

## Overview of Agricultural Health Study Data

Agricultural Health Study (AHS) master data files are organized by the AHS Phase in which the data were collected:

- Phase I – Questionnaire data collected between 1993 and 1997
- Phase II – Questionnaire data collected between 1999 and October 2005 – These files also contain cumulative exposure data constructed by combining information from both Phase I and Phase II questionnaire responses
- General – Demographic, registry, and biospecimen data collected over multiple project phases.

There are also two sets of analytic files which have been created using data from the master files:

- **Cancer Analytic Lag File (CALF)** – This file combines pesticide exposure data for all applicators in the study. It contains data from both Phase I and Phase II questionnaires and selected cancer data for incident cancers. Records represent total exposure to pesticides during sequential two-year time periods but are censored when respondents die, are diagnosed with an incident cancer, or move out of state (and are therefore no longer tracked by the Iowa or NC cancer registries).
- **General Uncensored Lag File (GULF)** – This file is similar to the CALF, but contains only exposure and mortality data. It does not contain cancer data and is censored only on mortality. **NOTE: THIS FILE HAS NOT BEEN CREATED AT THIS TIME.**

The Phase I and II files contain a wealth of exposure data, life-style data (e.g., smoking status), and respondent-reported health outcomes on a wide variety of illnesses.

The general files are usually used in conjunction with the questionnaire files:

- The **Demographic File** contains one record per respondent and contains our best information on demographic variables (e.g., birth date, gender, BMI), vital status, and AHS data collection activities in which the respondent has participated.
- The **Mortality File** contains one record for each person who has been reported as deceased by the state registries or the National Death Index. The data include date of death, underlying cause of death, and contributing causes of death.
- The **Cancer Registry File** contains one record for each cancer diagnosis for a member of the AHS cohort. Since a person can have multiple cancer diagnoses, there can be multiple records per person in the file. The file includes diagnosis date, cancer site, morphology, laterality, and stage as well as a number of other cancer-related variables.
- The **Buccal Receipt File** contains one record for each buccal collection kit returned by the respondents. Since some respondents returned multiple kits during Phase II, there can be multiple records per respondent. In addition, some of the kits which were returned were empty or did not have a signed consent form. Each record contains the participant identifier, the date the kit was returned to the processing lab, an indication of the amount

of material returned in the collection cup, whether an informed consent form was received. Note that since some specimens have been used for DNA analyses, the fact that a buccal sample was received does not necessarily mean that it is still available for analysis. This can be determined by looking in the BSI.