



## INCEPTION REPORT

### Data Entry Program, Data Management & Integration, & Field-Based Data Entry Training and Support

Contract # 7154940

Kimetrica International Limited  
2020 Pennsylvania Ave, NW #715  
Washington, DC 20006  
[www.kimetrica.org](http://www.kimetrica.org)

Tel: +1 202 470 6974  
Fax: +1 202 280 1272

William (Rob) Rose, Director  
Email: [rob.rose@kimetrica.org](mailto:rob.rose@kimetrica.org)

*Prepared for:*  
The World Bank  
Attention: Bertha Briceno & Claire Chase  
1818 H Street, NW  
Washington, DC 20433

Tel: 202-473-4111  
Tel: 202-473-2928

21 September 2010

## Table of Contents

List of Abbreviations and Acronyms .....	3
1 Introduction .....	4
2 Objectives .....	4
3 Scope of Work .....	4
4 Expected Outputs .....	5
5 Deliverables .....	6
5.1 Inception Report (Deliverable #1) .....	6
5.2 Data Handling Manual (Deliverable #2) .....	7
5.3 Review of Country-Specific Questionnaires (Deliverable #3) .....	8
5.4 A ready-to-use, fully documented Data Reduction System (Deliverable #4) .....	10
5.5 Assess LSF' Field Implementation Plan (Deliverable #5) .....	11
5.6 Data Entry System Testing, Training, and Support (Deliverable #6) .....	11
5.6.1 Data Entry System Testing .....	11
5.6.2 Data Entry Staff Training .....	12
5.6.3 Data Entry Supervisor Training .....	13
5.6.4 Post-Training .....	14
5.7 Weekly Summary Reports (Deliverable #7a) .....	14
5.8 Assessment Report on Data Quality Cleaning and Basic Descriptive Statistics (Deliverable #7b) .....	15
5.8.1 Field Editing .....	15
5.8.2 Manual Office Editing .....	15
5.8.3 Automated Office Editing .....	15
5.8.4 Cross Tabulations .....	16
5.8.5 Automated Quality Checks .....	16
5.9 Integrated Baseline and Endline Datasets (Deliverable #7c) .....	16
5.10 Documented Public Use Files (PUF) (Deliverable #7d) .....	17
5.11 Recommendations on Platform (Deliverable #7e) .....	17
6 Work Procedures (Deliverable #1, continued) .....	18
7 Roles & Responsibilities (Deliverable #1, continued) .....	20
7.1 World Bank/Washington, DC and Kimetrica .....	20
7.2 World Bank/Washington, DC and In-Country Principal Investigator (PI) .....	21
7.3 In-Country PI and Kimetrica .....	21
7.4 World Bank/Washington, DC and Local Survey Firm .....	21
7.5 Kimetrica and Local Survey Firm .....	21
8 Revision of Schedule of Payments .....	22
Appendix A: Project Timeline .....	24
Appendix B: Field Check Tables .....	29

## List of Abbreviations and Acronyms

ASCII	American Standard Code for Information Interchange
CSPPro	A census and survey data processing package developed by the U.S. Census Bureau, ICF Macro, and Serpro, funded by MEASURE
DDI	Data Documentation Initiative
DHS	Demographic and Health Survey (a MEASURE project)
IE	Impact Evaluation
IHSN	International Household Survey Network (a World Bank-supported project)
IRB	Internal Review Board
LSF	Local Survey Firm; i.e., the in-country firm who conducts the fieldwork and data processing tasks for the baseline and/or endline surveys
MICS	Multiple Indicator Cluster Survey (a UNICEF-funded survey)
MIS	Management Information System
OECD	Organisation for Economic Cooperation & Development
PI	Principal Investigators
PUF	Public Use File
SOP	Standard Operating Procedures
ToR	Terms of Reference
TSSM	Total Sanitation and Sanitation Marketing
WB	World Bank
WHO	World Health Organization
WSP	Water and Sanitation Program

## Foreword

In addition to presenting the topics requested for the Inception Report, Kimetrica has also included within this report the following two items:

- 1) A revision of the payment schedule, shown and discussed in Chapter 8.
- 2) Suggested tables to use for quality control of the fieldwork operation, as well as progress of the data entry operation; all can be found in Appendix B. These tables should be run weekly by the Data Entry Supervisor, and should be distributed to the Survey Management, local PI, World Bank/Headquarters, and Kimetrica.

## 1 Introduction

On 29 April 2010, Kimetrica International Limited was awarded a contract with the World Bank (WB) to support the Data Entry Program, Data Management & Integration, and Field-Based Data Entry Training and Support for the Bank's impact evaluation surveys related to the Water and Sanitation Program (WSP)'s programming and the scale-up of the Total Sanitation and Sanitation Marketing (TSSM) project. The project has a total of seven deliverables, starting with this Inception Report.

This inception report is based on the Terms of Reference (ToR), and is organized around the project deliverables. We have included a Gant schedule for each deliverable, although this schedule will vary for some deliverables depending on the local survey contracting. This report forms the work plan for the project. We have included from the ToR an overview of the project objectives and scope of work to underscore what the project is about and the process for completing all deliverables.

## 2 Objectives

Kimetrica will design and manage a Gates Foundation-funded, World Bank-directed, data entry system for use in the final round of a multi-round household survey in four countries (India, Peru, Senegal, and Vietnam). Specifically, Kimetrica will perform the following tasks:

- Create a data entry system that will minimize data error, ensure data integrity, and protect respondent privacy with practical methods.
- Provide guidance and support to the country-level survey firms, both with the data reduction system and with personnel and logistical requirements, to ensure sufficient training and high quality implementation of the data entry system.
- Produce preliminary analysis files, as well as final, fully documented, privacy-protected (i.e., anonymized) analysis files.

## 3 Scope of Work

Kimetrica will develop global protocols and standards for the data reduction system that will minimize data error, ensure data integrity, and protect respondent privacy. These global protocols will contain the following features:

- Range-checking of numerical and alpha-based questions
- Consistency checks among questions during data entry
- In-depth consistency checks of the data files made post-entry, run in batch mode
- Skip pattern checks to ensure consistency with the data entered at the point of the skip request to the skip location requested
- Visually agreeable screens that mirror the questionnaire layout
- Summary reports (executed in the office on data keyed to that point) which flag quality control issues encountered during field collection
- Data organized by respondent, in ASCII format, allowing for easy exportation to other data formats
- Frequency and summary tabulations of each country's dataset
- Secure data storage and reporting
- Appropriate specimen, document, and data handling via the adherence to protocols
- Country-specific source code and supporting documentation
- Final, anonymized data files

Kimetrica will then tailor the global-level components to the country-level, which will encompass the following tasks:

- Advice on country-specific survey firm requirements
- Ensure compliance with data handling and security requirements
- Review of the endline questionnaire modules
- Provide country-specific data entry program and user manuals
- Provide system user manuals and training support as needed
- Support survey implementation and management
- Be available to identify and quickly resolve software bugs with Local Survey Firm (LSF)

## 4 Expected Outputs

The following outputs will be submitted to the World Bank during the life of the project:

- 4.1 Inception Report (this document): a detailed list of activities, setting out a work plan highlighting key milestones.
- 4.2 Data Handling Manual: a clear set of rules at both the country and global level for data handling, storage, and backup protocols.
- 4.3 Questionnaire Review: a review of each country's questionnaire will be made for comparability to the other country questionnaires, as well as for internal coherence to a clear and logical question flow.

- 4.4 Data Reduction/Quality Assurance System: the fully tested, country-specific data entry system.
- 4.5 In-Country Field Implementation Plan: an assessment of the in-country survey firms' field implementation plan to ensure adequate existence and procedures for key tasks in the data processing cycle.
- 4.6 Data Entry Training and Support: Kimetrica staff will provide in-country training and in-country and remote support to the LSF's Data Entry Staff, as well as field test the data entry system prior to the system's implementation.
- 4.7 Datasets and Reports: Kimetrica will provide weekly reports on the progress of data entry in four countries, integrated datasets from the baseline and endline data as possible from all six countries (India, Indonesia, Peru, Senegal, Tanzania, and Vietnam), documentation on the datasets, and recommendations on best practices on data storage.

## 5 Deliverables

All activities discussed in this Chapter are visually presented in a Gantt chart, which can be found in Appendix A. Deliverables (milestones) are indicated by **green** boxes. Please refer to the Appendix for more details.

### 5.1 Inception Report (Deliverable #1)

The Inception Report (this document) presents several key topics in depth, in order to provide a more detailed sequencing of project activities, timelines, and milestones at the global level. These topics are addressed below, within this chapter. Other requested documents (Work Procedures and Roles & Responsibilities) can be found below in Chapters 6 and 7 (respectively).

As part of the Inception Report, the original ToR requested that the baseline questionnaires, data collected during that phase and baseline field documents be reviewed as part of the Inception Report. Whereas as part of Deliverable 5, the ToR requested that the LSF's implementation plan be reviewed by Kimetrica, with no mention of interviewer manuals, field training materials, or the LSF's Standard Operating Procedures (SOP). Kimetrica believes it would be more advantageous to review these field documents for the endline survey, where Kimetrica can impact the quality of the data collected, rather than reviewing baseline documents for which the data has already been collected. Refer to Appendix A, Deliverable 5, for the integration of these activities therein.

