<u>**1990 Census Measures**</u>
<u>**Geocoding Process**</u>

Fast Track Project Technical Report Addendum
Patrick S. Malone (919-668-6910; malone@alumni.duke.edu)
9-May-00

<u>**Table of Contents**</u>

<u>**Citation**</u>
Report
Malone, P. S. (2000). <u>1990 Census measures: Geocoding process</u> (Technical Report) [On-line].
Available: http://www.fasttrackproject.org/

I.   <u>Overview</u>

Target child home addresses from the Family Information Form (FIF, PxB) in each study year have been coded with respect to state, county, and tract/block numbering area (BNA) codes from the U.S. Census, and can thereby be linked to summary information from the Census or other datasets on households and individuals in the respective areas

The initial geocoding of Fast Track student addresses was conducted in 1999, based on address data from the FIF collected through 1997. Because parts of the coding process involved Research Coordinators' and interviewers' personal knowledge of student addresses, all available data were coded simultaneously (versus by cohort) to minimize delays.

The coding process will be updated each year as the FIF data are available, and this report updated. The Table at the end of this report summarizes coding successes each year.

II.   <u>Coding Through 1997</u>

By 1997, Cohort 1 had seven years of addresses conceivably available ($\underline{n}$ = 618, total = 4,326); Cohort 2, six years ($\underline{n}$ = 321, total = 1,926); and Cohort 3, five years ($\underline{n}$ = 260, total = 1,300); for a total of 7,552 student by year combinations. Of these, 7,537 (99.8%) had FIF records, and 6,994 (92.6%) had valid address entries (these 6,994 entries were comprised of only 2,694 unique addresses, but were treated as individual records in the matching process). Through a multi-stage matching process (detailed below), 6,951 addresses entries (2,671 unique addresses) were successfully coded by state, county, and tract/BNA (99.4% of the valid addresses, 92.0% of the total possible entries). The 43 unmatched records corresponded to 23 unique addresses (0.9% of total), including 1 address outside the United States.

The first stage of coding was automatic matching conducted by Geographic Data Technology (GDT), a firm specializing in map database development (http://www.geographic.com/). All street addresses were processed; rural route and Post Office box addresses do not reflect specific locations and can not be coded in this way. The initial automatic process resulted in codes for 5,823 (83.3%) matches. The Data Center and Research Coordinators reviewed and edited the unmatched addresses, allowing coding of an additional 222 (18.9% of the remainder) addresses (total 6,045; 86.4%). The results of GDT's matching included U.S. Census state, county, tract or BNA, blockgroup, and block codes, as well as latitude and longitude of the addresses (see Census documentation Appendix A, attached, for definitions).

The remaining 949 addresses included P.O. box and rural route addresses, non-U.S. addresses, invalid addresses, and valid addresses that were not in GDT's database. The latter could include, for example, special addresses in trailer parks or housing projects or addresses that were developed after the 1990 Census. Data Center staff searched for the remaining addresses in the TIGER/Census Tract Street Index

database published by the U.S. Census. This database includes the state, county, and tract/BNA codes for U.S. addresses, and was extracted from the Census records in early 1997, thus including more recently developed addresses. 241 (25.4%) addresses were coded from the TIGER files (total 6,286; 89.9%).

Finally, the remaining 708 addresses were coded by the Research Coordinators, based on the physical locations of the students' homes. Using Census Tract/BNA Outline maps from the U.S. Census, Research Coordinators identified the tract/BNA for each address (maps were not used for addresses located outside of the initial state, as in most such cases the Research Coordinator would not know the physical location of the student's home). 665 (93.9%) addresses were matched to tract/BNA in this stage (total 6,951; 99.4%).

Table 1. Record Matching Summary

	Entries	Matches
Through 1997		
No FIF record	15	
No usable address	543	
Initial batch-coding (GDT)	5,823	5,823
Second batch-coding (GDT)	222	222
Tiger/CTSI match	241	241
Map match	665	665
Unmatched	43	
Total	7,552	6,951