Appendix G

Calculations to Estimate Sample Size

Needed for a Long Term Cohort Study in

Fire Fighters
SAMPLE SIZE REQUIREMENTS FOR COHORT STUDIES

A cohort study is an epidemiological study in which exposed and unexposed groups of people are followed over time in order to measure rates of disease development in the two groups. By comparing the rates in the two groups, exposed and unexposed, the relative and absolute increases in risk associated with exposure can be estimated. One critical consideration in designing a cohort study is the size of the included population. Larger studies can more precisely describe the extent of the increased risk, compared with smaller studies. In fact, a study population that is too small may give potentially misleading results because results may not be statistically significant and are interpreted as “null” or “negative” as a result. Sometimes it is possible to enlarge the study population to the extent needed to be certain that a study is sufficiently large; sometimes, the sample size is fixed by the number of persons available for an investigation.

Sample size needs for a particular study reflect the magnitude of the expected increase in risk from the exposure under investigation, the size of the population available for study, and the resources (financial and time constraints, for example) available for the study. If the population numbers are fixed (e.g., all Anne Arundel County fire fighters), then statistical power is calculated to determine the likelihood of finding different levels of risk. Statistical power refers to the probability of finding an increase in risk of a particular magnitude to be statistically significant. If the population size is not fixed, than estimates are made of the numbers of participants needed to detect the anticipated level of risk with a reasonable probability. In this Appendix, we offer examples of such
calculations as a rough indication of the feasibility of carrying out a cohort study among Anne Arundel County fire fighters or among a broader group of fire fighters.

We offer example sample size calculations below for lung cancer and brain cancer. To make these calculations, we needed to have the rates of cancer incidence (new cases) in the population. For this purpose, we used routinely collected national figures. Incidence rates for lung cancer and brain cancer for males were obtained from the Surveillance, Epidemiology, and End Results (SEER) Program website (http://seer.cancer.gov/). The SEER Program tracks cancer incidence for a number of states, regions, and cities in the United States. We used them as the baseline rates that would occur in the population, absent an exposure that increased risk.

The SEER rates by age were assigned weights using census data for Anne Arundel County for males aged 35 to 64 so that they would be reflective of a population with age approximately equivalent to that of active and retired fire fighters at risk for cancer. Those age categories with a greater percentage of the population were assigned greater weight than age categories with a smaller percentage of the total population under study. To make the sample size calculations, we assumed that being a fire fighter would increase cancer incidence by 20, 50, 100, and 400 percent, corresponding to relative risks of 1.2, 1.5, 2.0, and 4.0. These were considered to cover a reasonable range of excess risk in fire fighters, from a small to a substantial increase. The expected incidence rates in exposed individuals (fire fighters) were obtained by multiplying these weighted rates with relative risks of 1.2, 1.5, 2.0, and 4.0.

The Poisson distribution formula was used to calculate the number of people needed for a cohort study; this formula is appropriate for incidence rates in a cohort
study. The formula uses the rate of disease in unexposed and exposed populations to
calculate the person years (number of people multiplied by years of follow-up) needed in
a study. This is then multiplied by the number of years of follow-up. Assuming, a ten
year follow-up, a cohort study on lung cancer would need 3,882, 2,972, 2,349, and 1,725
people to detect significantly increased relative risk values of 1.2, 1.5, 2.0, and 4.0,
respectively. To study brain cancer, a cancer with a lower incidence rate than lung
cancer, study populations of 27,861, 21,331, 16,854, and 12,382 would be needed to find
significantly increased relative risks of 1.2, 1.5, 2.0, and 4.0, respectively, in a 10-year
cohort study.

These calculations imply that a population comparable to that of Anne Arundel
County fire fighters exposed to the PCB-fueled training fires would have limited
statistical power to detect increased relative risks for either brain cancer or lung cancer.
Only relatively extreme risks would be found to be statistically significant.