IMATCHINE: Evaluation of the Chiranjeevi Yojana in Gujarat

Documentation of recode files

Introduction 2
Survey design and questionnaire content 2
Key constructed variables 2
Documentation of recode files 3
BPL items 4
Household recode 5
Construction of BPL date 5
Household members recode 5
Pregnancy recode 6
Last birth recode 6
Childrens recode 6
Additional notes on the recodes files 7
Constructed household covariates 7

Introduction

This note describes key changes to the raw input dataset and the preparation of the recode files. Additional details are described in the accompanying Stata do-file. Also see the questionnaire pdf file for annotations on the structure and specific issues.

The input data consists of the household questionnaire and the houselisting data for those households that were interviewed.

CY stands for Chiranjeevi Yojana.

Survey design and questionnaire content

The survey covered 600 villages across 25 districts and 75 blocks/talukas. 6,002 households were visited and 5,663 interviews completed (response rate 94%).

The analysis dataset includes 8,168 births between 2005 and the data collection in 2010, of which 65.5% were institutional deliveries.

The household survey included several sections:

- Household demographic and socio-economic characteristics. In particular, the questionnaire included retrospective questions from the BPL scoring scheme.
- Participation and awareness of government schemes
- Woman respondent background
- Reproductive details, including pregnancies that did and did not lead to a birth
- For the last 3 children born after January 2005
    - Pregnancy and post natal care
    - Child immunization, health and nutrition
    - Height and weight as measured by interviewer
- For the last birth:
    - Additional details on pregnancy and post natal care
    - Routine checkups
    - Acute and chronic health problems after deliver
    - Details on delivery and quality for the applicable delivery context, e.g. institutional or home delivery
    - Child care and safety
    - Costs of delivery and sources of payments

Key constructed variables

bpls	BPL score in 2002, retrospective from houselisting

bplconstr BPL status constructed with houselisting BPL score and eligibility cutoff.

bpls_today BPL score in 2010, from household survey
bpl Self-reported BPL status in 2010

The district rollouts were in Dec 2005 and Jan 2007. We use January 2006 and 2007 as cutoff dates, which allows the early adopters some extra time to implement and also helps with missing month information in the birtdates.

Documentation of recode files

The recode files are modified and reshaped versions of the raw data.

File description
Household Recode: one line for each household.
- All items that not specified elsewhere

=> merge using hhid

Household Member Recode: one line for every household member.
- Section A2 on living and deceased household members (with additional details for women who are deceased)

=> merge using hhid and pid

Pregnancy Recode: one line for every pregnancy of eligible woman
- Section B-1 to B-2 (women's schedule)
- Section B-3 on pregnancies not leading to births
- Section B-4 on pregnancies leading to births
- Section B-5 on pregnancy and post natal care for all children born >= Jan 2005

=> mother's ID is in mline
=> merge using hhid and pid

Last Birth Recode: one line for the last delivery/birth of eligible woman
- Section B-6 to B-13 on pregnancy and care at various times

=> merge using hhid and pid

Children's Recode: one line for the last 3 living children of eligible woman.
- Section B-14 on immunization of last 3 children
- Section B-15 on height, weight measurements

=> merge using hhid and pid

Note: In the spirit of questionnaire section B-5 these last 3 births were generally understood to mean “the most recent three births that occurred after 2005”. As result some women have details on fewer than 3 births although they may have had births

prior to 2005. In the final analysis dataset the last birth and last 3 births are restricted to have occurred after 2005.

Key variables for identifying cases

hhid	household ID
pid	hhold member personal ID from roster, unique within each hhold PID >=64 is for unsuccessful pregnancies and moved/deceased children (see below comments on pregnancy recode file)
pregid	pregnancy ID for all (successful/unsuccessful) pregnancies
mline	roster ID of the female respondent

Some other variables are only used to identify cases in tables. For example:

pregid_*	table-specific pregnancy ID, where * is the table section
cid_b5	child ID for table in sections B-5 and B-6 to B-13 (these children also have a pregnancy ID)

Other key constructed variables

age	Corrected age for children; see below for details.
q107	Age for all hhold members. This includes the children. For them, the ages have been harmonized with the corrected "age", so q107==age for all hhold members living in the hhold. 28 Children that have moved/deceased have "age" but not q107 since they are not in the hhold roster.
sex	Sex for children
q106	Sex for all hhold members

General notes

See additional comments in the questionnaire pdf.

- Unsuccessful pregnancies have PID = 64, 65, 66. These are not listed in the household roster.
- Successful pregnancies of children who moved or are deceased have PID = 74, 75, 76. These are also not listed in the household roster.

BPL items

- Item 4 on work/school status of children. It is unclear how to code the "n/a" cases for households where this question does not apply. The BPL guide does not provide instructions for such situations. We infer the official scoring by using the

official BPL lists.[1] If the item is applicable it is scored at the recorded value. If it is not applicable it is set to 4.

