



E X T E R N A L M E M O R A N D U M

TO: Scientific Advisory Board Members
 Genevieve Matanoski, M.D., Dr.PH., Chair
 Thomas Miller, U.S. Environmental Protection Agency

FROM: Pamela Mink, Ph.D., M.P.H.

DATE: October 20, 2005

PROJECT: BE02916.001 0101

SUBJECT: Follow-up Comments to the September 12–13 Meeting (Question C2: Selection of Data for Dose-Response Assessment)

The Science Advisory Board (SAB) was asked in its current charge to consider whether the data set from southwestern Taiwan remains the “most appropriate choice” for estimating human cancer risks associated with inorganic arsenic in drinking water given the recent epidemiologic studies from U.S. and other populations with typically low-level exposures. Preliminary summary comments by a subgroup of SAB members indicated that the panel supports the U.S. Environmental Protection Agency’s (EPA’s) continued use of the southwestern Taiwan dataset for estimating cancer risk in humans. Although the designated leader of this subgroup suggested that additional data from Taiwan, Argentina, and Chile could be used to “scale” the data, the case-control studies conducted in the United States and other areas with low-level exposure to arsenic in drinking water were largely dismissed.

It is our position that the data from the epidemiologic case-control studies are informative and should not be dismissed, particularly without formal evaluation or quantification of any potential limitations or biases. In addition, limited statistical power and misclassification of exposure, the two points most frequently raised by the committee, are not likely to operate in the manner or direction assumed in the EPA written reports, or as expressed by some members of the SAB. This memorandum provides additional comments regarding the strength of the low-level arsenic studies, the potential impact of exposure misclassification, and the statistical power of the studies.

Strengths of Low-Level Arsenic Studies

First, we consider the strengths of these studies. The most consistent finding from the case-control studies is a pattern of relative risk (RR) estimates that are predominantly below 1.0 for increasing levels of exposure above the reference group in analyses of the never smokers (Table 1; corresponds to Table B-4 in Exponent [2005]). When the increasing exposure levels

are collapsed into one category and compared to the reference group (lowest exposure), the RR estimates are all below 1.0 (Table 2; corresponds to Table B-3 in Exponent [2005]). The meta-RR (mRR) for either characterization of the data is 0.8, with an upper bound of the confidence interval of 1.1 (Tables 1 and 2). The high p-values for heterogeneity indicate that these findings are robust. The findings of the meta-analysis of never smokers are summarized graphically in Figure 1 (corresponds to Figure 2 in Exponent [2005]).

Thus, if we were to apply criteria analogous to the Bradford-Hill guidelines (Hill 1965) to this group of studies, we would conclude that for never smokers the findings are consistent for similar exposures (in different populations and when different measures and/or methods for characterizing exposure are used). Exposures for cases and controls in these studies were generally below 200 ppb overall, with the vast majority of participants exposed at levels lower than 50 ppb. The data presented in Table 1 do not show a consistent, inverse dose-response pattern, although the RR estimates in the highest dose categories are below 1.0 and lower than those at other exposure levels (with the exception of Bates et al. 2004). The mRR of 0.8 indicates a 20 percent reduction in risk for those “exposed” to low levels of arsenic in drinking water compared to the reference group, which was below 1 ppb in at least three of the studies (Karagas et al. 2004; Bates et al. 2004; Kurttio et al. 1999); exposure was reported as cumulative dose or dose-years in other studies. It is important to note that an inverse association does not necessarily mean a protective association, just as not all positive associations are necessarily causal. To further evaluate the association between arsenic and bladder cancer, particularly with respect to the current maximum contaminant level (MCL) of 50 ppb, we conducted an analysis that attempted to re-categorize the data to compare two groups: > 50 ppb vs. < 50 ppb. The purpose of this analysis was to determine if there was evidence of a significant excess risk for those with exposures above the current MCL (but still below approximately 200 ppb) as compared to those with exposures below the current MCL. Because we did not have access to the original data, we had to use the cutpoints that were used in the published studies; thus we were unable to make this comparison as precise as we would have liked, and all of the cutpoints were below 50 ppb. Results of this analysis of never and ever smokers combined, including the cutpoints used, are shown in Table 3. The mRR was 1.1 (95 percent confidence interval: 0.85–1.40).

