

COMMENTS TO EPA SAB ARSENIC REVIEW PANEL

(September 12-13, 2005)

Cancer Risk from Low Arsenic Concentrations in U.S. Drinking Water is Likely Overstated

Kenneth G. Brown

Introduction

It is well known that even with controlled animal experiments using genetically bred rodents, low-dose extrapolation of cancer risk may vary widely for statistical models that fit the observed data about equally well. An unusually good example of a dose-response model fit to actual data is shown in Figure 1: the observations are close to the fitted curve, the dose values are highly reliable, there is virtually no potential for confounding. One might feel reasonably comfortable about the validity and reliability of predicting response at low dose. Epidemiological data are different, however, and none of the characteristics described above apply to dose-response modeling of the southwestern Taiwan data and estimation of cancer risk at low arsenic concentrations in drinking water. We cannot check for potential confounding or interaction of arsenic with other risk factors for bladder or lung cancer (e.g., smoking), but we can examine how well dose-response curves fit the observations, get some idea of the reliability of the exposure data used for dose, and examine the potential for the ecological nature of the exposure data to bias risk estimates, particularly at the low arsenic levels most relevant to the U.S.

Results of Morales et al. (2000)

A plot of results from an exhaustive dose-response modeling effort of the Taiwan data (Morales et al., 2000) is shown for male bladder cancer in Figure 2, for selected models with (a) no comparison population, (b) comparison with all of Taiwan, (c) and comparison with an adjacent region in southwestern Taiwan (after conversion to apply to the U.S.) The plots clearly indicate that the observed data points are too disperse for reliable prediction by any of the dose-response curves. Morales et al. point to the

ecological exposure data as a possible explanation, noting that they had assumed the same arsenic concentration for all persons in the same village but that individual exposures could vary widely within a village (p. 660). Similarly, both NRC reports (NRC1, 1999; NRC2, 2001), as well as the current EPA toxicological review (EPA, 2005), have made that same assumption, specifically using the median well test of a village as the “dose” for the whole village. The validity of that assumption and its potential to bias risk estimates is discussed next.

Arsenic exposure in the Taiwan database

The Taiwan database used to develop the dose-response relationship in the current EPA draft, aside from the age at death, is given in Table A10-1 of NRC1. Arsenic concentration from villages are reported as arsenic well tests, presumably on all wells used for drinking water and only those, although that is not clear. Of the 42 total villages, 20 had only one well test. The arsenic levels for the remaining 22 villages with more than one well test are plotted in Figure 3, with the median values indicated. To consider just one example, the last village listed in Figure 3, Village O-G, has five well tests, with arsenic concentrations of 10, 10, 30, 259, and 770 $\mu\text{g/L}$, for a median of 30 $\mu\text{g/L}$. There are 11 bladder or lung cancer deaths, a large number for the 10,000 person-years of the village. The current EPA dose-response analysis, as well as those of both NRC reports and Morales et al., effectively assume that the 11 cancer deaths occurred at a dose of 30 $\mu\text{g/L}$.

The range of well tests is not so extreme across all villages, but it is readily apparent from the figure that the example just described is not an isolated case. Some villages with only one well test raise a similar concern, which will be discussed further. First we attempt to identify villages with unusually high cancer mortality, independent of the median arsenic levels used to represent dose.

Age-standardized mortality ratios (SMRs, for ages 20+ years, age-standardized to the 1976 world population) of each village were calculated for bladder, lung, and liver primary cancer mortality, by gender. The number of SMRs that are sufficiently high to occur by chance with probability 0.15 or less (i.e., significant at $p = 0.15$) is referred to as the “score” for the village (0, 1, 2, ..., 6). The chance occurrence of a score of 3, 4, or 5

and above, is approximately 0.05, 0.007, and 0.0005, respectively. Thus, a score of 3 is marginally significant overall ($p = 0.05$), and scores of 4 and above are highly significant.

The score for each village is plotted against its median arsenic concentration (dose) in Figure 4. If the doses are correct, and the excess cancer mortalities are largely attributable to arsenic, then one would expect a high association between scores and doses. There are clearly some low scores at high doses and visa-versa which raise questions about the reliability of the data, particularly with ecological exposure data.

Dose-response for bladder cancer

It is generally assumed that a dose-response relationship is non-decreasing and one or more parametric models is fit to the data, as in Morales et al., with that assumption. The reliability of the data are questionable, however, so a flexible smoothing spline (4 degrees of freedom) was used here to “let the data do the talking”, i.e., to see what shape the dose-response curve would take and if the data are consistent with a non-decreasing response. Using primary bladder cancer for males and females as an example, the age-standardized mortality (for ages 20 years and older) was plotted against dose, by village. The results, with a spline function fit to the data, are displayed in Figures 5 and 6. Age-adjusted mortality for all of Taiwan is indicated on the figures for comparison. The median arsenic concentrations plotted on the x-axis in the figures are the median well tests used as doses in the dose-response analyses of NRC1, NRC2, Morales et al., and the current EPA draft analysis under review.