- The cutoff for the BPL eligibility is a score total of 20. Households with a score <=20 are eligible.

Household recode

- Missing BPL date values in q140 are imputed with information from q141 on the BPL wave that is applicable. There is no date if the source is "village authority". In these cases the membership in CY (q151*_2) is used to obtain a date. Where all dates are missing, the district-wide rollout schedule is used as year. A missing month is set to January which is conservative since that way some non-BPL households are counted as BPL.

Construction of BPL date

For households that had BPL at the time of the survey, the year and month of BPL receipt are recorded in q140. Some date values are missing and are imputed as follows:

- Years are filled with q141 on timing on BPL survey
- For BPL from village authority, years are filled from q151 on participation in CY
- Missing month values are assumed to be January. That way some non-BPL households are counted as BPL – leading to a conservative estimate of the program impact.
- BPL households without any date information are assigned the applicable CY rollout date of Jan06 or Jan07

The final date of BPL receipt is a month-year value encoded in the variable "bpldate".

Household members recode

- For ~70 children the mother appears <= 12 years older than the child. It is unclear whose age is incorrect. The mother's ID in the hhold roster (q107a) has been set to missing (.m).
- Note that it is not possible to perfectly merge the mothers to all children, since some mothers are not recorded. The "mothers line number" in section B-4 does not always match with the line number in the household roster (variables mline and q107a_*). That may be legitimate in some circumstances. One explanation

1 http://ses2002.guj.nic.in/Home.aspx accessed on March 1, 2011. Note that in some cases the BPL lists always set this item to 4, replacing a available value. Yet in other instances the list uses the procedure we adopt. This is also in the spirit of the BPL guide, which uses different scoring.

is that the biological mother died <= Jan 05, so she is neither in the living nor deceased household roster. This seems to hold in at least some cases, e.g. hhid==10104.

Pregnancy recode

- The pregnancy and birth history is recorded for 1 woman per household.
- Some children may be classified as multiple births but have no sibling of the same age. That may be plausible, e.g. the sibling may have died. Also, the age for the children is imprecisely recorded. Multiple and single births may appear incorrectly coded.
- For analysis of children only use the variable "age" and "bdate" variables that are created in the pregnancy recode, as well as "anc_start" and "pnc_end" created in the analysis code. The age variables for living children (q406 and q107) coincided in the original data. However, the birth date in q404 was not consistent in some cases. The age and birth date for births in the pregnancy recode have been corrected. In 632 observations the age was available and used to impute the missing year. Age was always available when year was available, so there were no imputations in that direction. For deceased children the "age" is set to missing (.n) and the death and birth dates are unchanged (q407 and q404, respectively). Ages for other household members have not been corrected and are available in q107. The original age and date year/month are available in the data as _age_orig, _q404a_orig and q404b_orig.
- The facility name in text field q506 is not always consistent with the facility type in q505. E.g. q506=="other home" whereas q505 is 1 for "home" instead of 2. q505 was corrected after data entry. Disregard the text field q506 in case of inconsistency with q505.

Last birth recode

- 12% of observations have some mismatch between itemized government support in q1307 and the applicable subtotal in q1314b. For these cases the subtotal set to to sum of the items. In most cases, the sum of the items is less than the subtotal in which case this approach is conservative.
- Minor issues with subtotals are documented and corrected in the code.

Childrens recode

- While the questionnaire asks for mother’s height and weight, this data was not collected in the field work.

Additional notes on the recodes files

(1) Pregnancy recode has more children than hhold roster. As result the pregnancy recode may not perfectly merge with other recode files. There are legitimate reasons:

> (a) Unsuccessful pregnancies are in the pregnancy roster but are not in the hhold roster (by definition). They have a PID and pregid==64, 65 or 66
>
> (b) Deceased children are in pregnancy roster but may not be listed in the hhold roster (they may have died < Jan 2005 or not lived in hhold). They do have both the PID and pregid==74, 75 or 76
>
> => Roster is missing some PIDs so there should be no _merge==2
>
> use "$f_pregn" if !inlist(pregid, 64, 65, 66, 74, 75, 76), clear
> merge 1:1 hhid pid using "$f_hhmembers", keepusing(q106 q107) assert(1 3)
> tab pid _merge, m

Constructed household covariates

- The dates for unsuccessful pregnancies are missing in ~30% of cases, possibly due to respondent recall. These pregnancies are used to calculate the "total pregnancies" but are ignored for time-specific measures, e.g. pregnancies before/after the program rollout. Only successful/unsuccessful pregnancies with dates are used for the latter.
- The "total number of births" in q206c adds up exactly to the sum of individual pregnancies in about 91% of hholds. Another 6% have 1 or 2 births more in the total than in the sum of the births. One cause of the discrepancy is missing birth date information for the individual birth records, possibly due to recall problems. A similar issue applies to the "total births since 2005" in q412. If the date is not critical, these summary variables are used for total births (q206c) and total births since 2005 (q412).