In most of the studies, attempts were made to use historical data on arsenic in community drinking water and/or to sample water from drinking water sources based on self-reported residential histories. These efforts help to ensure the temporal association of exposure preceding disease, and in some studies, allowed analyses to be conducted according to duration or latency. Karagas et al. (2004) measured toenail concentrations of arsenic. Several studies (e.g., Tsuji et al. in press; Hinwood et al. 2003) have reported that toenails do not provide a valid biological marker for the purposes of quantifying sources of arsenic exposure (e.g., soil) because of external contamination of the toenails, which can confound the results. Toenails, however, have been suggested to be a reliable integrative biomarker of elevated arsenic in drinking water (Karagas et al. 2000, 2001). Karagas et al. (2000) reported a correlation coefficient of 0.65 ($p < .001$) for toenail arsenic and arsenic levels in drinking water in the range of 1–100 ppb. For levels below 1 ppb, however, the correlation was only 0.08. Garland et al. (1993) evaluated the

reproducibility of arsenic levels in toenails and found good serial correlation over the study period ($r=0.54$; 6-year study period), presumably resulting from consistencies in the dietary intake of arsenic (including from drinking water) over time.

Our interpretation is that these results are consistent with a threshold or sublinear association (i.e., to the extent that below a certain level the slope decreases to the point that it can no longer be measured) between increasing arsenic exposure and cancer risk. This would be consistent with alternative models presented for the southwestern Taiwan data (Brown 2005; Lamm and Kruse 2005), from a United States ecological study of arsenic in drinking water and bladder cancer (Lamm et al. 2004; Lamm and Kruse 2005), as well as the mechanisms of inorganic arsenic toxicity and carcinogenicity that were discussed at the SAB meeting. Thus, the body of evidence from the case-control studies of low-level exposure to arsenic in drinking water provides a coherent data set, particularly for the never smokers. As discussed at the SAB meeting, findings for ever smokers were heterogeneous, perhaps because of limited data on smoking status (former and current smokers grouped together), duration, and intensity.

Exposure Misclassification

Some of the members of the SAB suggested that the results of all of the epidemiologic case-control studies of low-dose exposure to arsenic in drinking water and bladder cancer should be disregarded because of the presence of information bias as a result of exposure misclassification. It appeared that misclassification was assumed to be non-differential and would result in bias in the direction of the null (1.0). If this scenario is accurate, it is important to consider the implications of this for the results of the analyses of never smokers in the case-control studies. The results for this group are neither confounded by nor modified by the effects of smoking, and may give the best indication of the independent effects of arsenic. As described above and shown in Tables 1 and 2, and Figure 1, the pattern of relative risks in the individual studies indicated an *inverse association* between low levels of arsenic in drinking water and bladder cancer. Furthermore, the results of the meta-analysis were $mRR = 0.8$. Thus, if these results are, indeed, biased toward the null, then the “true” relative risk estimates would be more extreme, or farther away from 1.0 than 0.8 (or the odds ratios in the individual studies). For example, if we were able to correct for misclassification, the true relative risk could be 0.7 or 0.5. Although the confidence interval for the mRR included values between 1.0 and 1.1, it is possible, but not probable, that the true RR would be greater than 1.0 (after adjusting for bias due to non-differential misclassification of exposure).

It should not be assumed, however, that misclassification is necessarily non-differential in all of the studies under consideration. There was no formal evaluation of information bias due to misclassification, nor did the SAB recommend sensitivity analyses to evaluate the potential impact of this type of bias on the results of any of the case-control studies. This was somewhat surprising, given that quantitative assessment of bias has been given increasing attention in recent years by epidemiologists, and it is becoming increasingly common to see presentations of results of sensitivity analyses in published papers and at scientific meetings. Furthermore, the SAB recommended sensitivity analyses to evaluate the impact of model choice, the values of

drinking water intake used, the dietary arsenic intake levels used, and other issues where there is uncertainty. Thus, it seems reasonable to conduct a more careful evaluation of potential bias in the case-control studies rather than simply dismissing them without a quantitative assessment of the potential impact of sources of bias (non-random error). Because epidemiologic studies are subject to bias from a variety of sources, it is important to evaluate how much bias is potentially present and how much of an impact it could likely have on the study results. These questions cannot be answered without a quantitative assessment, such as a sensitivity analysis.