The villages could be categorized into a lower-dose region (0.10 – 0.126 $\mu\text{g/L}$) and an upper-dose region (0.256 $\mu\text{g/L}$ and above), with no villages in between (0.127- 255 $\mu\text{g/L}$). The lower and upper regions contain 18 (43%) and 24 (57%) villages, respectively, so both regions are well-represented. Figures 5 and 6 both suggest that a rather flat, or even downward dose-response relationship, over the lower dose region is most consistent with the data, with very high bladder cancer mortality predicted at zero dose (approximately 34 for females and 31 for males, age-adjusted, per 100,000 persons), about 24 and 12 times, respectively, the equivalent figures for all of Taiwan.

If those figures are close to correct, they would imply exceedingly high SMRs for bladder cancer in the study region, *after* adjusting for arsenic in drinking water. If there

are no credible explanations for that outcome, then one is led to suspect the accuracy of the data and the possibility that bladder cancer mortality may be overstated in part or all of the lower-dose region (there may be problems in the higher-dose region as well, which are not being addressed here). Focusing on the lower-dose region, exposure misclassification might lead to either under-statement or over-statement of dose. From a statistical perspective, the only villages that we can identify as possibilities are those with statistically high scores in the lower-dose region, which would suggest potential understatement of dose.

Of the 18 villages in the lower-dose region, there are six with scores of 4 or higher, as shown in Table 1 (overall p-value < 0.01). Two of those have a single well test (Villages 3-5 and 3-H, previously identified in Brown and Chen (1995)) , three have five or more well tests (Villages 0-G, 0-E, and 0-I), and one has two wells in a narrow range of 53-58 $\mu\text{g/L}$ (Village 3-L). When the spline was re-fit to the data with the six villages of scores 4 or higher omitted, the bladder cancer mortality predicted at zero dose was much more in line with the figures for all of Taiwan, for both genders, indicating a strong influence of those villages. Those results are not included, however, because selectively omitting those villages also produces a bias, not so much from restricting attention to the lower dose

region, but from only being able to identify villages in that region for which the dose appears questionably low.

Table 1. Results from Villages in Lower-Dose Region

Vill.Name	No.Wells	As Median	As Lowest	As Highest	Score	p-value
3-H	1	10	10	10	5	0.001
0-G	5	30	10	770	4	0.008
3-5	1	32	32	32	4	0.008
3-L	2	56	53	58	4	0.008
0-E	5	110	10	686	4	0.008
0-I	7	110	20	590	4	0.008
3-N	1	32	32	32	3	0.050
2-I	1	11	11	11	1	NS
4-7	1	42	42	42	1	NS
6-A	1	45	45	45	1	NS
0-J	2	50	20	80	1	NS
4-D	1	60	60	60	1	NS
3-P	1	65	65	65	1	NS
6-C	1	73	73	73	1	NS
4-8	1	80	80	80	1	NS
0-O	1	100	100	100	1	NS
4-N	2	123	73	172	1	NS
4-J	1	126	126	126	2	NS

The villages noted above are just six examples where it was possible to identify data of questionable reliability. Since all six villages have scores of 4 or higher with an overall significance level (p-value in Table 1) less than 0.01, the evidence of high cancer mortality rates in those villages is not limited to bladder cancer used for illustration in Figures 5 and 6. The three villages with more than several well tests (0-G, 0-E, and 0-I) have arsenic doses (under As Median in Table 1) of 30, 110, and 110 $\mu\text{g/L}$, respectively, but there is also potential for arsenic exposure at much higher concentrations, up to 770, 686, and 590 $\mu\text{g/L}$, respectively (under As Highest in Table 1). Any arsenic-induced cancers that occurred in those villages could easily be from exposures of several hundred $\mu\text{g/L}$, instead of the much lower median values assumed in dose-response analyses.

One might anticipate that exposure misclassification might be limited to villages with a wide range of well tests, but the three remaining villages identified in the six described above suggest otherwise. Villages 3-H and 3-5 have single well tests of 10 and 32 $\mu\text{g/L}$, and Village 3-L has two tests in a tight range of 53-58 $\mu\text{g/L}$. Continuing with

bladder cancer an example, the respective age-adjusted mortality rates (ages 20+ years, per 100,000) for Villages 3-H, 3-5, and 3-L, respectively, for females (males in parentheses) are 99.5 (17.6), 49.3 (42.1), and 55.2 (62.3), compared to Taiwanese-wide rates of 1.4 (2.9). Those excess bladder cancer mortality rates appear much too large to be explained by sampling error or to attribute to arsenic levels of only 10, 32, and 53-58 $\mu\text{g/L}$ in drinking water. If attributable to arsenic and even close to accurate, those figures would make it difficult to explain the negative findings of studies at similar arsenic levels in the U.S. Adjustments for water consumption rates, weight, etc. would be needed for more precise comparisons with the U.S., but it is clear that the outcomes are extreme and raise doubts about the reliability of the exposure data, aside from treating a whole village as if exposed to a common arsenic level in drinking water equal to the median arsenic test .