To date, Kimetrica has received the baseline questionnaires for each country, but is still in need of the datasets and interviewer manuals (the latter for reference purposes) for each country. Kimetrica would prefer receipt of these materials before start up of Deliverable 4 (development of the data reduction system), but at the latest before start up of Deliverable 7c (integrated baseline, longitudinal, and endline datasets).

## 5.2 Data Handling Manual (Deliverable #2)

The Data Handling Manual is being developed in tandem with the Inception Report and will provide detailed step-by-step instructions and guidance on overall data management, training aspects of data handling, and field and office data handling at the global level. The manual will focus on a clear set of rules for data handling, storage, and backup protocols to protect against reasonable risks of data loss and corruption. It will be generic for all countries, with country-specific modifications being added as needed after finalization of the questionnaire and/or during assessment of the LSF's documents.

The procedures and guidance material we develop for this project will be based on bringing together guidelines and checklists developed by the International Household Survey Network (IHSN), a World Bank-supported project. Because our primary business is conducting surveys, we take survey quality control very seriously. Under this contract, we will provide procedures that will identify factors affecting the accuracy, validity, and reliability of survey data. We will provide tools on how to prevent and correct errors. We will provide examples and solutions on how to avoid common causes of misreporting. We utilize many tools and methods to provide the best approach to mitigating survey quality control issues, all of which will be addressed in Deliverable 2. A sampling of these can be found below:

Overall Confidentiality Issues: All persons engaged for the survey must take oaths of confidentiality and agree to follow general guidelines. For example, survey personnel should not discuss any details of their work with persons outside the survey team; this includes family members and friends. In addition, details of individual household interviews should only be discussed with other team members to resolve problems or concerns—they should not be the topics of idle conversation. Further, conversations amongst team members should not take place in public venues where non-survey team members could eavesdrop. These and other points of concern will be addressed in the Data Handling manual.

Field Handling of Questionnaires: On a day-to-day basis, field staff must be attentive to the security and confidentiality of the questionnaires. For example, questionnaires should never be left unattended in unlocked vehicles during the day, nor should they be left spread around in the field staff's lodging, where hotel workers could easily peruse the materials. These and other topics will be discussed.

Procedures for Returning Questionnaires to the Office: Detailed procedures on how questionnaires are to be prepared for data entry team will be given. This includes packaging and protection procedures, confidentiality of data, and transportation protocols, to avoid questionnaire loss.

Office Handling of Questionnaires: Similar to the field handling of questionnaires, protocols will be presented on how the data processing staff should handle the questionnaires and any biological specimens collected. In addition, persons not associated with the survey processing should not be given access to the data processing office(s); personal visits to the data processing team members should occur outside the office area.

Disposal of Questionnaires: The LSF should retain the questionnaires for a reasonable amount of time past the project's completion, six to 12 months is the norm (this should be dictated in the World Bank contract with the LSF). At the end of this six to 12 month period, they should be destroyed in a manner that retains the confidentiality of respondents. In addition, questionnaires will occasionally be returned from the field that will be deemed incomplete or otherwise un-keyable. These questionnaires should not be disposed of prematurely, but rather, must be retained with the other questionnaires in the cluster until the entire cluster's questionnaires are ready for disposal.

These and other issues will be discussed in full in the Data Handling Manual. Please refer to Appendix A for a timeline of Deliverable 2's activities.

### **5.3 Review of Country-Specific Questionnaires (Deliverable #3)**

To accelerate completion of this deliverable, the Principal Investigators (PI) in each country should have already begun to review and revise the baseline questionnaires as necessary to produce the endline questionnaires, as this task is not dependent on contract award to either Kimetrica or a local survey firm.

In order to produce a single, functional dataset as requested in the ToR for the six participating countries, the baseline and endline questionnaires in all six countries must adhere closely to the global questionnaire. However, during meetings held in June and July 2010 between World Bank and Kimetrica staff, it was acknowledged that several of the countries had deviated from the global baseline questionnaire, with at least one quite significantly (Vietnam), rendering a final common dataset questionable. It was also acknowledged that the local country PIs would be reluctant to adopt the global questionnaire for their endline survey (and unless the global baseline questionnaire had been originally used, it would still fail to provide for comparative data). Rather, the local PIs will most likely adopt their country-specific baseline questionnaire as their endline questionnaire, with little or no changes. The only anticipated changes would be due to correcting mistakes in the original questionnaire wording or responses to effect a more streamlined interviewing; to adjust for date/time changes; to introduce new questions necessary for the endline round of data; or to make or introduce other minor variations as needed to facilitate the baseline to endline questioning. Kimetrica anticipates these activities should not take more than a week on the part of the PI, as indicated in Appendix A's Deliverable 3.

In order to facilitate the WB & PI reviews of the questionnaire, Sherrell Goggin (Kimetrica) has already reviewed the global baseline questionnaire and made substantive suggestions to the WB staff. She reviewed the questionnaires both in terms of the overall structure, content, flow, and wording, and in terms of revisions necessary to facilitate the endline survey. Goggin then met mid-August 2010 with WB staff to review these suggestions, which will now be finalized and incorporated within the baseline questionnaire by WB staff. Once this is done, it will then be sent to the Berkeley reviewers, who will then make their final review before the revised baseline questionnaire can be sent to the in-country PIs. Once the



local PIs receive this revised baseline questionnaire, they should make review and finalization of the endline questionnaire a top priority, so as to not further delay start up of the individual surveys. While the interruption of end of the year holidays should not present a problem to fieldwork, these holidays should not be allowed to intervene between enumerator training and the fieldwork start-up.

Once the PI has revised the questionnaire as necessary, it then will pass back to the World Bank staff for review, after which Kimetrica can then make a final review of the English language version of the country-specific questionnaires, addressing the following points:

- Ensure comparability between the baseline and endline questionnaires
- Ensure that the introduction of a question(s) does not alter the universe of any existing questions, unless it is to correct for a mistake or omission in the baseline questionnaire
- Ensure that the removal of a question(s) does not alter the universe of any existing questions due to skip pattern changes, unless it is to correct for a mistake or omission in the baseline questionnaire
- Ensure the questionnaire flows in a clear, logical manner
- Ensure the questionnaire is free of typographical or other errors
- Draft a report detailing all problems and inconsistencies noted, with recommended solutions

For each country, Kimetrica will submit to the World Bank a report of the above, making recommendations to resolve any problems noted. From this the World Bank and local PI must produce the final country-specific questionnaire. Kimetrica can then proceed to document commonalities and differences between the baseline and endline questionnaire, in preparation for creating the finalized dataset.

At the conclusion of the entire WSP project, Kimetrica will attempt to create a master dataset using the common data fields across all surveys. Since Peru's questionnaire was closest to the global version, it is suggested that it be used as the starting point for data field comparisons. To facilitate the production of this master dataset, Kimetrica asks that the World Bank indicate, preferably in a tabular (Excel) format, if data fields are considered of primary or secondary importance to the survey. In this way, Kimetrica can compare each country-specific dataset to this list and determine what percentage of primary and secondary data fields were collected, and hence, whether or not an attempt should be made to include the country in the master dataset. For example, if there are 100 data fields identified as having primary importance to the survey, and only 30 fields were collected for a given country, analysis for that country at the global level might be so comprised as to render global-level analysis for this country as ineffective. Final determination for inclusion in the master dataset will be at the discretion of the World Bank.

Details on the timing of these events can be found in Appendix A's Deliverable 3 entry.

#### 5.4 A ready-to-use, fully documented Data Reduction System (Deliverable #4)

Once Deliverable 3 is finalized, Kimetrica will commence work on the Data Reduction System. Kimetrica will follow documented best practices in order to minimize data error, ensure data integrity, and protect respondent privacy. The system will facilitate timely, transparent, and secure access to documented analysis files. We understand from the ToR that the data entry program must have these characteristics:

- Allow for the update and reconciliation of key information between the baseline and endline surveys
- Organize data by respondent in ASCII format
- Provide range, skip-pattern, and consistency checks
- Alert data entry operators to out-of-range values
- Present visually agreeable screens that mirror the questionnaire layout

In addition, the overall Data Reduction System must provide for the following features:

- A data format consistent with WSP requirements
- Ensure that panel data can be retrieved from the baseline study and modified as necessary
- Weekly error summary reports detailing errors occurring during field collection
- Consistency reports run on a batch (cluster) level, detailing errors between data fields
- Basic tables containing the sample size, frequencies, and, for quantitative variables, information on the means, minimum, and maximum values

Finally, as part of this specific deliverable, Kimetrica will provide the following items:

- Documentation and test results exhibiting the system's compliance with the ToR
- Full source code, and any global- or country-level documentation developed
- As needed, a lab-based data entry system to support the registration of specimen processing
- Documentation to support the above

Given that data entry from the baseline surveys were entered using CSPro, Kimetrica will continue that trend and use CSPro for the endline round.