In the case of cohort studies, it might be reasonable to assume non-differential misclassification, which, on average, will produce bias in the direction of the null. For case-control studies, however, there is a greater probability that exposure misclassification will vary according to case/control status. For example, if determining past exposure to arsenic in water depends on ascertaining a residential history, relying on proxies in situations where cases have died or may be too sick to respond to questionnaires may result in greater misclassification. Alternatively, cancer cases may be more motivated than population controls to take the time to provide a complete and accurate residential history. In addition, it has been shown that non-differential measurement error in a continuous variable may give rise to differential misclassification when that variable is categorized (Flegal et al. 1991). The main point is that non-differential misclassification should not be automatically assumed.

Statistical Power

The other main limitation raised by EPA and members of the SAB regarding the epidemiologic case control studies cited was limited statistical power because of small sample size. The Exponent meta-analysis addressed this potential problem by improving the statistical power of the analyses and the precision of the estimates. The power of the meta-analysis was presented in the original report (Exponent 2005; Table 3). Briefly, we had approximately 80 percent power to detect an mRR of 1.4 for ever and never smokers combined, and 80 percent power to detect an mRR of 1.5 for never smokers. Given that results for never smokers were consistently in the opposite direction (i.e., less than 1.0), and that ruling out a positive association of a certain magnitude seems more important than establishing the statistical significance of an inverse association, we conducted a *post hoc* analysis of the power of the meta-analysis, based on the results for the never smokers. These results are shown in Table 4. Thus, if the true RR for low-level exposure to arsenic (vs. lowest exposure category or <1 ppb) and bladder cancer is 1.1, then the probability of observing a mRR of 0.8 (or more extreme) is 4 percent. Similarly, if the true RR is 1.0, the probability of observing an mRR of 0.8 (or more extreme) is 11 percent. Based on this analysis, it does not appear likely that the true RR is greater than 1.0.

Finally, with respect to statistical power and sample size, we are in agreement with the limitations and cautions presented by Checkoway et al. (2004). These limitations are as follows:

- Power and sample size calculations are “based on the presumption that the purpose of the study is to make a decision solely on the bases of the

information obtained by the study, whereas in practice the study findings are usually evaluated in the context of the findings of previous studies” (e.g., all of the case-control studies of low-level arsenic exposure and bladder cancer).

- These calculations “assume that the purpose of the study is simply to distinguish between two, and only two, competing hypotheses: the null and a single alternative. In practice, this is rarely why epidemiologic studies are performed.”
- Power calculations “depend on an arbitrary definition of ‘statistical significance’ (the choice of the alpha error rate...) which is increasingly discouraged in epidemiology, in favor of a ‘weight of evidence’ approach to data interpretation.”
- “The choice of beta error rate is also arbitrary.”
- “The choice of the alternative value of the relative risk is often little more than a guess.” (pp. 86–87)

To use statistical power as a “litmus test” for considering or dismissing a study or a group of studies will inevitably result in dismissing information that is potentially informative and useful. If the data were indicative of an increased risk, then the ability to distinguish the RR from 1.0 and to estimate the RR precisely would be more important. But, in the case of low-level arsenic exposure and bladder cancer, the weight of evidence does not even suggest that the risk is significantly increased for never smokers.

In our opinion, the major limitation of the epidemiologic case-control studies is that they do not provide data across a wide range of exposure levels. Their strength, however, is that they do provide data on the relative risks of bladder cancer for humans exposed at low levels, typical of the levels in the United States. Thus these data are relevant and applicable to the question before the SAB, specifically, which data are appropriate for estimating cancer risk in humans? We submit that these data, in combination with other data, are appropriate and worthy of consideration by the SAB. Ignoring these data, while considering only data from an ecological study with obvious limitations from a population that has unique characteristics and is clearly not representative of the United States population, is inconsistent with good scientific practice.

