Arsenic data from Southwestern Taiwan

Louise Ryan and Knashawn Morales

December 20, 2003

This document describes the arsenic data from Southwestern Taiwan that were described and analyzed in two separate reports from the National Academy of Science. The data were also analyzed by Morales et al (2000) and also discussed by Ryan (2004).

The files *mlun.sw.txt, mblad.sw.txt, flun.sw.txt* and *fblad.sw.txt* contain the data files for male and female lung and bladder cancer from Southwestern Taiwan. The naming conventions are fairly self-explanatory: files starting with 'm' are for males while files starting with 'f' are for females. Data are available from 42 villages.

Note that the data sets also include the population data based on the Southwest region of Taiwan (hence the abbreviation sw in the filenames). This is identified in the file as a village (number 1) with exposure equal to 0.

Each file has the following variables:

village id (coded 1, 2, etc)
median arsenic level for the village
midpoint of the agegroup (in 5 year increments)
person years at risk in that agegroup
number of cancers in that agegroup

To illustrate, here are the first few lines of the male bladder cancer data from the village exposed to 10 ppb. We have coded this village as village number 2:

2	10	22.5	1128	0
2	10	27.5	634	0
2	10	32.5	389	0
2	10	37.5	313	0
etc				

So, for example village 2, whose well had a median of 10ppb had 0, this says that there were 0 bladder cancers among the 1128 person years at risk in the agegroup 20 to 25.

The final agegroup, indicated at 82.5, refers to the group of individuals older than age 80.

The files $usdth\_mlun, usdth\_mblad, usdth\_flun$ and $usdth\_fblad$ contain the numbers of deaths from male and female lung and bladder cancer in 1996. The files $usdth\_mtot$ and $usdth\_ftot$ contain the numbers of deaths from all causes in the United States, in 1996. The naming conventions are as above. The files $usmalepyr.txt$ and $usfemalepyr.txt$ are the person years at risk in the same year. These data will be needed for calculating benchmark doses.

The file $wells.txt$ contains the individual arsenic measurements of the wells in the 42 villages. There is one row of data for each measured well. The first line of the file lists the variables which are:

village	Village number
nwell	number of wells in the village
conc	arsenic concentration (PPB) measured in the well

These data are illustrated in Figure 3 of Ryan (2004).

References

Morales KH, Ryan L, Kuo TL, Wu MM, Chen CJ (2000). Risk of internal cancers from arsenic in drinking water. Environmental Health Perspectives 108: 655-661.

NRC (1999). Arsenic in Drinking Water. Subcommittee on Arsenic in Drinking Water, National Research Council

NRC (2001). Arsenic in Drinking Water: 2001 Update. Subcommittee to Update the 1999 Arsenic in Drinking Water Report, Committee on Toxicology, Board on Environmental Studies and Toxicology, National Research Council.

Ryan LM (2004). Epidemiologucally based environmental risk assessment. Statistical Science (to appear).