The ToR states as part of Deliverable 4 that the Data Entry System should be tested in each country. However, this can only occur if there are "live" questionnaires to test the system against, and this can only occur once Deliverable 5 is complete and Deliverable 6 is underway, i.e., that the enumerators have either piloted the questionnaire or have begun actual fieldwork. Furthermore, if there were problems beyond Kimetrica's control, such as a delay in the fieldwork startup, this would unjustly penalize Kimetrica's completion of this deliverable. Therefore, Kimetrica will test each country-specific data entry system in their

offices prior to arrival in the local country through the use of sample questionnaires as completed by Kimetrica staff.

As part of Deliverable 6b, once in country and prior to Data Entry Staff training, Kimetrica will test the system using fielded questionnaires. In addition, training of the Data Entry Staff will also act as a test of the data entry system, for the data entered during this period is discarded before actual startup of the data entry cycle begins. This is the identical method used by the data processing staff of the DHS.

See Deliverables 4 and 6 in Appendix A for more information on timing of these specific events.

## 5.5 Assess LSF' Field Implementation Plan (Deliverable #5)

Deliverable 5 start-up will hinge upon the completion of Deliverables 2 and 3. Once the LSF has developed their field implementation plan, Kimetrica will review their materials with the following targets in mind:

- To ensure an adequate recruitment and training methodology for field and office staff
- To ensure adequate interview procedures and guidelines to field staff during the enumeration
- To ensure that best practices for data management are in place
- To ensure the adequate existence and/or recommend procedures for labeling of specimens and lab-based data entry requirements when needed (in applicable countries)
- To ensure adequate existence and/or recommend procedures for primary data handling while survey teams are in the field to secure against unauthorized access or data loss
- To ensure adequate existence and/or recommend formats and procedures for field data reporting requirements, data transmission schedules, and recommended data backups

Refer to Appendix A, Deliverable 5, for an estimated timeframe for the LSF and Kimetrica's activities.

## 5.6 Data Entry System Testing, Training, and Support (Deliverable #6)

Deliverable 6 marks the start-up of fieldwork and data processing activities, and should commence once Deliverable 5 is complete. Kimetrica's role will include testing, installation, and setup of the data entry system, as well as training of the Data Entry Staff.

### 5.6.1 Data Entry System Testing

Deliverable 6b, per the ToR, stipulated "field DE (data entry) test reports". It is Kimetrica's understanding that this referenced testing of a field-based data entry system, as the timing

on this activity was one month prior to the startup of actual field enumeration. Therefore, since field-based collection will not take place for any of the four countries Kimetrica is supporting, this deliverable is not directly supportable.

However, Kimetrica will make every effort to test the office-based system prior to launch. Further, once Kimetrica staff arrive in country for the data processing training, the initial 2-3 day training period will also serve as a rigorous test of the system. Corrections will be made during this period before "live" entry begins. The remainder of the two week trip will allow for further opportunities to make any other corrections should problems arise. By the end of the two week period, the system should be considered perfect or nearly so; should any further modifications be necessary, these can be handled via email or using Kimetrica's FTP site. This is the same mechanism used by the DHS, and has proven quite serviceable.

### 5.6.2 Data Entry Staff Training

Our training approach will include direct classroom training, interactive workshop discussion, hands-on and practical experience training, survey enumeration and data management problem identification, and mitigation measures. As we stress the importance of continuous training, we intend to provide data entry managers with regular training updates on techniques to reinforce training topics. Finally, once Kimetrica has completed the in-country training, on-going support will be given to the LSF throughout the duration of the project in the form of email, Skype, and telephone contact. Based on the type of issue or question raised, the appropriate Kimetrica staff member will contact the LSF to resolve the problem. Given the geographic spread of our team, we will be able to respond to queries raised in all four survey countries immediately, and to the Washington, DC based WSP team from 0900 to 1800 local time.

Shortly after field staff training is completed by the LSF and fieldwork has begun, training of the Data Entry Staff will begin in earnest by Kimetrica staff. We anticipate a Kimetrica staff member to be in-country approximately two weeks to accomplish software installation, training, and initial supervision of the data processing activities.

The CSPro data entry software will include all the necessary validation rules, errors, and consistency checks as stipulated in Deliverable 3. We will provide guidance to the LSF on the best approach to data entry, based on local country conditions and project sample size. Based on our findings, we will recommend specific procedures for data handling and entry.

Before deploying the data entry software and training materials to the LSF, our team will spend at least eight hours entering sample questionnaire data for each country at the Kimetrica offices. This will identify any significant problem areas in the data entry forms so they can be corrected before deployment. It will also allow us to provide the LSF with an optimum time estimate to enter one questionnaire, important for their scheduling and budgeting purposes.

The materials for in-country data entry training will include the following modules:

- 1) *An introduction to the survey.* Familiarization with the project and questionnaires.
- 2) *Defining roles of the Data Entry Supervisors and Data Entry Staff.* Explaining the roles and responsibilities to both groups.
- 3) *Understanding the survey and survey questions.* Question-by-question explanation and discussion.
- 4) *Understanding the Data Entry System and Controls.* The data entry screens will resemble, as much as possible, the questionnaires. During this training module, Data Entry Staff will be familiarized with the data entry software and controls.
- 5) *Practice Data Entry Sessions.* Each data entry trainee will be provided with a computer workstation and actual questionnaires from the field. Note that this training session is an important method to determine which Data Entry Staff should be retained and which should be dismissed.

The duration of practice session will be dependent on the following factors:

- 1) *Quality of data collected from the field.* Questionnaires that are incomplete, that made incorrect selections of eligible children and/or caregivers, that have missing data, or have improperly executed skips, will all contribute to delays in data entry and data processing.
- 2) *The ease with which trainees learn the system.* Quality of staff varies in each country; some companies maintain a permanent staff of experienced data entry operators; others must hire operators on a temporary basis. While the latter group will normally take longer than the former to learn the process, a permanent staff of data entry operators who do not follow best practices for data handling can take just as long or longer, for they must be "untaught" undesirable habits.
- 3) *The speed with which trainees can enter a cluster.* Clusters in most countries should average 15 households (India is the exception at 25). It is suggested that the trainee complete one cluster as part of their training.
- 4) *Site Conditions.* Computers should run on a local network. If there are network problems forcing the manual copying of files, or power fluctuations/loss preventing access to the system, these will also delay the training session.

Data Entry Staff training should run approximately three days.

### 5.6.3 Data Entry Supervisor Training

In addition to participating in all Data Entry Staff training, supervisor(s) will be given additional training. Depending on the Data Entry Staff size, whether one shift or two is maintained, and other local conditions, Kimetrica may suggest that the services of two or more supervisors be engaged. In addition to training these supervisors, at least one additional, permanent staff member of the LSF should be trained to cover in the event of

attrition or temporary absence by the primary supervisor(s). Topics for the in-country supervisor training will include the following modules:

- 1) *Supervision and Monitoring*. A review of roles and responsibilities, explaining the communication expectations among the data entry, supervisor, and survey manager staff. Data entry quality control will be stressed.
- 2) *Review of Data Entry Process*. Overview of possible problems, causes, solutions, with example problems and how to resolve them.
- 3) *Identifying Problem Questionnaires*. If questionnaires contain so many problems as to render them unkeyable, an entire household may be omitted from entry (and hence, analysis). Guidance will be given on what constitutes this level of severity.
- 4) *Error Reporting*. Supervisors will be trained on how to produce quality control reports for each Data Entry Staff. The reports, which should be run on a bi-weekly or weekly basis, will help detect data entry problems early so that corrective actions can be taken.
- 5) *Quality Control*. The Data Handling Manual will provide guidelines related to quality control.

As with enumeration training, the actual amount of time required for supervisor training will vary; however, two days should be sufficient.

#### 5.6.4 Post-Training

After training for all data processing staff members is complete (anticipated to last 5-6 days), Kimetrica staff will remain in-country for an additional week to ensure the smooth operation of the staff, software, and procedures.

Timing of these activities can be found in Deliverable 6 of Appendix A.

### 5.7 **Weekly Summary Reports (Deliverable #7a)**

Deliverable 7a, weekly summary reporting, will commence once data entry training has ended and "live" entry has begun.

Database management capabilities will be programmed using CSPro. A significant aspect of the project's MIS is that it will be able to provide quality control reports and country-level progress reports to the LSF, Kimetrica's managers, and designated World Bank staff in near real-time. The Kimetrica team will work with the World Bank to ensure that reports are designed to capture progress against all key performance indicators.

This will allow for any data quality or data entry management issues not captured at the country-level to be flagged and addressed very quickly. It will also allow managers to review countries' data entry progress.

Suggested table templates to be used for the fieldwork's data quality process can be found in Appendix B. These tables can be revised, expanded, or modified to suit the needs of the World Bank or local country. It is suggested that the Data Entry Supervisor execute these tables on a weekly basis. During the in-country training visit to the LSF (Deliverable 6a), Kimetrica staff will spend time to train the Data Entry Supervisor in the usage of this system. See Appendix A for the scheduling of Deliverable 7a.