Although the data from these case-control studies cannot by themselves provide a basis for dose-response modeling, because of lack of data at higher exposure levels, neither can the southwestern Taiwan data serve this purpose on their own. The case-control studies can be used to validate dose-response models derived from other data sets, as we have illustrated with the application of the results of the meta-analysis (Exponent 2005). If the dose-response curves cannot accommodate the results of epidemiologic case-control and cohort studies conducted in low exposure populations, then the validity of the models for accurately estimating cancer risk in humans is questionable.

References

Bates MN, Rey OA, Biggs ML, Hopenhayn C, Moore LE, Kalman D, Steinmaus C, Smith AH. Case-control study of bladder cancer and exposure to arsenic in Argentina. *Am J Epidemiol* 2004; 159(4):381–389.

Brown KG. Cancer risk from low arsenic concentrations in U.S. drinking water is likely overstated. Comments to EPA SAB Arsenic Review Panel. September 12–13, 2005. Available at: http://www.epa.gov/sab/pdf/miller_and_twc2_arsenic_review_panel.pdf

Checkoway H, Pearce N, Kriebel D. *Research methods in occupational epidemiology*, 2nd edition, New York: Oxford University Press, 2004.

Exponent. Epidemiologic studies of low-level arsenic exposure in drinking water and bladder cancer: a review and meta-analysis (unpublished technical report). Exponent, Inc., Washington, DC, 2005 Available at: http://www.epa.gov/sab/pdf/exponent_and_wpsc3_arsenic_review_panel.pdf

Flegal KM, Keyl PM, Nieto FJ. Differential misclassification arising from nondifferential errors in exposure measurement. *Am J Epidemiol* 1991; 134(10):1233–1244.

Garland M, Morris JS, Rosner BA, Stampfer MJ, Spate VL, Baskett CJ, et al. Toenail trace element levels as biomarkers: Reproducibility over a 6-year period. *Cancer Epidemiol Biomarkers Prev* 1993; 2(5):493–497.

Hill AB. The environment and disease: Association or causation? *Proc R Soc Med* 1965; 58:295–300.

Hinwood AL, Sim MR, Jolley D, de Klerk N, Bastone EB, Gerostamoulos J, Drummer OH. Hair and toenail arsenic concentrations of residents living in areas with high environmental arsenic concentrations. *Environ Health Perspect* 2003 111(2):187–193.

Karagas MR, Tosteson TD, Blum J, Klaue B, Weiss JE, Stannard V, Spate V, Morris JS. Measurement of low levels of arsenic exposure: a comparison of water and toenail concentrations. *Am J Epidemiol* 2000 152(1):84–90.

Karagas MR, Le CX, Morris S, Blum J, Lu X, Spate V, Carey M, Stannard V, Klaue B, Tosteson TD. Markers of low level arsenic exposure for evaluating human cancer risks in a U.S. population. *Int J Occup Med Environ Health* 2001; 14(2):171–175.

