2] was computed for these groups (rather than for individual CBs), and then applied uniformly to all CBs in each group.

*Weight scaling*

Adopting the practice followed by the Demographic and Health Survey (DHS) conducted by the DoS in 2017, the HEIS sampling weights were finally affected by appropriate scaling factors, to have the sum of weights match the sample size in each governorate.

## Annex 2: INFORMING THE DESIGN OF THE FOOD RECALL MODULE, HEIS 2017/18

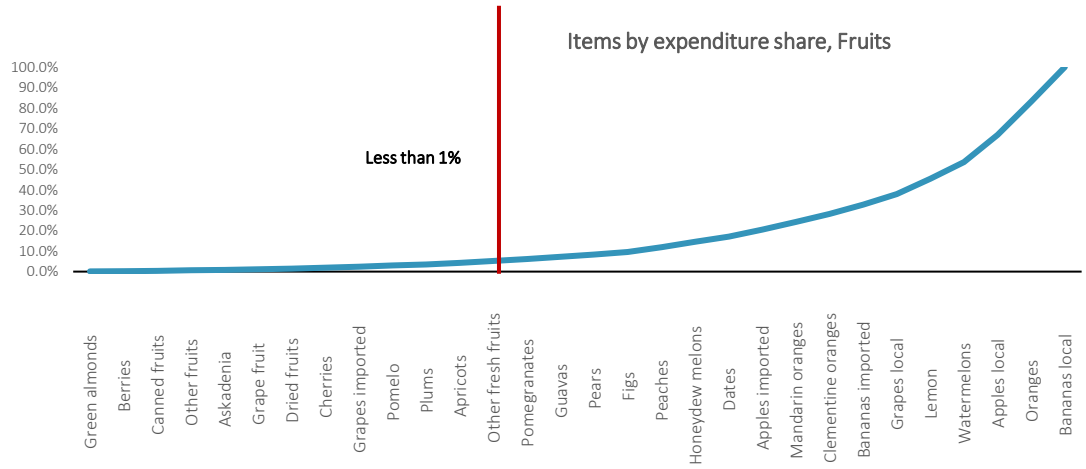
The 2010 HEIS collected food expenditures and consumption through a 7-day diary, and included 227 food items, grouped into 14 broad categories. This data provided the basis for informing the design of the food recall module implemented in the HEIS 2017/18. In particular, this data was examined from the point of view of reducing the number of items to a more manageable list, by being able to aggregate and map these items into new, homogenous categories.

Item homogeneity was considered across 4 dimensions: (i) price (unit value); (ii) caloric content; (iii) share of consumers; and (iv) expenditure shares. In addition, these dimensions were considered over time (survey months and quarters) for consistent patterns. Principal component analysis (PCA) was employed to identify "clusters" of items in a scatterplot, using the first two principal components. In addition, several items were considered on a case-by-case basis. For instance, eggs and milk are clearly distinguishable and were kept as separate categories, while fairly homogeneous items such as canned tuna and canned sardines were grouped into canned fish. Items with very low consumption, with less than 1% of households reporting any consumption, were grouped to other items (if a homogeneous item was available) or added to the residual category within the overall group (other cereals, other vegetables etc). Some items were grouped based on additional discussions with the DOS team.

For example, the table below includes items that have very similar caloric content but vary based on frequency of purchase. In this case, only beetroot would be reassigned to the residual category or grouped with a similar item.

| Item | Unit value | Calories | Frequency |
|---|---|---|---|
| Blackeye peas | 0.90 | 460 | **0.5%** |
| Green beans | 1.10 | 460 | 6.9% |
| Beetroot | 0.67 | 470 | **0.1%** |
| Carrots | 0.59 | 420 | 11.1% |

Finally, within each of the major categories, only those items whose combined expenditures represented about 95% of the total expenditure for the category were kept. This procedure was implemented from the smallest expenditure share items, for example, excluding starch, which accounted for 0.13% of expenditures in the Cereals and products category, then the next largest item, until the remaining items in the category accounted for 95% of the total expenditure in that category. The figure below shows items that would qualify for elimination under this rule, within the broad category of fruits.




Following this case-by-case analysis guided by the PCA, and the elimination of small expenditure share items by category further 233 items that were reported in food diaries in 2010 were consolidated into 191 items used in the recall module 2017.

