

CHILD HEALTH AND DEVELOPMENT STUDY

3. Data Collection

First observe that the sample of 680 infants used in the study are white, male infants. Employing a subset of data homogeneous with respect to sex and race increases the sensitivity of the analyses by removing sources of variation. The practice of examining data separately for sex-race categories is a typical analytic strategy.

It is obvious that the results of any statistical analysis are affected by the method the data were collected. It is obvious that some measurements are inaccurate. Nevertheless, as long as there is no clear evidence of a systematic bias in some groups, there is no reason to expect that the conclusions based on the model will be affected by the measurement errors considering the large size of the group under study.

As part of the study, a lengthy prenatal interview was conducted with each female participant. Extensive information was gathered on a large number of pregnancy-related variables. During the interview, which generally took place early in pregnancy, great effort was made to determine as reliably as possible the date of the first day of her last normal menstrual period. From this date and the date of termination of pregnancy, an estimate of the length of gestation was derived. The estimate is widely recognized to be very accurate. Moreover, the units of measurement (weeks) make the occurrence of the error very unlikely.

The number of cigarettes smoked per day is the value reported by the mother during the interview. It is reasonable to expect that the number might be underestimated (unless zero). As a consequence, the effect of smoking on infant birth weight may be overestimated.

The birth weight is subject to measurement error, but as long as there is no reason to suspect that there is a systematic measurement bias in some groups of interest (for example, the infants born to smoking mothers), it is very unlikely that our inferences will be affected.