**Improving Maternal and Child Health in India: Evaluating Demand and Supply Side Strategies**

**Impact Evaluation of Chiranjeevi Yojana in Gujarat**

# Documentation on Data Cleaning and Data Entry

This note describes the data cleaning and changes made to address data entry error that was identified in the data. The changes described fall into three different broad categories: a) Data Modification issues, b) Data entry issues and c) Table linkage issues.

## Data Modification Issues

- Variables for Cluster and Strata were combined to create a new Unique ID variable.
- Codes for one of the BPL survey question data was recoded (see documentation on recode file for details).

## Data Entry Issues

In certain questions such as question related to indebtness i.e. q127 and q128, data entry errors were identified based on consistency checks (q127 indicate debts (1=yes), while q128 reported "no indebtedness"). In such instances, the inconsistency was resolved by cross checking with the specific survey schedules. If the error was made during data entry, the data was modified to reflect the responses recorded in the completed survey schedule. If the data was inconsistent in the survey schedule, it was not changed. Similarly in case of questions about migration and ownership of land study corrected the few cases of data entry by cross checking with the survey instrument.

## Table Linkage Issues

There were two issues that came up during data entry with regard to linking of data tables. First issue was related to linking the table on deceased members (starting with q104ii) with the table on women who have expired (starting with q109). As the two tables did not have a linking ID number, the study team linked the two tables by considering gender, age and marital status in both the tables.

Second issue was that the field team collected series of information from households in different tables such as the household roster table, the mother's pregnancy outcome table etc. Although these tables collected different set of information from the households but they were linked with each other by the ID number of the household members and particularly by the ID number of the mother. In a few cases, the field investigators did not copy the same line id number which created problems in linking these tables during data entry. The data entry team corrected this problem by referring to the particular survey schedules to check the line ID number in order to match it up in each table.