## **5.8 Assessment Report on Data Quality Cleaning and Basic Descriptive Statistics (Deliverable #7b)**

As part of the Data Handling Manual, Kimetrica will outline for the LSF various steps that should be implemented to ensure data quality in both the field and the office procedures, and to reduce data loss. As part of Deliverable 7b, Kimetrica will provide an assessment of the LSF's data quality cleaning for the endline dataset. In order to do this, we will cover the following areas.

### **5.8.1 Field Editing**

During Kimetrica staff's in-country visit, questionnaires will be monitored during the data entry operation to determine how much field editing is taking place, what type of editing is taking place, and whether the corrections being made are being done properly. Too much editing signals poor field enumeration practices, and should be corrected with the enumeration activities rather than corrected by field editing. Likewise, manual editing of data, which cannot possibly be known to the field staff (such as dates of birth) must be halted the moment it is discovered. Kimetrica staff will work with the Data Processing Supervisor(s) to look for and report these types of field problems.

### **5.8.2 Manual Office Editing**

During Kimetrica staff's in-country visit, a discussion of office editing will take place with the Data Processing Supervisor. Essentially, no office editing should be allowed, unless the mistake or omission of data is such that the system will not allow entry to continue. In these instances editing of the data must be allowed. For example, if in one of the child rosters the interviewer failed to list all eligible children, the system will not allow entry to continue until all eligible children are given. In this situation the Data Entry Supervisor must write the missing child's identification code in the roster, and the operator must then enter a row of blanks for the data. Kimetrica can then look for these types of entries in the data (i.e., a row with all missing/blank values) and assume that the interviewer failed to list the person (or listed the person but failed to collect any data for them, an equally egregious situation).

### **5.8.3 Automated Office Editing**

Kimetrica staff will review all automated editing programs written by the LSF to ascertain whether improper editing rules were applied. Automated editing by the LSF should only be done at the request of the World Bank and local PI, and should be made in strict accordance with set guidelines and editing rules.



#### 5.8.4 Cross Tabulations

Kimetrica will conduct a series of cross-tabulations on key data fields to look for problem areas. For example, a cross tabulation of sex with age in the household should show a proportionate number of responses in each age category. Likewise marital status crossed by age should not produce questionable data, such as widowed 5 year olds.

#### 5.8.5 Automated Quality Checks

Kimetrica will develop a series of automated quality control checks to probe for inconsistencies, discrepancies, omissions, or otherwise problem data. For example, in the household roster, Kimetrica will check the person's date of birth against their age to ensure consistency between the responses. We will also make basic checks of the household structure to ensure, for example, that there is one and only one head of the household, that ages of parents and children appear reasonable, and that educational levels attained are reasonable considering the person's age. As part of this process, if there are any key fields (i.e., data contributing to indicators) that the World Bank wants particular attention paid to, then the Bank should provide a listing of these data fields and the check to be made.

The second half of this deliverable will be the compilation of basic descriptive statistics for each country-specific unedited data set. (If the data are edited, this will also be produced for the country-specific edited data set.) Using these statistics, it will be possible to determine the frequency of missing data, out-of-range or otherwise invalid data, or other conditions warranting attention.

### 5.9 Integrated Baseline and Endline Datasets (Deliverable #7c)

After assessing the quality of the data and resolving any anomalies in the data with the LSF, Kimetrica will create a single integrated dataset of the baseline, longitudinal (if applicable), and endline datasets using CPro. CPro utilizes ASCII data files, but can produce STATA, SAS, or SPSS files as well. CPro is being used in all data processing activities within this project, and was chosen for its widespread usage in statistical agencies worldwide, as well as its ease of use.

All six countries (India, Indonesia, Peru, Senegal, Tanzania, and Vietnam) will be included in this deliverable, and each will have their own dataset. The integrated dataset will include data pairs (or triplets) of common data found in both the baseline and endline (and longitudinal, if applicable) datasets. It will also contain solo data values for those data fields that cannot be matched in any of the other survey periods.

At this time, Kimetrica is unaware of whether or not the baseline or longitudinal data were edited in each country, nor whether those edits improved the overall quality of the data. Therefore, if there are two datasets for a given country, the World Bank or local PI must determine which data to use—though regardless of which data are used, it is strongly suggested that they all be in the same state, i.e., either edited or unedited. In addition, it is Kimetrica's understanding that weights may be used for the baseline reports that are



currently underway. If true, then these weights must be included in the final integrated dataset, in order for analysts and users of the data to be able to reproduce these results.

This deliverable will begin after completion of Deliverable #7b, except in the case of Indonesia and Tanzania, which can begin whenever the World Bank provides the data sets to Kimetrica.

See Appendix A for the timing of this deliverable.

## 5.10 Documented Public Use Files (PUF) (Deliverable #7d)

The World Bank and the IHSN have developed tools for data documentation and data archiving. This suite of tools is collectively known as the Microdata Management Toolkit. Kimetrica is contracted by OECD to provide technical support and training to countries in Africa and Asia in the adoption of the World Bank / IHSN data archiving standards.

The Toolkit comprises three modules. The Metadata Editor is used to document data in accordance with international metadata standards (DDI and Dublin Core). The Explorer is a free reader for files generated by the Metadata Editor. It allows users to view the metadata and to export the data into various common formats (STATA, SPSS, etc). The Metadata Editor and Explorer are based on Nesstar technology and developed by the Norwegian Social Science Data Services (NSD). The CD-ROM Builder is used to generate user-friendly outputs (CD-ROM, website) for dissemination and archiving.

For this project we feel the best documentation and data dissemination strategy is to use the data archiving tools developed by OECD and the World Bank. The documentation process includes standard templates provided by IHSN to assist in documentation. They include over 50 elements related to the survey, including factors such as sampling technique, documentation of non-response, etc. The tools also capture questionnaires and stand alone manuals. This includes all the syntax, programs, manuals, and design documents.

However, we may explore including the World Health Organization's (WHO) statistical data and metadata exchange (SDMX) standard for statistical information and indicator computation. Because it is health sector specific, using this data archiving standard could also promote further harmonization and integration into the international effort to manage various health sector indicators. The World Bank and Kimetrica will make final selection of an archiving protocol via mutual agreement. Please refer to Appendix A for the timing of this Deliverable, 7d.

## 5.11 Recommendations on Platform (Deliverable #7e)

For our final Deliverable, #7e, we will provide recommendations for a data platform to use for the storage and sharing of datasets, while ensuring compliance with best practices and safeguards in compliance with Deliverable 2.

In addition, Kimetrica will attempt to create an integrated dataset of common data from all six countries; however, there will likely be too few data fields in common across the countries and across all surveys in each country to make this a viable end product.

Deliverable 7e is global in scope and will begin once the project is underway. It will be submitted after Deliverable 7d has been completed for all six project countries.

## 6 Work Procedures (Deliverable #1, continued)

Kimetrica has developed a draft work plan that has been referenced throughout this document, which can be found in Appendix A in the form of a tabular timeline. Therein, we have attempted to make our best estimates on time projections for specific tasks that we are obligated to under the terms of our contract with the World Bank. In addition we have included key events performed by other entities that will impact our delivery schedule. Please note that there is no "downtime" between events—the timeline represents an idealized project flow. If there are any delays between events, this will most likely result in delays to the overall project. Please refer to Appendix A during the discussion of the work procedures given below.

Beginning with Deliverable #1, the Inception Report (this document), we have listed several tasks within this deliverable that are addressed within this document. Timing of this deliverable is dependent only upon contract award of the project with the World Bank, which has now commenced. There was some deviation from the topics listed in the ToR, which has been addressed above in Chapter 5.3. We hope this meets with the World Bank's approval. The remaining task under this deliverable, Roles & Responsibilities, has been addressed below in Chapter 7.

Deliverable #2 corresponds to the Data Handling Manual. This document is being drafted in tandem with Deliverable #1, and should be ready at the same time or shortly thereafter Deliverable #1 is complete.

Deliverable #3 consists of review of the country-specific questionnaire modules. As discussed above in Chapter 5.3, this should have begun already by World Bank staff. Once the internal review processes by World Bank, Berkeley contributors, and in-country PI are over, Kimetrica can begin their comprehensive review of the near-final local questionnaire. At the time of this writing, no contracts have been awarded to any local survey firms. However, should an award be made, the LSF cannot begin any work until Deliverable #2 is complete, at which time they can review the Data Handling Manual to familiarize themselves with its content. Only once Deliverable #3 is complete can they begin the more weighty tasks of developing their managerial and field materials.

Deliverable #4 entails the development of a country-specific Data Reduction System by Kimetrica staff. This will commence once Deliverable #3 is complete for the country. For this deliverable, Kimetrica will also develop system documentation and/or other materials to help guide the end-users.

The project gets in full swing for the LSF with the onset of Deliverable #5, a review by Kimetrica of the LSF's field implementation plan. In order for Kimetrica to begin this task, the LSF must accomplish several tasks, among them reviewing the Data Handling Manual and familiarizing themselves thoroughly with the questionnaire modules. Only at that time can they properly develop a field implementation plan, field training manual, field reference guide, and standard operating procedures. This effort can occur in tandem with Deliverable #4, as it does not depend on the software system being complete. However, Deliverable #6, onset of the actual survey, cannot commence until the LSF has completed the documents listed above, nor until Kimetrica has signed off on the LSF's materials as being acceptable.