Figures 5 and 6 suggest unrealistically high bladder cancer mortality rates at low arsenic levels (even at zero) and a negligible (or negative) slope factor. This latter concern is one of the motivations for imposing a comparison population for dose-response analysis, which predictably produces supralinear response at low dose and a high slope factor.

Adding a comparison population

Referring to Figures 5 and 6, adding all of Taiwan as a comparison population is equivalent to adding a large number of hypothetical villages at the location of the box symbols in those figures with the Taiwanese-wide bladder cancer mortality rates. The more weight given to the comparison population at zero dose, the more the spline curve will bend downward toward the box symbol at zero dose. EPA (2005) uses a region neighboring the Taiwan study area as a comparison population. We do not have the age-standardized bladder cancer mortality rates for that region, so all of Taiwan is used for illustration instead. That should make little difference for this example. NRC2 cites evidence that SMRs for the area where arsenic is endemic based on regional population rates are similar in magnitude to SMRs based on rates for the national population (NRC2, p.191) and the results of Morales et al. for the two comparison populations appear indistinguishable in the plots shown in Figure 2. The correct dose to use for the comparison population is unknown; zero is used for this example.

Imposing a comparison population with such hugely different background mortality rates than predicted from the study region alone suggests a lack of credibility in the results from the Taiwan database. Sparsity of data is not a plausible explanation, because the Taiwan database is large, with considerable data in the lower-dose region (18 villages with a large number of person-years of exposure). Just repeating the same analysis of the Taiwan database with the addition of a comparison population, however, does not address the problem and it has a large impact on the outcome.

Adding the comparison population with bladder cancer mortality rates of 1.4 and 2.9, for females and males, respectively, essentially anchors the dose-response curves at or close to those values at zero dose (depending on how much weight is given to the comparison population). Since the study data are more consistent with much higher mortality rates at low dose, the effect on the dose-response curve is for it to increase upward sharply from its anchor. This effect is demonstrated in Figures 7 and 8, with the previous plots from Figures 5 and 6 without a comparison population included for comparison. Predictably, the effect is dramatic at the low dose end, which determines a slope factor and low-dose risk estimates for extrapolation to the U.S. population.

Conclusions and recommendations

It is apparent from Figures 5 and 6 that the Taiwan data indicate high bladder cancer mortality rates in the lower dose region, with flat or decreasing rates up to about 100 µg/L. A plausible explanation is dose- misclassification, and the potential for bias, from treating the median arsenic well test of a village as the dose for the whole village. Six villages in the lower-dose region (below 127 µg/L) where overall cancer mortality rates are significantly high ($p < 0.01$, based on scores taking into account SMRs of bladder, lung, and liver cancer, by gender) are identified as examples. These examples are just suggestive; it was not possible to identify villages where doses might be too high.

Three of the six villages have 5-7 well tests, are in the dose range 30-110 µg/L, but also contain wells that tested in the range 590-770 µg/L. Any cancer mortality due to arsenic could easily be associated with arsenic levels much higher than the median for the village. The remaining three villages are at doses of 10, 32, and 56 µg/L, based on a single well test at the lowest two doses and two tests in a tight range (53-58 µg/L) in the third.

Using bladder cancer as an example, the excess cancer mortality rates in those villages appear much too large to be explained by sampling error or to attribute to arsenic levels of only 10, 32, and 53-58 $\mu\text{g/L}$ in drinking water. If attributable to arsenic and even close to accurate, those figures would make it difficult to explain the negative findings of studies at similar arsenic levels in the U.S. Adjustments for water consumption rates, weight, etc. would be needed for more precise comparisons with the U.S., but it is clear that the outcomes are extreme and raise doubts about the reliability of the exposure data, aside from treating a whole village as if exposed to a common arsenic level in drinking water equal to the median arsenic test.

Imposing a comparison population, such as all of Taiwan used as an example, does not correct for a potential bias from dose-misclassification of villages in the Taiwan data. Instead, it shifts the potential bias from likely understatement to likely overstatement of the slope factor. This is because the Taiwan data alone estimate background bladder cancer mortality rates that are on the order of 10-25 times those of the comparison population. To accommodate the comparison population at zero dose and a much lower response level (i.e., cancer mortality rate) than indicated by the Taiwan study data, the smoothing spline has to drop sharply as dose decreases toward zero. The more the Taiwan data without a comparison group may tend to overstate risk at low dose, from exposure misclassification or otherwise, the steeper it would tend to make the fitted smoothing spline at low dose, and the higher the slope factor.

Given that EPA is required to set regulatory guidelines for arsenic concentrations in drinking water, and that it may have to rely on the Taiwan database in some form, it is recommended that the quality of the exposure data be closely examined, and that sensitivity analyses be conducted to identify villages that might be statistical outliers or otherwise particularly influential to risk estimation of cancer from arsenic at low levels common in the U.S. and elsewhere. The basic problem is that mortality data is known on an individual basis but exposure is only known at the village level. Further thought needs to be given to the validity and potential bias of representing dose for a village by the median arsenic well test and alternatives sought that might provide more reliable risk estimates for extrapolation of risk to the U.S.

