Data Appendix for "Short, Medium, and Long-Term Consequences of Poor Infant Health: An Analysis Using Siblings and Twins"

P. Oreopoulos, M. Stabile, L. Roos, and R. Walld

This data appendix describes the data sources used in the paper in greater detail. A table listing the data sources used and the sample years is included below. The birth data originate from Manitoba Health hospital records. The registry contains information on births in Manitoba since 1970. Siblings are linked to mothers using hospital birth record information. The registry also attaches to every birth a family identification number (called the Registration Number or REGNO), which links the infant to the 'family head', usually the father. Therefore, the registry data allow us to identify the mother in all cases. Fathers are identifiable in 85 percent of cases. When an individual turns eighteen years old, he or she receives his or her own REGNO. On marriage, a female receives the REGNO of her husband. Both the mother's identification number (an encrypted PHIN or Personal Health Identification Number) and REGNO are used to define siblings[1]. Several checks on this algorithm as applied to the seven years of birth cohorts (looking at missing data, the number of children designated as having the same mother and father, and complicated blended families) have indicated it to be highly accurate.

Two siblings with the same birth date are designated as twins. The birth records do not allow us to distinguish between monozygotic and dizygotic twins. Based on

[1] Siblings are noted as "full siblings" if they are children of the same mother (as noted on the birth record) and the same man is noted on the research registry (using the child's REG NO) as 'family head' at the time of the child's birth. Slightly over 85 percent of those identified as siblings (from having the same mother) meet the criterion set out above.

earlier descriptive studies [e.g. Conley et al., 2003], our twins data are likely comprised of roughly 25 percent monozygotic pairs and 75 percent dizygotic pairs.

Information on the provincial language arts test is taken from education enrollment records and linked to the provincial registry. Taken in grade 12, these tests contribute 30 percent to the students' final course grade. Individuals pass the language arts test by scoring 50 percent or more on a comprehensive exam. The test focuses on reading comprehension, exploring and expanding on ideas from texts, the management of ideas and information, and writing and editing skills. For each birth cohort, we record the test score in 5 percentage point categories (13 in total, with a residual 14^{th} for students scoring between 0 and 35 percent) in the year that most students write the test. Within each birth cohort, approximately 35% of test scores are missing. For these students we impute test scores based on the reason for missing information (ranking them below the lowest scoring category among those who wrote the test). These additional categories, listed by highest to lowest rank are: absent (about 1 percent of each birth cohort sample); In grade 12 but not tested (about 8 percent); In grade 11 or lower (about 19 percent), Not enrolled (about 2 percent), and Withdrawn from School (about 10 percent). For the entire sample, we therefore have 19 test score categories. Following methods forwarded by Mosteller and Tukey (1977) and Willms (1986), we then compute a standardized score for each individual by assuming an underlying logit distribution, which is divided into pieces according to the percentage of cohort members in each category. Scores are calculated separately for each birth cohort because of small changes in the categories available and in the percentage distribution each year. In a typical year, the highest scorers are given an index score of 2.96, while those withdrawn from school are given a

score of -1.84. The logit transform produces an index with an overall mean of zero and a standard deviation of one. The ordering on this index is closely correlated with the student's eventual graduation status.

The postal code from the family head's address identifies the street or building where the family lives. The address of the family is updated about every six months. To proxy for general social economic background, family income in the 2001 Census was aggregated and averaged over Enumeration Areas, which were in turn matched to corresponding postal code addresses in our sample. Enumeration areas contain a population of about 400 to 700 persons. The areas were ranked from highest to lowest income and then grouped into five population quintiles. Mustard et al. (1999) and Roos et al. (2005) show a substantial correlation (0.435) between our measure of persons' neighborhood average income and self-reported household income (not available in our data).

The province of Manitoba was chosen because of the unique ability to link the sources of data used in this paper. With a population of 1.17 million, Manitoba has the 5th largest population among Canada's provinces and territories. Over half the population lives in the capital of Winnipeg, making it the 9th largest city in Canada. Manitoba has a relatively large aboriginal population (12.7%) Within Canada, Manitoba has generally ranked in the mid-range of a series of indicators of health status, socioeconomics, and health care expenditures.

References:

Shanahan M, Gousseau C. "Using the POPULIS framework for interprovincial comparisons of expenditures on health care," Med Care, 37 (6 Supplement), JS83-100.

Statistics Canada (2006). Canadian Statistics. Retrieved February 27, 2006 from the Government of Canada, Statistics Canada Web site: http://www40.statcan.ca/l01/cst01/index.htm

Years of Coverage by Data Type:

Data	Years
Physician claims:	1978-2003
(for total costs to age 12-17)	
Hospital claims:	1970-2003
(birthweight, apgar, gestation).	
Registry:	1978-2003
(coverage, mortality and mobility)	
Vital stats:	1978-1986
(cause of death in perinatal period)	
Education:	1995-2003
(attainment and age-appropriate grade)	
Income assistance:	1995-2003
(income assistance 0/1 and duration age 18+)	