Deliverable #6 marks the start-up of field activities. It cannot begin until Deliverable #5 is complete. Numerous tasks for both Kimetrica and the LSF have been listed in order to show the event dependencies. As mentioned above, the timeline shows an idealized schedule, with no room for delays—slippage between any of the events will cause delays to the overall survey timeline. Further, delays of any individual event will cause a delay to the overall project. For example, a minimal amount of time has been allocated for the LSF to print all materials. If more time is taken, then fieldwork and data processing will be delayed.

Deliverable #7a will occur in tandem with Deliverable #6a, as this task involves providing weekly summary reports to the World Bank and local PI on data quality in the field, as well as progress of the data processing operation.

Deliverable #7b will occur after completion of Deliverable #7a, as this task involves making an assessment of the data quality cleaning and compiling basic descriptive statistics on the country-specific unedited data. If the data were edited, this will also be produced for the country-specific edited data.

After assessing the quality of the data and resolving any anomalies in the data with the LSF, Kimetrica will create a single integrated dataset of the baseline, longitudinal (if applicable), and endline datasets for Deliverable #7c. All six countries (India, Indonesia, Peru, Senegal, Tanzania, and Vietnam) will be included in this deliverable, and each will have their own dataset. At this time, Kimetrica is unaware of whether or not the baseline or longitudinal data were edited in each country, nor whether those edits improved the overall quality of the data. Therefore, if there are two datasets for a given country, the World Bank or local PI must determine which data to use—though regardless of which data are used, it is strongly suggested that they all be in the same state, i.e., either edited or unedited. This deliverable will begin after completion of Deliverable #7b, except in the case of Indonesia and Tanzania, which can begin whenever the World Bank provides the data sets to Kimetrica, but before the end of the contract term.

For the penultimate deliverable, Deliverable #7d, Kimetrica will produce public use files for each of the six country-level datasets. It will include annotated questionnaires and codebooks, and will be done for each country once Deliverable #7c for that country is complete.

The majority of the above deliverables were local in scope, i.e., they will be executed in turn for each of the four countries participating in this project (or six in the case of Deliverables

7c and 7d). Deliverable #7e, however, marks a return to a global-level deliverable, and as such, can only be initiated once all deliverables for all countries have been completed. The deliverable includes an outline of recommendations for storage platforms. Kimetrica will also attempt to create an integrated dataset of common data from all six countries for this deliverable; however, there will likely be too few data fields in common across the countries and across all surveys in each country to make this a viable end product.

## 7 Roles & Responsibilities (Deliverable #1, continued)

The WSP project is a Gates Foundation-funded project directed by the World Bank in Washington, DC. Baseline and longitudinal surveys have already taken place in the six countries participating in this project. For the final portion of this survey (to which this document pertains), an endline survey will occur in the six countries, at which point consolidation activities can take place. As part of the endline survey, the World Bank will contract local survey firms (LSF) in each of the six countries to conduct the fieldwork and data processing. Kimetrica's services have been engaged for these six countries to:

- Prepare an integrated country-level dataset; and
- Prepare public use files, which will include annotated questionnaires and codebooks.

In four of these countries, Kimetrica's services have been engaged to:

- Review the proposed endline questionnaires;
- Develop the data reduction system;
- Provide general oversight and guidance to the LSF's data processing activities;
- Provide an assessment report on data quality cleaning and descriptive statistics;
- Provide recommendations on storage plans; and
- Create an integrated global dataset, if feasible.

In addition to these LSFs and Kimetrica, the World Bank is also working with Principal Investigators (PI) in each country who will help direct the survey locally, colleagues from the University of California/Berkeley who are helping with the questionnaire design, and others, not all of whom are known to Kimetrica. This section will clarify the duties and roles of these disparate entities.

### 7.1 World Bank/Washington, DC and Kimetrica

Kimetrica will report directly to the Senior Impact Evaluation Specialist (SIES) in the World Bank's Washington, DC offices. The SIES will provide guidance on the sequencing of all country activities, monitor the quality control of Kimetrica's deliverables, and provide overall project oversight. Should Kimetrica encounter any discrepancies among the local PI, LSF, or other interested parties activities, the SIES, having primary authority over the project, will serve as arbitrator.

Kimetrica will not engage in communication with any interested parties other than the local PI and LSF, unless explicitly requested by the SIES.

## **7.2 World Bank/Washington, DC and In-Country Principal Investigator (PI)**

The WSP SIES will provide overall quality control and support to the country PIs, who in turn will work as the SIES's local representative, responsible for management of the LSF and monitoring the quality and timely delivery of their outputs. The WSP SIES will hire the LSF, however, these firms will report directly to the local PI.

## **7.3 In-Country PI and Kimetrica**

When in-country and as otherwise needed, Kimetrica will work closely with the in-country Principal Investigator to coordinate country activities. When in-country, Kimetrica will work primarily on the LSF premises to conduct training and overall data processing oversight. When working onsite or remotely, should any problems occur with the LSF (such as failure of the LSF to submit weekly file updates), Kimetrica will work closely with the PI to resolve the problem.

Kimetrica will include the PI on all correspondence with the LSF, and will update the PI on any telephone or Skype communications. Once the data processing operation has commenced, Kimetrica will provide weekly progress reports to the LSF, local PI, and WSP SIES.

Any information or requests made by the local PI that appear conflicting in nature with the SIES's direction or the overall project deliverables will be submitted to the SIES for resolution.

## **7.4 World Bank/Washington, DC and Local Survey Firm**

As mentioned in Chapter 7.2 above, the WSP SIES will be responsible for contracting the LSFs, who will then report locally to the in-country PI. The local PI will be in charge of ensuring overall project timetables with the LSF, and their compliance with the terms of their contract. The LSF will be responsible for the hiring of all field and data processing staff, the printing of questionnaires and other required materials, data collection activities, implementation of the data reduction system provided by Kimetrica, data cleaning, and monitoring the quality of the data being collected.

## **7.5 Kimetrica and Local Survey Firm**

Kimetrica will work closely with the LSF to coordinate their data processing activities, provide guidance and advice on the conduct of the data processing activities, and provide the data reduction system and training for it. Kimetrica will be responsible for ensuring that data entry system functions properly, and will respond quickly to fix problems with the software to ensure timely collection of data.

Although there will not be a contractual relationship between Kimetrica and the LSF firm, to ensure a close and effective working relationship between Kimetrica and the LSF, Kimetrica recommends that the WSP include language in the LSF's contract that require them to collaborate with Kimetrica in the areas of data management and training. If contracts are performance based, one of the measurements of performance should be a successful working relationship. Kimetrica requests that any contracts between the WSP and LSFs be provided to Kimetrica, so that Kimetrica is aware of all contractual obligations, especially as it overlaps with and pertains to the data processing operation.

In our experience, the best method to ensure successful working relationships is to understand and accept each party's ToR, and ensure that all communication is based on mutual respect.

## 8 Revision of Schedule of Payments

Following the issuance of the data reduction project contract and meetings with World Bank staff, Kimetrica requests that the payment schedule be amended to better reflect the timing of costs associated with the implementation of the projects. The original payment schedule, as presented in the World Bank's Terms of Reference, is shown below:

Payment (# and proportion of total payment)	Deliverable	Breakdown of Total Payment
Payment 1: 10%	Signing of Contract	10%
Payment 2: 15%	Deliverable 1	1.5%
	Deliverable 2	1.5%
	Deliverable 3	5%
	Deliverable 4	7%
Payment 3: 15%	Deliverable 5 for all countries	5%
	Deliverable 6a for all countries	5%
	Deliverable 6b for all countries	5%
Payment 4: 20%	Deliverable 7a	10%
	Deliverable 7b	10%
Payment 5: 40%	Deliverable 7c	35%
	Deliverable 7d	5%

After discussions with the World Bank staff and gaining a better understanding of the timing of country surveys, among issues, the following quandaries exist:

- 1) Payment 2 contains deliverables that are both global and country in scope. Specifically, Deliverables 3 and 4 (questionnaire review and development of the data reduction system respectively) are highly country-dependent, and if these tasks must be completed in all countries before Payment #2 can be made, this payment will not likely be processed for several months, and most likely not processed until early 2011.
- 2) Deliverables 5, 6a, and 6b (assess the local survey firm's (LSF) implementation plan; provide training to the LSF at the start-up of data entry; and field testing of the data

entry system; respectively), clearly state that work must be completed "for all countries" before Payment #3 can be made. However, we again face the situation where the time-lag between the start-up of Deliverable 5 in the first country, and the completion of Deliverable 6b in the last country, could be several months or more, causing the specter of another payment slipping into 2011.