References

- Brown, K., and Chen, C.-J. 1995. Observed dose-response for internal cancers and arsenic in drinking water in the blackfoot disease endemic region of Taiwan. In: *Arsenic Exposure and Health*, pp. 153-170, ed. W.R. Chappell, C. O. Abernathy, and C. R. Cothorn, Science and Technology Letters, Northwood, Middlesex, England.
- EPA (U. S. Environmental Protection Agency). 2005. *Toxicological Review of Ingested Inorganic Arsenic*. EPA. In support of summary Information on the Integrated Risk Information System (IRIS). (CAS No. 7440-38-2). An IRIS Review draft for the Science Advisory Board (SAB).
- Morales, K.H., L. Ryan, T.L Kuo, M.M. Wu, and C.J. Chen. 2000. Risk of internal cancers from arsenic in drinking water. *Environ. Health Perspect.* 108(7): 655-661.
- NRC1 (National Research Council). 1999. *Arsenic in Drinking Water*. Washington, DC: National Academy Press.
- NRC2 (National Research Council). 2001. *Arsenic in Drinking Water: 2001 Update*. Washington, DC: National Academy Press.

Figure 1. Example of a dose-response curve from a controlled animal experiment (rats exposed to hydrogen sulfide for 4 hours at various concentrations).

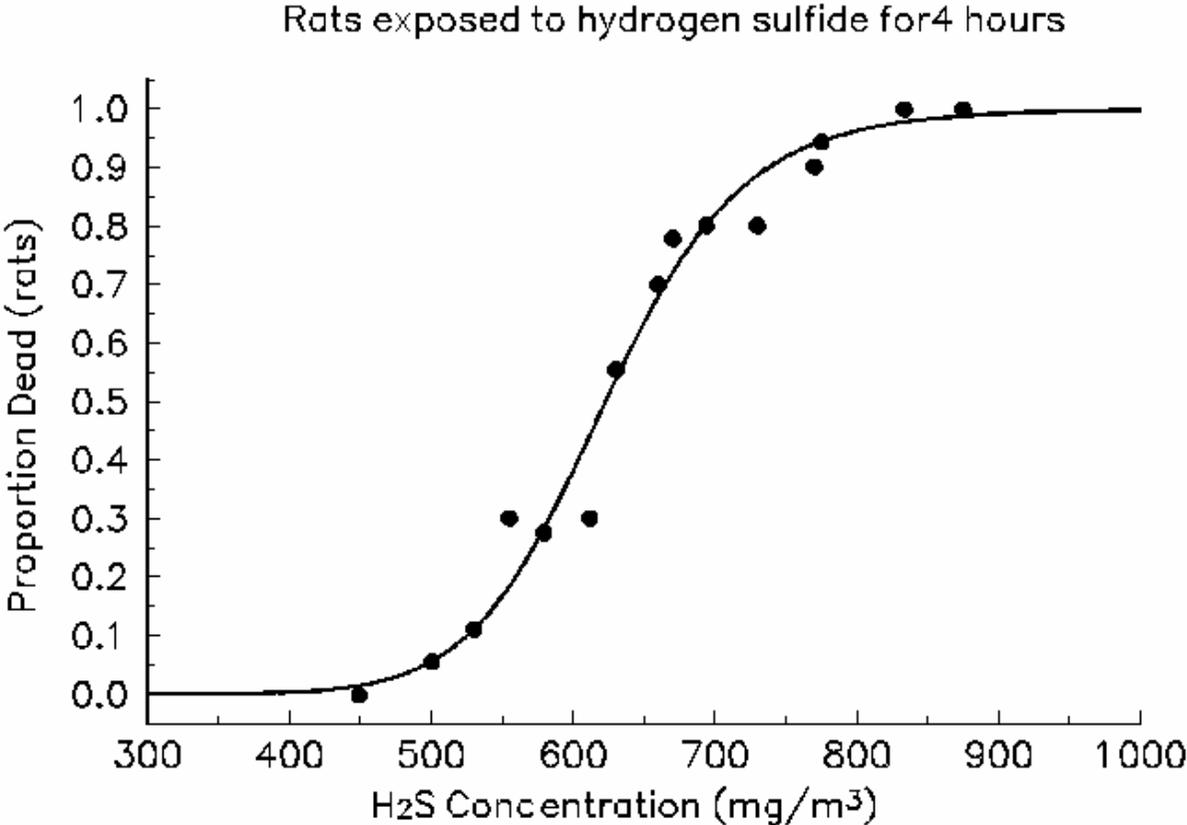


Figure 2. Male bladder cancer. Estimated lifetime death risk over background rates in Taiwan (a) without comparison population, (b) with Taiwanese-wide comparison population, (c) with Southwestern Taiwanese region comparison population. Reprinted from Morales et al. (2000) (With permission of Environ. Health Perspect).