Throughout the process, care was also taken that the elimination of items or their grouping would end up with information consistent with existing CPI categories. For instance, for the category of cereals, extra care had to be taken to ensure that the aggregation procedure did not lead to a loss in information. For instance, Wheat grains, Packed Flour, Local wheat flour, Semolina, would all be grouped into the other cereals category based on their low expenditure shares, but this would have been difficult to implement because of the diversity of items in the residual category and because these were separate items in the CPI.

| Code | Description | Case by case | Share | CPI |
|---|---|---|---|---|
| 101 | Wheat grains | | | Wheat grains |
| 102 | Rice, ordinary grain | Rice, ordinary grain | Rice, ordinary grain | Rice, ordinary grain |
| 103 | Rice, long grain | Rice, long grain | Rice, long grain | Rice, long grain |
| 104 | White flour | White flour | | White flour |
| 105 | Local wheat flour (Baladi) | | | Local wheat flour (Baladi) |
| 106 | Packed flour | | | Packed flour |
| 107 | Whole wheat bread (Baladi) | Whole wheat bread (Baladi) | Whole wheat bread (Baladi) | Whole wheat bread (Baladi) |
| 108 | White wheat bread | White wheat bread | White wheat bread | White wheat bread |
| 109 | Bread other types | Bread other types | Bread other types | Bread other types |
| 110 | Burghul (boiled crushed wheat) | Burghul (boiled crushed wheat) | | Burghul (boiled crushed wheat) |
| 111 | Freekeh (burned wheat) | Freekeh (burned wheat) | Freekeh (burned wheat) | Freekeh (burned wheat) |
| 112 | Semolina | | | Semolina |
| 113 | Pasta/ vermicelli (angel hair pasta) | Pasta/ vermicelli (angel hair pasta) | Pasta/ vermicelli (angel hair pasta) | Pasta/ vermicelli (angel hair pasta) |
| 114 | Ka'ak (Local cakes & biscuits) | Ka'ak (Local cakes & biscuits) | Ka'ak (Local cakes & biscuits) | Ka'ak (Local cakes & biscuits) |
| 115 | Biscuits | Biscuits | Biscuits | Biscuits |
| 116 | Children's biscuits | Children's biscuits | Children's biscuits | Children's biscuits |
| 117 | Starch | Starch | | Starch |
| 118 | Baby food (Cerelac) | Baby food (Cerelac) | | Baby food (Cerelac) |
| 119 | Qatayef (local sweets) not stuffed | Qatayef (local sweets) not stuffed | Qatayef (local sweets) not stuffed | Qatayef (local sweets) not stuffed |
| 120 | Custard | Custard | | Custard |
| 121 | Corn (assorted) | Corn (assorted) | | |
| 122 | Fee for grinding grains and foodstuff | Fee for grinding grains and foodstuff | | |
| 123 | Other | | | |
| | | Wheat grains, Packed Flour, Local Wheat Flour, Semolina | Wheat, flour, corn, other grains, and grinding fees | |

## Annex 3: MEMO ON HEIS DATA QUALITY: THE CHALLENGE OF INCREASING SAMPLE SIZE AND IMPLEMENTING TABLET BASED SURVEYS

The government of Jordan has requested the World Bank for technical assistance in a full overhaul of the design and implementation of the upcoming round of the Household Expenditure Income Survey (HEIS), with the overarching objective of producing high quality data to support the government in its efforts to measure, monitor and ultimately reduce poverty. In the context of a wrap-up meeting to update His Excellency the Minister of Planning and International Cooperation

(MOPIC) on progress made so far in preparatory work, two important issues were raised. The World Bank was asked to prepare a short memo on the following two issues:

(i) The pros and cons of expanding the sample from governorate level representation to sub-district level representation; and

(ii) The pros and cons of using tablets to implement Computer assisted Personal Interviews (CAPI) to directly enter responses of households and individuals as they are interviewed versus Pen and Paper interviews (PAPI). This is to ensure the decision on data management systems that the Department of Statistics (DOS) is considering at the moment is an informed choice that explicitly tries to minimize risks. DOS currently has a strong preference for a CAPI approach.

The bottom line question in both cases is: What is the best way to arrive at the *highest quality data* that supports credible estimates of poverty and other living conditions in the Kingdom? Below we address each of these issues in turn:

## Expanding the household sample to be representative at the sub-district level

During the current mission, the team worked with DOS's sampling division to examine the master sample frame based on the most recent census of 2015. The preparatory work done for sampling included the construction of a very detailed list of every single block of the Kingdom (approximately 20,000 blocks) along with the geographic identifiers (from governorate to block level) and population shares (Jordanians and non-Jordanians[3]). This list defines the master sample frame. Based on this frame, the team worked to first select a household sample that will be representative at the governorate level.