- 3) Deliverables 7a and 7b (dissemination of weekly reports to WSP staff, and assessment of data quality cleaning, respectively) are both country-specific activities, and as such, dictate that Payment #4 cannot be made until the work is finished in all countries.
- 4) Per Chapter 5.3 above, it could be argued that Deliverable 7c is a country-specific activity as well, further delaying the final Payment (#5) until the work is finished in all countries.
- 5) In the original ToR there were five tasks under Deliverable #7, tasks a through e, but in the chart above, tasks c & d got combined into a single task, c, while task e got renamed as task d. We have separated these out into their original tasks, in part because task 7c, delivery of the integrated datasets, will now be done on a country level and is therefore local in scope.
- 6) It will place an extreme financial burden on Kimetrica to front the lion's share of the project costs that will be incurred in the next few months for a potentially lengthy period.

Given these very real concerns, we suggest that the contract payment terms be amended according to this revised schedule

<b>Payment (# and proportion of total payment)</b>	<b>Deliverable #</b>	<b>Breakdown of Total Payment</b>
<b>Payment 1: 10%</b>	Signing of Contract	10%
<b>Payment 2: 30%</b>	Deliverable 1	15%
	Deliverable 2	15%
<b>Payment 3: 10%</b> (Country #1)	Deliverable 3	1%
<b>Payment 4: 10%</b> (Country #2)	Deliverable 4	2%
<b>Payment 5: 10%</b> (Country #3)	Deliverable 5	1%
<b>Payment 6: 10%</b> (Country #4)	Deliverable 6a	1%
	Deliverable 6b	1%
<b>Total Outlay: 40%</b>	Deliverable 7a	1%
	Deliverable 7b	1%
	Deliverable 7c	2%
<b>Payment 7: 20%</b>	Deliverable 7d	10%
	Deliverable 7e	10%

We have striven for a balance between allowing Kimetrica to receive payments within a reasonable timeframe that can cover our project operating costs, and not causing the World Bank to incur numerous small payments. We hope the above is an equitable solution.



## Appendix A: Project Timeline

			Month 1 (week)				Month 2 (week)				Month 3 (week)				Month 4 (week)				Month 5 (week)				Month 6 (week)				Month 7 (week)				Month 8 (week)				Month 9 (week)			
	Tasks	Who	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Deliverable 1 activities are contingent upon contract award to Kimetrica. Tasks are global in scope.																																						
1.1	Review all baseline questionnaires	KM																																				
1.2	Define project activities for each component	KM																																				
1.3	Develop detailed workplan	KM																																				
1.4	Define work procedures and roles and responsibilities	KM																																				
1.5	Internal review and finalization of Inception Report	KM																																				
1	Deliverable 1 submission (Inception Report)																																					

Who's Who:

KM: Kimetrica  
WB: World Bank, Headquarters  
PI: Principal Investigator in local country  
LSF: Local Survey Firm

Who's Who:

KM: Kimetrica

WB: World Bank, Headquarters

PI: Principal Investigator in local country

LSF: Local Survey Firm

			Month 1 (week)				Month 2 (week)				Month 3 (week)				Month 4 (week)				Month 5 (week)				Month 6 (week)				Month 7 (week)				Month 8 (week)				Month 9 (week)				
	Tasks		Who	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Deliverable 2 activities are contingent upon contract award to Kimetrica. Tasks are global in scope, and will be performed in tandem with Deliverable 1.			KM																																				
2.1	Draft Data Handling Procedures (Data Management)																																						
2	Deliverable 2 submission (Data Handling Manual)																																						



			Month 1 (week)				Month 2 (week)				Month 3 (week)				Month 4 (week)				Month 5 (week)				Month 6 (week)				Month 7 (week)				Month 8 (week)				Month 9 (week)			
	Tasks	Who	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
	Deliverable 3 activities are <b>local</b> in scope. Tasks 3.1 and 3.2 are not contingent upon contract award to Kimetrica or the LSF, and should be initiated as soon as possible. Ideally, Deliverable 3 will be completed at the time of contract award to the LSF.																																					
3.1	Review of baseline questionnaires to revise for endline	PI																																				
3.2	Review of revised endline questionnaires	WB																																				
3.3	Review of revised endline questionnaires. Assess comparability with country's baseline questionnaires. Ensure clear wording and logical flow of all questions. Submit report detailing problems and recommended resolutions.	KM																																				
3.4	Reconciliation of KM comments	PI+WB																																				
3.5	Document commonalities/differences between baseline and endline QREs																																					
3	Deliverable 3 submission (Questionnaire Review)	KM																																				

			Month 1 (week)				Month 2 (week)				Month 3 (week)				Month 4 (week)				Month 5 (week)				Month 6 (week)				Month 7 (week)				Month 8 (week)				Month 9 (week)			
	Tasks	Who	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
	Deliverable 5 activities are <b>local</b> in scope. Activities are contingent upon contract award to a LSF and completion of Deliverables 2 & 3.																																					
5.1	Review Deliverable 2 materials	LSF																																				
5.2	Develop Fieldwork Standard Operating Procedures (SOPs)	LSF																																				
5.3	Develop implementation plan	LSF																																				
5.4	Develop Interviewer (reference) guides (contingent on Deliv #3)	LSF																																				
5.5	Review of local survey firm's recruitment, training, SOP, and interviewer (reference) guides and their implementation plan.	KM																																				
5.6	Reconciliation of Kimetrica comments	LSF																																				
5.7	Final review of LSF materials	KM																																				
5	Deliverable 5 submission (LSF Field Plan Assessment)																																					

			Month 1 (week)				Month 2 (week)				Month 3 (week)				Month 4 (week)				Month 5 (week)				Month 6 (week)				Month 7 (week)				Month 8 (week)				Month 9 (week)			
Tasks			1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
<b>Deliverable 6a activities are local in scope, and are contingent upon completion of Deliverable 5.</b>																																						
6a.1	Recruitment of local field staff	LSF																																				
6a.2	Piloting of questionnaire--occurs after task 6a.1	LSF																																				
6a.3	Finalization of questionnaires, field training manuals, and field reference manuals based on Pilot--occurs after task 6a.2	LSF																																				
6a.4	Printing of questionnaires, field training manuals, and field reference manuals--occurs after task 6a.3	LSF																																				
6a.5	Training of local field staff--occurs after task 6a.4	LSF																																				
6a.6	Field collection (~2 months)--occurs after task 6a.5	LSF																																				
6a.7	In-country setup, training, and support to LSF's data entry staff--this can begin one+ week after task 6a.6 commences	KM																																				
6a.8	"Live" data entry--this will commence ~one week after task 6a.7 commences	LSF																																				
6a.9	Remote Support to LSF--occurs after task 6a.7	KM																																				
6a	Deliverable 6a submission (Report on local training)--will submit after completion of task 6a.9																																					
<b>Deliverable 6b activities are local in scope, and are contingent upon completion of Deliverable 6a.3.</b>																																						
6b.1	Test DE system with Pilot QREs--occurs after task 6a.2	KM																																				
6b	Deliverable 6b submission (Data Entry test reports)																																					

			Month 1 (week)				Month 2 (week)				Month 3 (week)				Month 4 (week)				Month 5 (week)				Month 6 (week)				Month 7 (week)				Month 8 (week)				Month 9 (week)			
	Tasks	Who	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
Deliverable 7a activities are <b>local</b> in scope, and will occur in tandem with task 6a.8.																																						
7a.1	Weekly data entry summary reports	KM																																				
7a	Deliverable 7a submission (Weekly data entry reports)																																					
Deliverable 7b activities are <b>local</b> in scope, and are contingent upon completion of Deliverable 6a.																																						
7b.1	Assessment Report on data quality cleaning and descriptive statistics	KM																																				
7b	Deliverable 7b submission (Data assessment Report)																																					
Deliverable 7c activities are <b>local</b> in scope and are contingent upon completion of Deliverable 7b.																																						
7c.1	Country-level integrated dataset for the baseline, longitudinal, and endline data files	KM																																				
7c	Deliverable 7c submission (Final integrated country-specific, country-level dataset)																																					
Deliverable 7d activities are <b>local</b> in scope, and will be performed upon completion of Deliverable 7c.																																						
7d.1	Public use files, including annotated questionnaires and codebooks	KM																																				
7d	Deliverable 7d submission (PUF)																																					
Deliverable 7e activities are <b>global</b> in scope, and will be performed at project close-out																																						
7e.1	Recommendations on Storage Plan	KM																																				
7e.2	Creation of a global-level integrated dataset if feasible	KM																																				
7e	Deliverable 7e submission (Recommendations)																																					

## Appendix B: Field Check Tables

This section includes suggested data quality tables to be run on a weekly basis to assess the performance of fieldwork (henceforth referred to as "field check tables"). They provide summary information of the teams' performance on a range of data quality indicators. These can greatly assist in identifying weak or underperforming teams. These tables should be distributed weekly to the local Survey Management, local PI, World Bank/Washington, DC, and Kimetrica.

To help identify "problem" teams, most tables have a target value that each team should meet—those falling below it will be flagged. For example, Field Check Table 1 presents the Household Response Rate. The desired rate is 90%—any team failing to garner a 90 percent response rate will be flagged. The targets shown are example targets only; WSP staff, in light of local country conditions, should determine actual targets.

