

Glossary of important terms and names:

@	holds the pointer in SAS for conditional input :, here by card number
ACPP	Archives CPP - data as transferred from mainframe to PC format by National Archives and Records Admin., Center for Electronic Records
ASCII	A standard plain text format for PC electronic files with no formatting or database aspects to selectronic storage.
Definition of codes, 1978 CPP hardcopy document	In modern jargon a combination data dictionary and codebook finalized in April 1978 after the final version of the CPP electronic data was created (MDF0378.ASC). It was reformatted and incorporated into the microfiche documentation. Pre PC and word processing, it exists in hardcopy only (parts have been word-processed), often old xerox copies several generations removed from the original. This hardcopy is probably the most useful and definitive document in understanding the CPP electronic data . It did not scan for OCR.
EBCDIC	
fiche	Microfiche. Reduced scale photographs of documents common before the advent of electronic documents. CPP documentation (about 6,000 pages) was microfiched at 29x on 75 microfiche.
Field	Numbered paragraph in 'definition of codes' which may have more than one item of information/variable.
Forms	Standardized instruments and schedules used for data collection in the CPP. Not to be confused with normal forms - the principles of modern database design. Forms came in series and are digits 2-4 of each punchcard record. They were revised often and the version/revision is digit 5 of each punchcard record. A copy of each form is supposedly in the microfiche documentation. Many did not photograph well and are blurred.
Master Data File MDF0378P.ASC	All CPP data ; 6.1 million punchcard records, are in the final version of this electronic file created April 1978. At least 6700 variables.
merging or match-merging	linking or combining datasets with common individuals to add variables only. Can be a complex process due to duplicate individuals (not keyed/indexed) and individuals only in one dataset. Inexperienced analysts are advised to master merging two datasets before attempting 3 or more in one step. Not the same as concatenating or appending -

adding individuals (sometimes variables too).

NARA, CER	National Archives and Records Administration, Center for Electronic Records, College Park, Maryland.
N CPP	In this context, the version of the CPP ASCII data supplied by Dr. Mark Klebanov, NICHD, NIH on two CD-ROM disks. It is the A CPP EBCDIC master data file mdf0378.asc and/or work files either subsetted/converted into 62 ASCII datasets by topic or form or unchanged. N CPP data was used as the basis for SAS input programming because it was obtained first, removes a step, and was used by others outside JHU.
NINDB (id #)	ID number assigned when the CPP's institutional home was the National Institute for Neurological Diseases and Blindness . It is 7 digits for the mother/family, 8 digits for the pregnancy, 9 for the child(ren). Context supplies which is apropos but the varying length of the key/index causes problems in match-merging. See CD insert for more.
Normal forms	The rules of modern database design and best practice if followed carefully. Any analyst/manager is warned that the original CPP data breaks them except for the compendium variable file and a few others.
pdf	Well known Adobe Acrobat portable document format legible with the freely available Acrobat Reader.
punch-card	Physical paper card with holes punched according to values of a variable used to input data into computers before the advent of direct electronic keyboard input.
SAS	Statistical Analysis System. De facto standard software package in health research. for manipulating large datasets. SAS Institute, Cary, North Carolina. Version 8.2e, Windows..
SPSS	Most used software package for analysis in psychosocial and related health research. Less data management features than SAS. Conceptual Software Inc., Houston, Texas. Version 11.0, Windows.
STATA	Software package used at JHMI for training. Version 7.

VARFILE.ASC

A useful dataset containing a compendium/**selection of 1200 variables** from the total 6,700. Sole original dataset with substantial information about mother and child from the womb through age 8.

**** It has **one record per child** or pregnancy. ****

Work files

Subsets of CPP data (30) from the electronic master file used on an ongoing basis for analysis. The original CPP documentation on 'fiche recommends using these files, not the Master File, for analyses. NARA CER staff used the DOS 8.3 naming convention for the ASCII versions so these filenames are shorter than the IBM mainframe filenames.