Analysis Tips

Collapsing Variables

The medical history form contains some condition/disease status (e.g. emphysema, asthma) variables coded as "never been told", "first told during the past year", "first told more than one year ago". For analytical purposes these variables should be collapsed into "never" or "ever". Due to small sample sizes for non-white/non-black races, RACE01 is often collapsed into White/Non-white or Black/Non-black.

Self-Reported Variables

In cases where confirmed or calculated variables are available, these are preferable over self-reported variables. All calculated variables are defined in the CALCVAR.DOC file. This is especially true for diabetic status, hypertension and baseline event status.

For diabetic status, DIABADA or DIABWHO (the American Diabetes Association and World Health Organization criteria respectively) should be use rather than self-reported diabetes from the medical history or other forms. These variables are only available for years when blood was drawn. For analyses that require diabetic status at other years, DIABADA/DIABWHO closest in time can be used.

The calculated hypertension status, HYPER, is available at all years except year 8 when blood pressure was not measured.

There are two versions for all baseline events status variables, *BASE and *BLMOD. The *BASE variables are self-reported and some of those events could not be confirmed. The *BLMOD variables are events confirmed at baseline and updated retroactively during events surveillance. The *BLMOD variables reflect who is at risk for an incident event and should be used to indicate baseline status in an incident event analysis.

Hypertension

The calculated hypertension variable has always been coded as Normal/Boderline/Hypertensive. These levels are now outdated. Borderline and hypertensive should be combined into one level of hypertension.

Adjusted values

Some blood measurements have been adjusted for laboratory drift. More detail on adjusted variables can be found in CALCVAR.DOC. Though most of these adjustments are minor, drift may have an effect on analyses that estimate changes of blood levels over time. For this reason adjusted values should generally be used rather than raw values when available. This is particularly important for longitudinal analyses and when the baseline years for the two cohorts are combined.

C-reactive protein (CRP) was initially measured only on baseline bloods. Later, it was measured on year 5 bloods using a different assay. The values obtained from the two assays were not directly comparable without adjustment. For the New Cohort, while all three CRP values (adjusted, original and year 5) contain valid data, they all originate from the same blood sample. It is only appropriate to consider change in CRP from baseline (year 2) to year 5 for the Original Cohort.