It should be noted that flagged teams are not necessarily doing a poor job. Following are some of the considerations that should be given before taking any (corrective) action:

- ◆ *Enumeration Conditions:* There may be variation in team results due to the region in which they are working (a mountainous rural area with poor roads versus the capital city with paved roads).
- ◆ *Number of cases being evaluated:* Oftentimes a team will not meet the target because only a small number of cases (households) have been entered by the data entry staff for the team in question. This is especially true at the start of fieldwork. Even several weeks into the fieldwork, there may be a low number of cases entered into the system for a given team. This could be due to the team failing to return questionnaires to the office, or it could be due to data entry staff failing to select a cross-section of clusters from each team.
- ◆ *Reasonableness of the target and closeness of the team to it:* First, review all the teams' performances as a whole. Are they all hovering in the same vicinity to one another, but much lower than the target? Possibly the target was set too high, and the rates you are receiving is what you can reasonably expect. However, it could be that the teams are underperforming. Generally, there will usually be a few teams that outperform the others in the quality of work they do. These teams should be used as the benchmark to which the underperforming teams are measured, and their rates of achievement should be used as a target. Markedly underperforming teams should be considered for dismissal.
- ◆ *Has a team been given enough time to improve?* Even after a team is informed of mistakes, improvements will not appear in the tables immediately because of the time lag involved in finishing their subsequent clusters, submitting these to the main offices, and waiting for data processing to enter them.

The following is a discussion of each of the field check tables, followed by the actual tables with hypothetical results, for illustrative purposes only. Each table follows a similar layout, listing each Team in the stub and topics being measured across the columns.

### **Field Check Table 1**

**What:** The results of the overall household interview in the columns, in order to ascertain percentage of successful interviews and level of non-response.

**Why:** This is an important measure of the team's performance. A team must be able to gain a high rate of successful interviews in order to ensure the validity of the survey itself.

**How:** Note when calculating the HH response rate that the denominator for the number of households does not include those households that were vacant, were temporarily away, or were otherwise absent. This is because the interviewer cannot be faulted if the household is away. On the other hand refused interviews are included in the denominator. This is subjective, for certainly the household has the right to refuse to participate. However, oftentimes it is the manner in which the team asks for the interview that causes a failure to gain permission to interview, and so even if this column is left out of the denominator, this column should be reviewed to ascertain if some teams have a much higher refusal rate than other teams.

It should be clear from the table that Team 4 is having significant problems, as they have a very low rate of successfully interviewed households, have a high rate of failure to find an eligible person at home to interview, and have a very high refusal rate. All of these points should be immediately addressed with the individual team members.

### **Field Check Table 2**

**What:** Was the team able to find/identify household members from the baseline survey?

**Why:** Failure to list and identify household members from the baseline survey in the endline survey will make comparability of baseline and endline personal data impossible.

**How:** The first three columns show the total number of persons who were listed in the endline survey, those persons who were found in the baseline household, and whether the target was met for that team. Note when setting a target that there will always be some attrition in the household due to deaths, marriages (causing persons to leave the household), persons coming of age and hence, leaving the household for work or education, or other reasons. Therefore, a reasonable target must be set for overall number of persons.

The next two columns track whether the child who was under 2 years of age during the baseline survey was found in the endline household. Due to the delay in the baseline and endline surveys, children under 5 but over 2 will most likely be over 5 years of age at the time of the endline survey and hence, no longer of interest. However, it is hoped that children under the age of 2 years at the time of the baseline survey will still be under the age of 5 during the endline survey and hence, of interest to the survey. Therefore, attempts

to ascertain whether or not the team is identifying those children will be of utmost importance. When setting this target, the prevalence of childhood diseases and other health issues needs to be taken into consideration.

Finally, it is suggested that the WSP may wish to track the number of new persons introduced in the household since the baseline survey, and this data is shown in the final table columns.

### **Field Check Table 3**

**What:** Is the team manipulating children's ages, or dropping them entirely, in order to drop them out of the under 5 age category?

**Why:** Intentionally changing a child's age, or omitting the child entirely from the household, will introduce significant bias into the data. This table attempts to determine this predicament with a table showing all children age 0-9 by single year of age.

**How:** This is a very simple, but very telling table. Oftentimes the interviewer manipulates data in order to conduct fewer interviews. This table presents the total number of children found (to date far within the keying cycle) of children for each of the ages from zero to nine years of age. Generally, one would expect an equal distribution of children across these ages, unless a significant event has occurred during this period such as war or the outbreak of a disease. Therefore, on the whole, the age ratio of children age 0-4 years (those eligible for selection in the survey) to be the same as those aged 5-9. However, if it isn't, then further investigation is warranted. Again we find that Team 4 is severely underperforming, with an age ratio of only 0.5. Frequently this is confirmed in the household schedule when it is apparent that the interviewers are changing dates of birth or age to make the child older, or even striking a child from the household and writing "moved" or some other such reason so as to omit the eligible person from the interview cycle.

### **Field Check Table 4**

**What:** Reviewing the accuracy of date of birth reporting for the child.

**Why:** Problems in accurate date of birth reporting can lead to inaccurate calculations of stunting and wasting.

**How:** This table shows various combinations of the child's month of birth, year of birth, and age being reported. A column is not included but could be shown that presents whether or not the child's reported age agrees with the reported date of birth. Note that even though day of birth was reported, this was not considered for the table, as that level of accuracy is not needed for calculating stunting and wasting.

### **Field Check Table 5**

**What:** Is the team taking correct anthropometry measurements for children under 2 years?

**Why:** Failure to accurately take a child's measurements will result in inaccurate estimations of stunting and wasting.

**How:** This is a partner table to Table 4. This table provides a column for each of the key measurements shown in Module 33. At a minimum, each column will report for each measurement of a child whether or not a measurement was taken—if no measurement was found, no tally will be given to that team. However ideally, the World Bank should provide a reference table that shows for a given child's age, what their head/arm/etc measurement should be, so that out of range values can be more easily determined.

### **Field Check Table 6**

**What:** Did the team succeed in recording most or all information for each child under the age of 2 and 5 years?

**Why:** Failure to collect key fields on a child will reduce the quality of the final dataset.

**How:** For each module that applies to children under 2 or under 5 years of age, key fields will be identified as necessary to be keyed. If all or the majority of those fields are blank, then this will count against the team.

It should be noted that the data entry program will not be at the mercy of the interviewers to determine whom the eligible children are for each of the modules pertaining to children. Rather, the software will calculate on its own which children are eligible, and not allow the data entry operator to proceed until all eligible children have been listed on the relevant module rosters. The data entry operators will further be trained so that if a child's entry is missing from a roster, they should provide missing values for the fields.

### **Field Check Table Summary**

In order to facilitate identification of underperforming teams, the Survey Management should review this summary table. It presents in a consolidated manner the targets from each of the previous tables in the columns and the teams in the stub, allowing easy detection of struggling teams.

### **Weekly Data Processing Status Reports**

The last table in this Appendix shows the suggested chart to use for tracking the progress of data processing. It records entries according to cluster number, such that only clusters that have entered the data processing cycle are shown. Operators responsible for the main entry and verification phases of each cluster are given, along with their operator identification code, date of entry, number of households entered for the cluster, and number of children age 2 and 5 years, respectively.

The very last line of this chart shows the total count and percentage to date of the various phases being tracked. For example, in this survey there are 181 clusters, which represent 100% of the total. To date, 92 clusters have been registered into the data processing



operation, which represents 51 percent of the total workload (registration means the expected totals (HH count, etc) for the cluster have been supplied). The next column shows that 81 clusters have been assigned to main entry, meaning they have either finished the main entry phase, or are actively being keyed—this represents 45 percent of the entire workload. The subsequent column indicates that 78 clusters have been accepted, meaning they have finished the main entry phase—this represents 43 percent of the entire workload. The subsequent two columns similarly track the verification phase, and finally, whether the unedited cluster has been saved—which can only occur once all keying differences between the main and verification operators have been resolved.

The final three columns display the total number of households entered in all clusters to date, the total number of children found under the age of 2 years, and the total number of children found under the age of 5 years. If there are estimates for the total counts expected for the survey (which there should be, from the listing operation), these can be used to quickly ascertain whether or not the survey is on track. In our chart, there is an underlying expectation of 2,715 total households for the survey—therefore the 977 households found to date represent 36 percent of this total. For this number to be meaningful, it should be compared against the "Date Accepted" entry, i.e., the number of clusters that have completed main entry. In our chart, 43 percent of the clusters have been keyed once, so one would anticipate the approximate same number of households to have been keyed—however, ours is a bit short, clocking in at only 36 percent. This difference should be investigated, as it is too far off the target number to be accounted for by slight variations between urban and urban clusters, etc. Target numbers were not used for the under 2 and under 5 columns, but similar estimates can be made if these numbers are given.

If this chart is run on a weekly basis and progress of each of the key indicators (main entry, verification entry, backup of unedited data, and backup of edited data (if editing is conducted)) is recorded in an Excel file, this will allow Survey Management to track the overall progress of each phase of the data processing operation.

One caveat about this chart should be made. If progress lags in the data entry operation, particularly main entry, it could be due to other reasons beyond the control of the data processing staff; for example, fieldwork failing to return completed clusters, forcing operators to sit idle. Data Processing Supervisor(s) and Survey Management must be in close coordination with one another throughout the fieldwork to ensure the steady return of clusters from the field to avoid this situation.