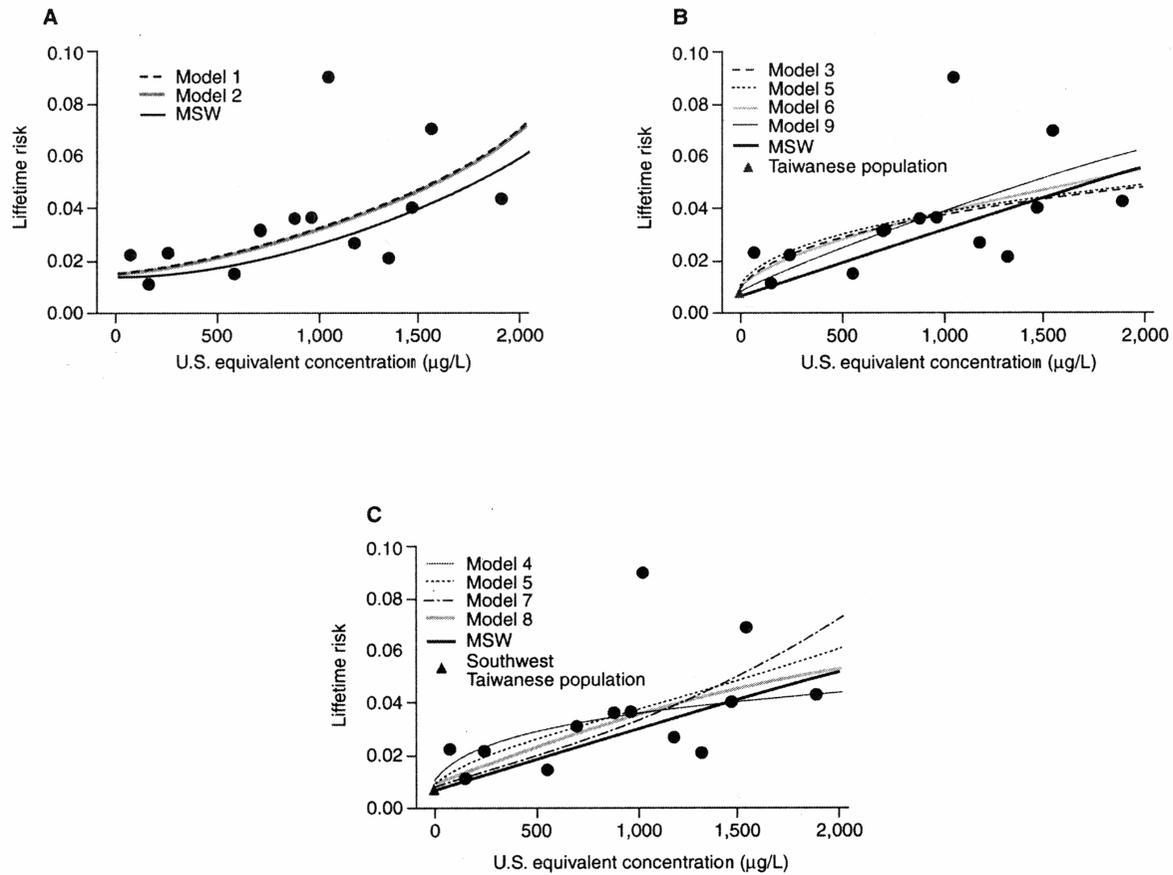


Figure 3. Arsenic well tests for villages with multiple well tests. Cross marks are at medians. (NRC1, 1999, Table A10-1).

