

## Report On Data Cleaning

### *Random Data Checking Procedure*

Following the dual data entry of all records by Nigerian consultants and the merging and cleaning of the data files(as outlined below) by World Bank staff, the hard copies of the questionnaires were randomly checked against the entries in the data files for errors by World Bank staff. Five LGAs were selected at random in both the Kogi and Lagos states. In each of these ten LGAs, the hard copy of the PHC Coordinator Questionnaire, the hard copy of the LGA Treasurer Questionnaire, and up to five hard copies of both the Staff Questionnaires and the Health Facility Questionnaires were randomly selected and checked against the entries in the data files<sup>1</sup>. While in several instances parts of the alphanumeric entries were abbreviated or omitted, no substantive differences between the hard copies of the questionnaires and the entries in the data files were found.

### *Merging and Cleaning of Data Files*

In order to facilitate the evaluation of data across the different levels of the Nigerian primary health care system surveyed in this study, the four survey data files were merged. Three types of problems arose: **(1)** In matching the Health Facility Questionnaire data file to the LGA Questionnaire data file, there were six facilities (out of 252) that shared three unique codes.<sup>2</sup> In matching the Health Facility Questionnaire data file to the Staff Questionnaire data file, **(2)** there were 7 facilities (out of 252) with no corresponding staff in the Staff Questionnaire data file **(3)** and 13 staff (out of 719) with no corresponding facility in the Health Facility Questionnaire data file.

Problem **(1)** was resolved for two of the six facilities by checking the hard copies of the questionnaires against the respective data files and correcting discrepancies. Problem **(1)** was resolved for the remaining four facilities by requesting an updated list of facility codes from the Nigerian consultants and correcting discrepancies.

Problem **(2)** was resolved for four of the seven facilities by checking hard copies of the questionnaires against the Health Facility Questionnaire data file and correcting discrepancies. Problem **(2)** was resolved for the remaining facilities by looking at the comments on the hard copies which revealed that conditions did not give the respective interviewers the opportunity to interview staff at the facilities in question. These facilities were retained in all the data files since they are indicative of the functionality of the respective facilities.

Problem **(3)** was resolved for three of the thirteen staff by checking hard copies of questionnaires against the Staff Questionnaire data file and correcting discrepancies. The remaining ten staff still cannot be matched to any existing facilities. In one case, the hard copy of the Staff Questionnaire cannot be found. We conjecture that these ten records were misplaced prior to data entry. These ten facilities were removed from the merged data file yet retained for evaluation in the Staff Questionnaire data file.

---

<sup>1</sup> For example, in one of the LGAs randomly selected in the Kogi state – Ijumu – only three health facilities were surveyed (two original facilities and one replacement facility), and, thus, only three Health Facility Questionnaires were selected and checked against the entries in the data file.

<sup>2</sup> These unique codes were the result of a systematic combination of a unique state code, a unique LGA code and a unique facility code.