### Field Check Table #1: Household Response Rate

Percent distribution of sampled households by Interviewer Team # by Results of Household Interview

Team #	Result of household interview									TOTAL (10)	Number of Households (11)	Household response rate (%)* (12)	Target Not Met (< 90%) (13)
	Completed Interview (1)	Incomplete Interview (2)	HH Present, no elig. Respondent found (3)	Rescheduled (4)	HH Absent (No one home) (5)	HH Temp. Away (6)	Refused (7)	Dwelling Vacant (8)	HH Not Found (9)				
Team 1	87.1	0.9	1.5	1.2	3.2	1.1	2.5	1.0	1.5	100.0	478	92.0	-
Team 2	85.6	0.6	2.2	0.0	6.1	1.9	0.3	1.0	2.3	100.0	502	94.1	-
Team 3	82.4	0.2	8.2	0.0	1.3	0.8	6.5	0.0	0.6	100.0	491	84.2	84.2
Team 4	68.3	1.9	11.7	0.0	2.0	5.1	10.8	0.2	0.0	100.0	488	73.7	73.7
:													
Total	94.8	0.5	1.8	0.0	0.1	2.3	0.1	0.2	0.2	100.1	1,959	97.5	

\* HH Response rate = (1) / (1+2+3+4+7+9) \* 100. World Bank HQ should provide Kimetrica with a target response rate, numbers below which will be flagged.

## Field Check Table #2: Ability to Find Former HH Members

Percent distribution of women eligible for height and weight by result of height and weight measurement, and percentage of valid data for anthropometry, according to interviewer team

Team #	Number of Persons listed in the Endline HH	Baseline Person Still Living in Endline HH	Target Not Met for Baseline Persons Not Living in Endline HH (1)	Baseline Child < 2 Living in HH (2)	Target Not Met for Children < 2 Not Living in Endline HH (3)	Number of Households	Number of new HH members added to Endline HH	Average # of new HH members in Endline HH (4)
Team 1	438	98.2	-	99.1	-	90	13	0.14
Team 2	294	97.6	-	96.7	-	67	4	0.06
Team 3	581	99.1	-	97.3	-	102	7	0.07
Team 4	319	94.2	<b>94.2</b>	93.5	<b>93.5</b>	44	2	0.05
:								
Total		97.3		96.7				

(1) A target value needs to be assigned. Table uses 95%.

(2) I use 2 rather than 5, as those children age 2-5 in the baseline are most likely older than 5 now, whereas children under 2 are probably still under the age of 5 years for the endline.

(3) A target value needs to be assigned. I'm using a lower value than the HH, 96% (to account for higher childhood mortality-- though maybe I shouldn't?).

(4) A target value has not been assigned for this, but it's a number Survey Managers should be aware of.

### Field Check Table #3: Age Displacement of Children

Number of all children age 0-9 years listed in the household schedule by Interviewer Team # by Age, and Age Ratio 0-4 years/5-9 years

Team #	Children age 0-9 years										Number of Children Age 0-9 yrs	Age ratio (children 0-4/ children 5-9)	Target not met (< 0.80)
	0	1	2	3	4	5	6	7	8	9			
Team 1	10	11	12	9	14	13	12	8	13	14	117	0.94	-
Team 2	14	16	15	12	13	14	13	14	15	7	133	1.10	-
Team 3	12	13	11	13	11	9	16	17	19	20	140	0.73	<b>0.73</b>
Team 4	6	7	4	4	3	5	9	13	11	10	72	0.50	<b>0.50</b>
:													
Total	43	47	42	38	41	41	50	52	58	51	462	0.90	

Note: To meet the target, the age ratio of (children age 0-4 / children age 5-9) > 0.8. World Bank HQ should provide Kimetrica with a target number.

#### Field Check Table #4: Accuracy in Date of Birth Reporting for Children Under 5

Percent distribution of children under 5 by Interviewer Team # by Completeness of date of birth/age information

Team #	Data Collected					TOTAL	Number of Children Under 5 years	Target not met
	Month and Year of birth (1)	Year of birth only (2)	Age and Year of birth (3)	Age only (4)	Other/ No data (5)			
Team 1	100.0	0.0	0.0	0.0	0.0	100.0	656	-
Team 2	97.4	2.1	0.1	0.4	0.0	100.0	732	-
Team 3	98.6	0.6	0.0	0.8	0.0	100.0	890	-
Team 4	88.3	7.3	2.4	1.1	0.9	100.0	623	<b>88.3</b>
:								
<b>Total</b>	99.0	0.3	0.4	0.2	0.1	100.0	3,707	

Note: To meet the target, the number of births with both year and month of birth reported > 95%. World Bank HQ should provide Kimetrica with a target number.

### Field Check Table #5: Anthropometry Measurements for Children Under 2 years

Percent distribution of children under 2 years by Interviewer Team # by Valid Anthropometry Measurements Taken\*

Team #	Child's Weight Taken (G.33.6)	if child weighed with mother, was mother weighed (G33.7)	Height taken (G.33.9)	Arm Circumference (G.33.10)	Head (G.33.11)	Hemo-globin level (G.33.14)	Number of children < 2	Percent of children with all data recorded **
Team 1	97.2	92.3	98.9	65.7	98.6	93.6	1,883	59.4
Team 2	100.0	95.6	100.0	99.5	99.4	91.4	1,188	88.4
Team 3	93.2	98.0	94.2	91.4	55.1	93.5	1,297	46.7
Team 4	88.9	85.2	86.5	91.2	90.4	89.5	1,612	81.8
:								
Total	95.1	0.8	0.2	0.0	3.9	100.0	5,980	69.1

\*The World Bank or local country PI must define what the valid range is for each anthropometric measurement.

\*\*A target rate must be set so that under-performing teams can be flagged.

(8) = Number of children with valid data present in columns 1 through 6. This can never be higher than the lowest percentage shown for the team. For example, Team 1 only recorded the arm circumference 65.7% of the time, so the overall percentage will therefore be lower than this 65.7.

# Field Check Table #6: Children < 2 and < 5 accounted for in all Modules

Percentage of times that a team provided key information for a child

	For modules pertaining to Children under 5 years of age, percentage of times the correct children were listed						For modules pertaining to Children under 2 years of age, percentage of times the correct children were listed**				Target not met***
Team #	Module 14	Module 15*	Module 16	Module 17	Module 19	Module 20	Module 21	Module 22	Module 24	Module 25B	
Team 1	99.8	98.2	97.4	98.4	100.0	100.0	98.3	99.7	100.0	96.2	
Team 2	80.7	73.5	82.4	81.1	83.9	77.6	80.7	73.5	78.4	91.1	
Team 3	77.0	67.8	69.3	70.1	72.2	71.5	77.0	67.8	69.3	76.0	
Team 4	87.3	90.1	92.3	93.2	89.0	88.5	87.3	91.2	93.1	92.7	
:											
Total	86.2	82.4	85.4	85.7	86.3	84.4	85.8	83.1	85.2	89.0	-

\*Module may get dropped.

\*\*Modules 23, 26, and 29 have been dropped from the baseline survey.

\*\*\*A target (or targets, if any) needs to be set.

"correct" child means the child's info isn't missing.

## Field Check Table Summary

Interviewer Teams not meeting targets in selected field check tables

Team #	FC #1	FC #2, Target #1	FC #2, Target #2	FC #3	FC #4	FC #5	FC #6	FC #7
Team 1	-	-	-	-	-	59.4	target(s)	anything
Team 2	-	-	-	-	-	88.4	need	else WB
Team 3	84.2	-	-	0.73	-	46.7	to be	wants to
Team 4	73.7	94.2	93.5	0.50	88.3	81.8	set	see
:								tracked?



## Suggested Data Processing Status Report to be run Weekly

Cluster	Date Register	Entry Operator	Date Assigned	Date Accepted	Verif. Operator	Date Assigned	Date Verified	Backup Unedit.	Date Final	# HHs	# kids <2 yrs	# kids <5 yrs
1	80802	JOHN	07 80909	80830	LANCE	04 80901	80902	80902	0	15	18	31
2	80730	MARY	12 80815	80815	BRUCE	11 80815	80826	80826	0	15	19	32
3	80729	MARK	05 80814	80814	KAREN	06 80814	80819	80819	0	14	22	37
8	80728	LILIAN	16 80805	80805	BRUCE	11 80805	80812	80812	0	15	14	24
10	80902	PAT	03 80925	80925	KAREN	06 80925	80926	80926	0	15	8	16
11	80918	KAREN	06 80924	80924	JOHN	07 80924	80927	80927	0	15	9	17
12	80828	BRUCE	11 80902	80902	FLOYD	10 80902	80905	80905	0	13	11	19
13	80828	MIGUEL	16 80830	80830	JOHN	07 80830	80902	80902	0	16	5	13
:		:			:							
to max # clusters												
181	92		81	78		76	72	72	0	977	1109	1542
100%	51%		45%	43%		42%	40%	40%	0%	36%		