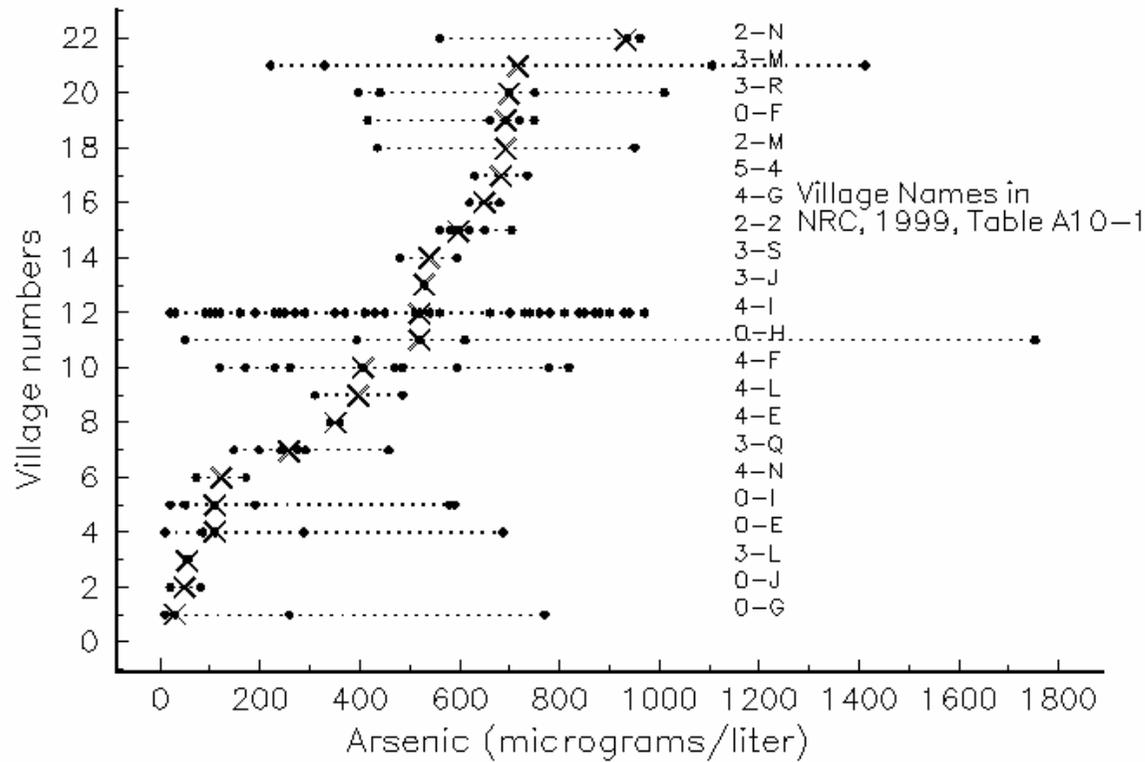


Figure 4. Significance of age-standardized mortality ratios for several cancers by dose (median arsenic concentration in well tests).

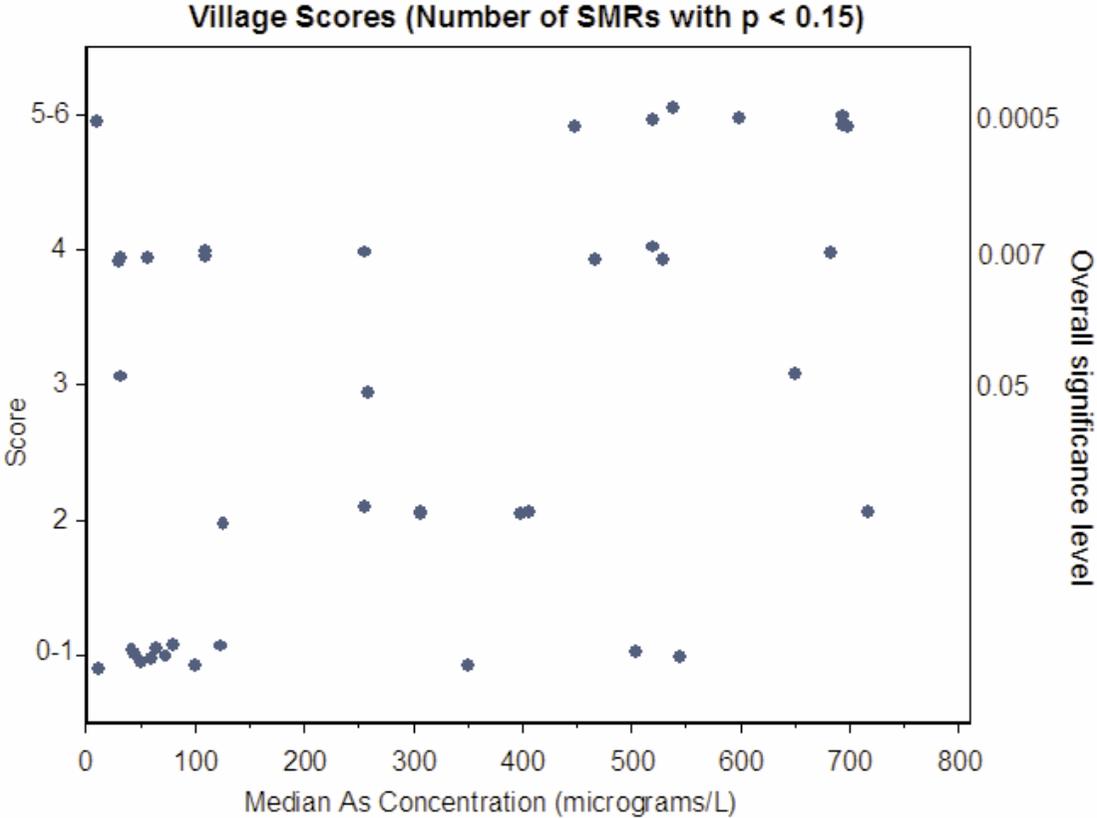


Figure 5. Dose-response for female age-adjusted bladder cancer mortality in Taiwan data, fit by smoothing spline without comparison population.