The previous two rounds of the HEIS in 2010 and 2013/14, with a sample size of approximately 14,000 and 24,000 households respectively were claimed to be representative at the sub-district level. During the analytic phase of deriving poverty estimates with the World Bank team, these samples proved to be inadequate at the district and sub-district levels. Furthermore the 2013/14 round also yielded very noisy estimates even at the governorate level for some governorates.

For governorate level representation of Jordanians and non-Jordanians, the sample size calculations suggest that the HEIS should visit a total sample of about 20,000 households, randomly selected in two stages. In the first stage – which was completed during this mission – 2,500 blocks were selected, using the master sample frame. In the second stage, 8 households – 6 Jordanians and 2 Non-Jordanians – will be selected in each block, using a full updated listing of all households in the block as a sample frame.

The 2,500 blocks were allocated into the 12 governorates according to Kish's allocation, in order to bring about accurate estimates both for each governorate and for the Kingdom as a whole, as shown in the first columns of Table 1 below. The last columns give maximum margins of errors that

[3] Excluding the Non-Jordanian population in camps

this sample is expected to deliver in each domain, for the estimation of any prevalence (such as the poverty rate).

*TABLE 6: ALLOCATION OF THE 2017 HEIS SAMPLE BY GOVERNORATE AND TYPE OF POPULATION, AND MAXIMUM MARGINS OF ERROR EXPECTED FOR THE ESTIMATION OF PREVALENCE*

| | | Sample size (households) | | | Margins of error | | |
|---|---|---|---|---|---|---|---|
| Governorate | | Jordanians | Non-Jordanians | Total | Jordanians | Non-Jordanians | Total |
| 11 | Amman | 4239 | 1379 | 5618 | 1.8% | 2.8% | 1.5% |
| 12 | Balqa | 937 | 279 | 1216 | 3.9% | 6.2% | 3.4% |
| 13 | Zarqa | 1539 | 478 | 2017 | 3.1% | 4.7% | 2.6% |
| 14 | Madaba | 845 | 219 | 1064 | 4.1% | 6.9% | 3.7% |
| 21 | Irbid | 1940 | 573 | 2513 | 2.7% | 4.3% | 2.3% |
| 22 | Mafraq | 899 | 261 | 1160 | 4.0% | 6.4% | 3.4% |
| 23 | Jerash | 842 | 238 | 1080 | 4.1% | 6.7% | 3.5% |
| 24 | Ajloun | 864 | 200 | 1064 | 4.1% | 7.3% | 3.8% |
| 31 | Karak | 890 | 222 | 1112 | 4.0% | 6.9% | 3.6% |
| 32 | Tafilah | 887 | 153 | 1040 | 4.0% | 8.3% | 3.9% |
| 33 | Maan | 880 | 176 | 1056 | 4.0% | 7.7% | 3.7% |
| 34 | Aqaba | 815 | 245 | 1060 | 4.2% | 6.6% | 3.6% |
| Total | | 15577 | 4423 | 20000 | 1.1% | 1.8% | 0.9% |

Source: Census **2015– Jordan**

For smaller geographic areas, a much larger sample size will be required - we will require a sample of around 65,000 households to deliver credible poverty estimates for each of the Kingdom's 51 districts; and a sample of over 100,000 for each of its 89 sub-districts.

In summary moving from governorate level representation to sub-district entails a five-fold increase in sample size. Even if resources to field a survey of that size were available, the effort is more than likely to be counterproductive. While sampling errors might be acceptable, that would not compensate for the increase in non-sampling errors, which will emerge on account of the difficulties of managing and implementing such a large field operation. Non-sampling errors can manifest themselves in various ways, all of them damaging to data quality and ultimately to the credibility of the HEIS poverty estimates. These errors include but are not limited to interviewer-induced non-response, misreporting or underreporting due to perverse incentives, poor understanding or misinterpretation of the questionnaire, and inadequate or untimely supervision.