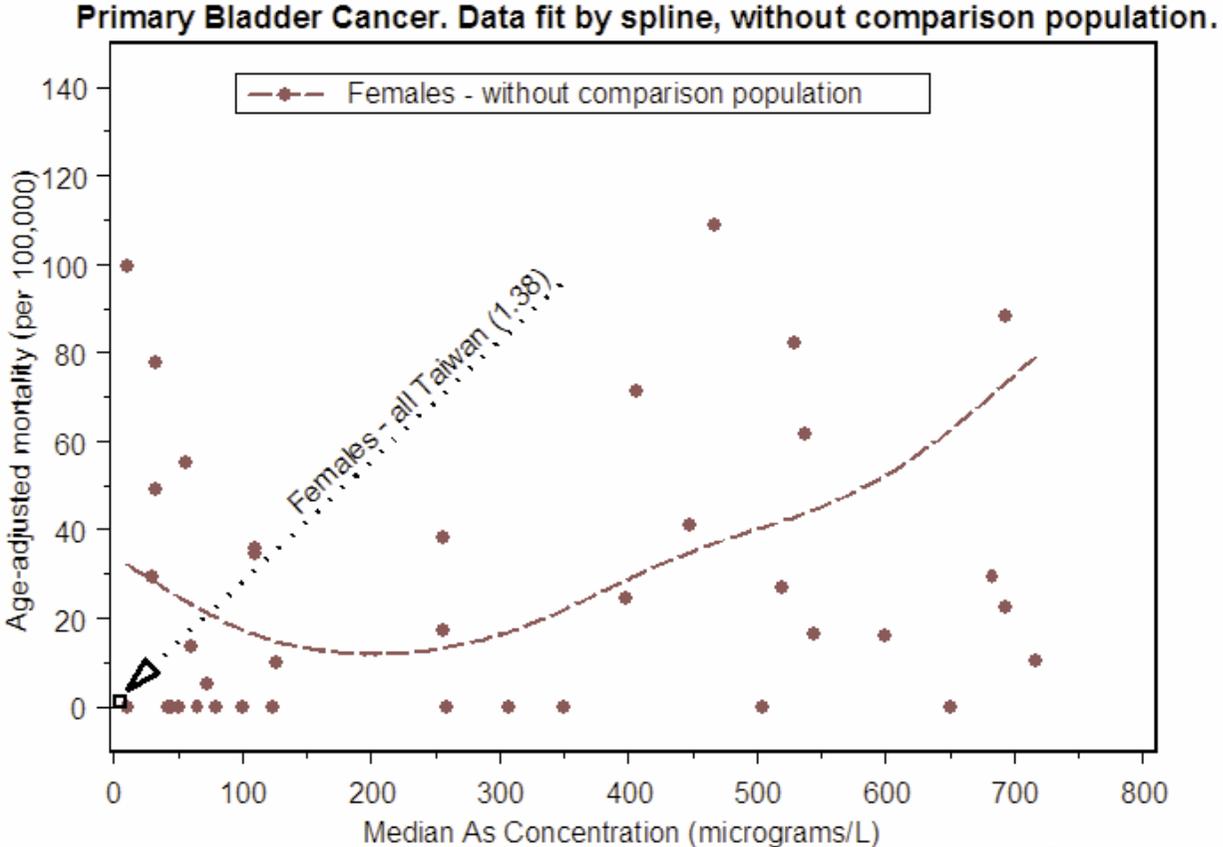


Figure 5.

Figure 7. Dose-response for female age-adjusted bladder cancer mortality in Taiwan data, fit by smoothing spline, with and without Taiwanese-wide comparison population.

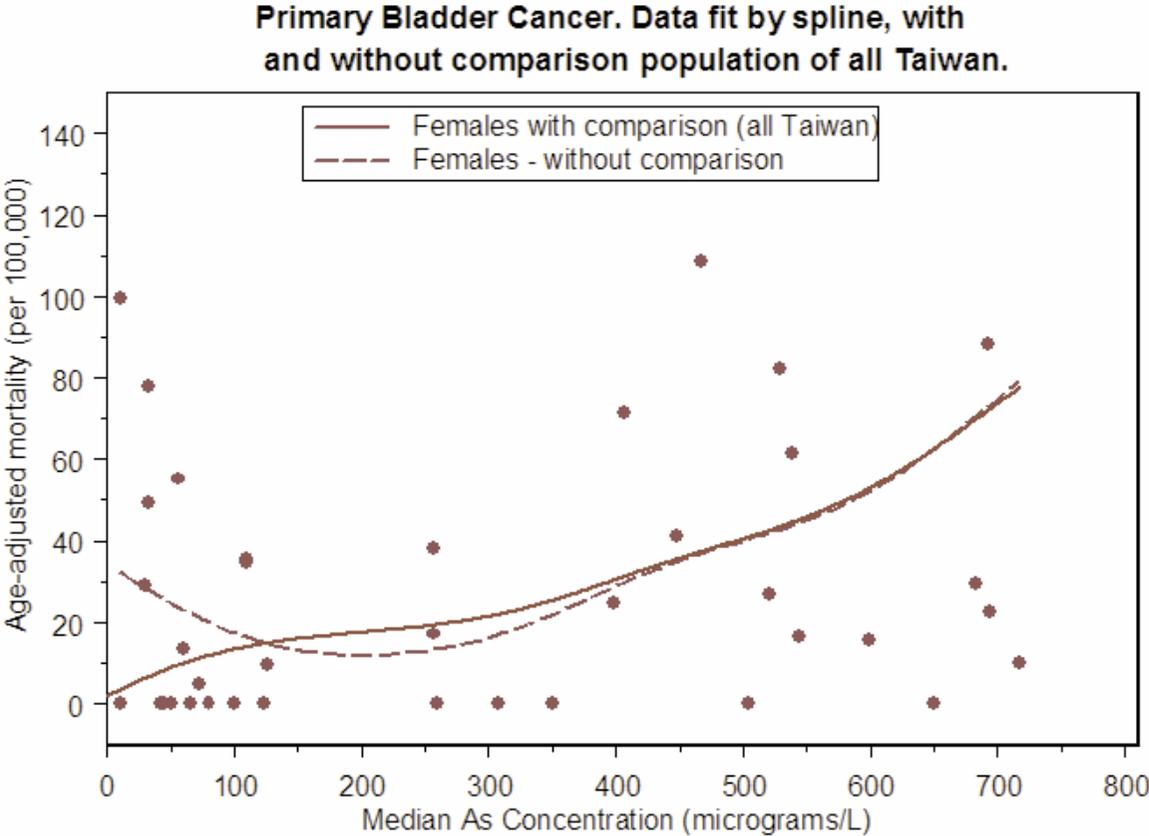


Figure 7.

Figure 8. Dose-response for male age-adjusted bladder cancer mortality in Taiwan data, fit by smoothing spline, with and without Taiwanese-wide comparison population.

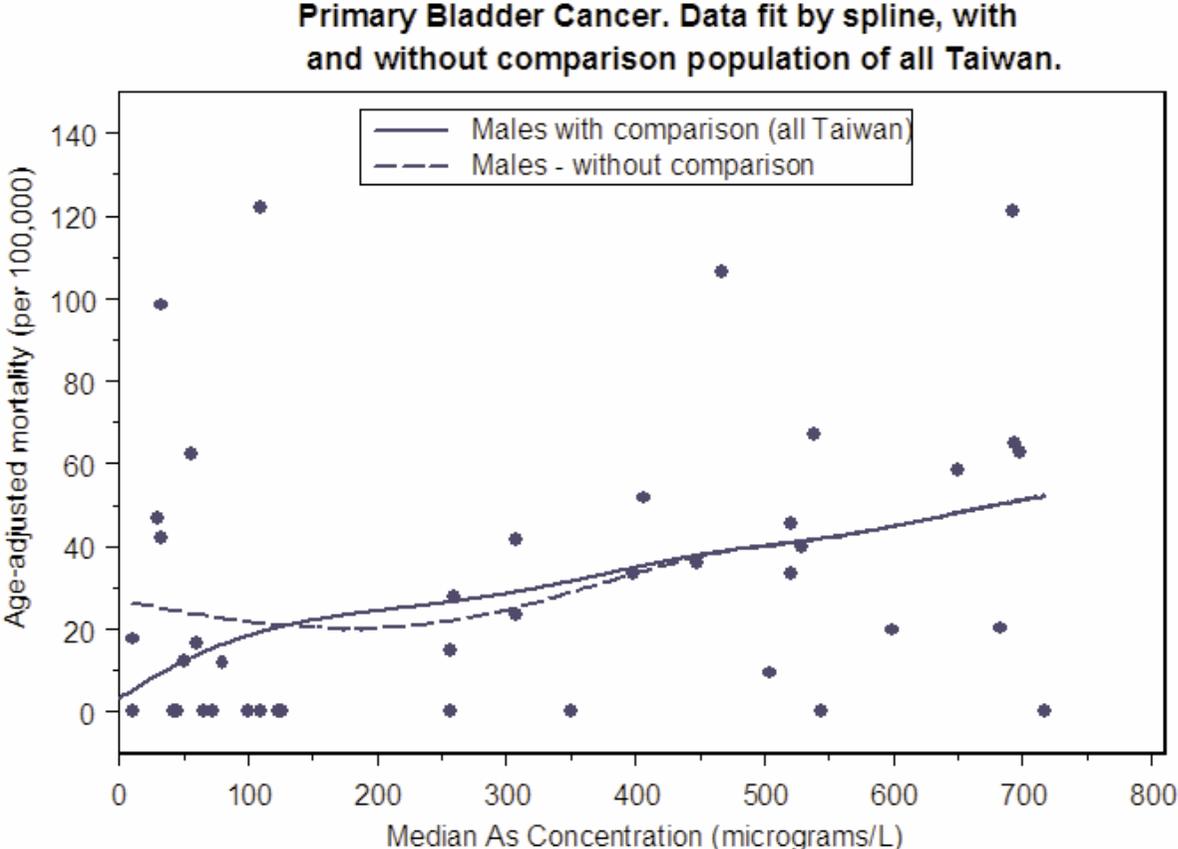


Figure 8.