While the risk of non-sampling error is present in a sample of any size, these risks grow exponentially with sample size. The essence of the risk comes from the increase in the number of enumerators required to field this survey within a one-year period; and managing the quality of the data collection process in the field. For the case of 20,000 households, which will yield representation at the governorate level, a high-quality data collection effort can be managed with around 50 interviewers. For a survey that is five times larger, the required increase in effort will not just be a linear increase in the number of enumerators to 250. This is because data quality depends critically on survey management. DOS is aware of this challenge from its previous

experience in 2013/14 when 200 enumerators were used to be able to visit around 24,000 households 17 times on average. Survey management involves three components- recruiting and training enumerators, supervising enumerators in the field, and managing the entire field implementation process. The human capital required for each of these three components is large, and moreover the management challenge grows exponentially with sample size.

The process of hiring enumerators and training them will require recruiting a larger pool (about double the required number), as many refuse or drop out midway. Identifying as many suitable candidates will not be easy, given the demands of a survey such as the HEIS. The candidates should have the required background and motivation, the stamina to face the hardships of fieldwork, and the willingness to do it for a long year. All recruited enumerators will also need to be trained for an estimated three-week period, which will require serious logistics and investment. In addition, enumerators will need to be supervised intensively throughout the year. The proposed sample at the governorate level will require around 10 supervisors, who will also need to be trained as master trainers. The management of this larger number of field supervisors would also likely require an additional administrative layer for quality control.

The survey management process is therefore very complex. Scaling up the field operation to cover a significant increase in sample size is not a linear problem; and it is likely to induce a massive increase in non-sampling errors. Many countries have opted out of large increases in sample size, either for budgetary reasons or because they have witnessed the negative implications of doing so. Global experience has also demonstrated that drastic increases in sample size can be dangerous. Two illustrative examples are provided below:

- Tanzania decided in the early 2000s to expand the sample of its Household Budget Survey (HBS) from 5,000 to 26,000 households, in order to deliver sub-national estimates for each of its 30 regions. Management became so difficult that the rural part of the survey was interrupted after about six months. A post-mortem examination of the datasets revealed that, among other indicators of unsupervised fieldwork, the average household size had dropped by more one person per household between the first and the last survey month.
- Honduras, Guatemala, el Salvador and other Latin-American countries decided in the late 1990s to incorporate consumption modules to their Labor Force Surveys (LFS-which are simpler surveys, typically conducted on samples of around 10,000 households), converting them into de facto Household Budget Surveys (HBS). Unfortunately, management standards could not rise to challenges imposed by the added complexity of reporting incomes and expenditures, and the quality of both the consumption and the labor data decreased dramatically as a result. Most of these countries have now reverted to conduct their regular LFSs and periodic HBSs as separate exercises.

The demand for district and sub-district level information is however understandable for policy makers. Poverty mapping techniques recently developed by researchers at the World Bank and elsewhere offer the option of delivering detailed sub-national figures without resorting to oversized samples.

Briefly, the poverty mapping methodology involves combining the depth of information in a survey like the HEIS with the complete spatial coverage available from a census. Basically, we impute into population census data—which does not have consumption data—a measure of per capita consumption from household survey data—which has a sample too small for small area disaggregation. We estimate an expenditure or consumption model using a set of explanatory variables that are common to both the household survey and the census (e.g. household size, education, housing and infrastructure characteristics and demographic variables). These estimations are first limited to the lowest geographical level at which the survey data is representative. Then these estimates from the model (including the estimated error terms associated with those coefficients) are used to predict expenditure or consumption for every household in the census. These household-unit data are then used to compute poverty estimates for small areas.

The quality of the output of the poverty mapping methodology is constrained by *the quality of the inputs*, that is, the census and household survey data sets. If there are problems with either data set, these problems will not disappear when the data sets are used to produce the map. Many countries have used small area estimation to create local welfare estimates and poverty maps. The list of countries that are using the poverty mapping methodology has been growing over time and it has been implemented in countries far afield as Palestine, China, Indonesia, Brazil, South Africa and the European Union. A subset of these countries has been improving their census and surveys to refine the modelling and predictions and some countries in Asia and Latin-America have also produced poverty maps overtime.

In Jordan, we have a unique opportunity to create the appropriate links between the 2015 census and the next round of the HEIS and strive towards a high-quality poverty map. This may be the only option to get sub-district poverty estimates without compromising the quality of the HEIS on account of a massive expansion in sample size.

## Pros and Cons of paperless Computer Assisted Personal Interviews (CAPI)

Across the world, rapid progress in technology and its increasing affordability have many governments clamoring to use electronic data collection tools based on potential advantages associated with time, cost and quality. Most of these perceived advantages are often based on the performance of the technology aspect alone. Data quality however depends on the interface of field management systems with technology, which rely heavily on human capital and management.

Most computer-assisted survey applications in developing countries have typically been in simpler surveys (e.g.: market research, opinion polls), or in surveys that are smaller in scale or scope (e.g.: labor force surveys, census). However, household surveys like the HEIS are complex multi-topic surveys with very detailed food and non-food consumption and expenditure modules. These complex surveys typically need to be administered in several visits, and they often collect data on different respondents in each household. In such contexts, there is both very limited experience with evidence of success in computer-assisted personal interviews (CAPI).

In large part, this is because tablet-based CAPI requires an enormous amount of preparatory work and extensive beta testing of the survey application. The United States, which started

implementing CAPI for household consumption and expenditure surveys took several years to prepare for such a transition. Countries like Ethiopia that are currently considering transitioning, have postponed their survey round by 18 months in order to carefully pilot, beta-test and assess its feasibility.

Absent such a careful and elaborate preparatory phase, adopting tablet based CAPI is likely to be accompanied by important drawbacks:

- While the time lag between data collection and availability is reduced by CAPI, preparing and beta testing necessitates a longer set up time. In addition, the need to continually adapt computer-assisted formats presents a challenge and increases the likelihood of failures if adaptions to the newer devices and programs are not fully successful.
- Data-entry errors by interviewers cannot be caught (unless an invalid entry is caught by a range or consistency check). However, these types of checks are difficult to specify in complex modules like consumption expenditures, which are used for estimating poverty.
- Because of pressures to complete the survey quickly when in the household, CAPI can also result in more data entry errors compared to other modes of data entry.
- Interview times for the consumption modules are also likely to be longer with an electronic questionnaire that with a paper questionnaire. This is an important consideration, because these modules are vulnerable to respondent fatigue, and longer interviews might trigger an upsurge of non-response.
- Certain important quality assurance actions, such as supervision check-up visits to the households are much harder to implement with an electronic questionnaire than on paper.

In the case of Jordan, the next round of the HEIS is proposing many fundamental changes. These include changes in sampling design, moving from a diary to recall method for food and non-food consumption, and moving away from centralized data entry in order to improve data quality and reduce the time lag from data collection to its availability.

In this context, adopting a quick move to CAPI will pose a critical challenge- it will not allow adequate time for planning, development and beta testing; and this will also pose a risk to the entire field operation. Instead, there are other ways to gradually introduce the efficiencies associated with a decentralized data entry system, by integrating computer-based quality controls to fieldwork. This approach will not eliminate paper questionnaires entirely; and the paper surveys will serve as backup to the electronic data. The quality of the data reported on paper questionnaires can also be checked while the enumerators are still in the field and inconsistencies can be solved by direct verification with the households.

In this integrated approach, interviewers will be provided with laptops or tablets with keyboards along with a well-developed data entry program. An integrated approach will also help to:

- Improve the quality of the information, because errors and inconsistencies will be corrected factually rather than by guesswork by an auditor; and it will foster the application of uniform criteria by all the interviewers throughout the whole period of data collection.
- Eliminate the need to “clean” the datasets. Data cleansing should be avoided because it is lengthy and error-prone, due the myriad of undocumented decisions involved in the process.

- Generate databases that are ready for tabulation and analysis in a timely fashion. In fact, this can be done as the survey is conducted. This will also give survey managers the ability to effectively monitor field operations.

Global experience suggests that the success of paperless CAPI will depend substantially on the effort spent on programming, piloting and testing the application; and on the attention paid to the underlying data management and transfer systems. This requires considerable resources and time to refine the application *before* taking it to the field. Without such preparation, paper-based surveys may well outperform CAPI. Decentralized data management systems which integrate computer controls to field work and deliver data within a similar time frame as a paperless CAPI approach have been implemented successfully for complex questionnaires like the HEIS for over three decades across more than 50 countries. Most countries that have successfully transitioned to CAPI for complex surveys have done so after a significant period of testing; and with risk management strategies that do not completely do away with paper surveys overnight. This is because while technology can be an important enabling tool, it cannot function well without a well-managed human interface.

Jordan needs new poverty estimates within a short time frame, in order to update the previous estimates which are now 6 years old, and to better understand the welfare and living standards of Jordanians in a time of a large population influx. The significant revisions in the design and field implementation of the next HEIS round, and the risks associated with a rapid transition to paperless CAPI within a time frame that is not likely to allow for sufficient testing pose significant risks of compromising the quality of the HEIS.