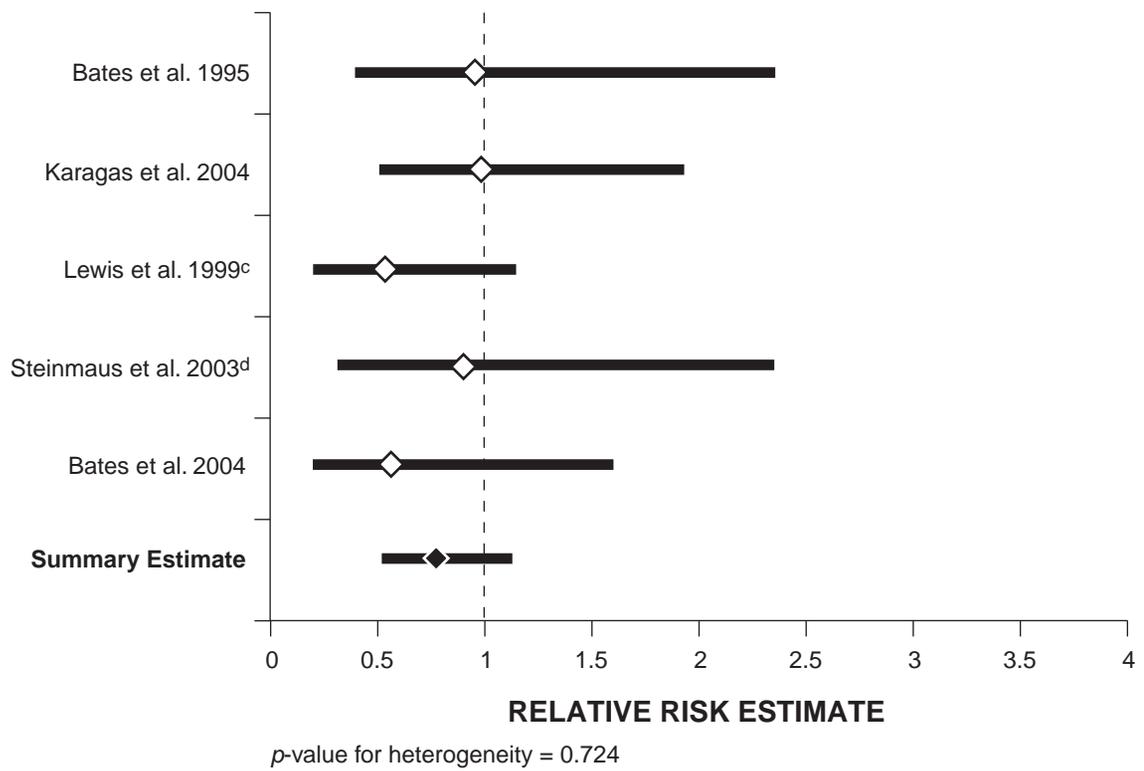
Karagas MR, Tosteson TD, Morris JS, Demidenko E, Mott LA, Heaney J, Schned A. Incidence of transitional cell carcinoma of the bladder and arsenic exposure in New Hampshire. *Cancer Causes Control* 2004; 15(5):465–472.

Kurttio P, Pukkala E, Kahelin H, Auvinen A, Pekkanen J. Arsenic concentrations in well water and risk of bladder and kidney cancer in Finland. *Environ Health Perspect* 1999; 107(9):705–710.

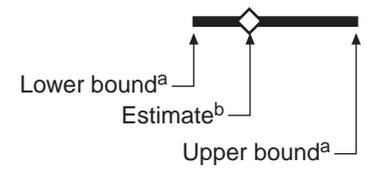
Lamm SH, Kruse MB. Arsenic ingestion and bladder cancer mortality—what do the dose response relationships suggest about mechanism? *Hum Ecol Risk Assess* 2005; 11:433–450.

Lamm SH, Engel A, Kruse MB, Feinleib M, Byrd DM, Lai S, Wilson R. Arsenic in drinking water and bladder cancer mortality in the United States: An analysis based on 133 U.S. counties and 30 years of observation. *J Occup Environ Med* 2004; 46(3):298–306.

Tsuji JS, Van Kerkhove MD, Kaetzel RS, Scrafford CG, Mink PJ, Barraj LM, Crecelius EA, Goodman M. Evaluation of exposure to arsenic in residential soil. In press, *Environ Health Perspect*.



LEGEND



Notes:

- ^a Mid- p confidence limits.
- ^b Individual crude relative risk estimates based on recalculated collapsed exposure category data; calculated using *Episheet*.
- ^c Cohort presumed to be nonsmokers, combined men and women and recalculated SMR.
- ^d Analysis based on 40-year lag.

Figure 1. Arsenic exposure and risk of bladder cancer among NEVER SMOKERS: Study-specific collapsed exposure categories

Table 1. Arsenic exposure and risk of bladder cancer among NEVER SMOKERS: Study-specific exposure categories analyzed

Study (Data Source)	Study Country	Study Design	Study-Specific Arsenic Exposure Range	Observed Cases or Deaths Among "Exposed"	Type of Relative Risk Estimate	Estimate	Lower Bound	Upper Bound	Relative Weight of Study
Bates et al. 1995 (Table 3)	U.S.	Case-control	Cumulative dose (mg):		OR				
			< 19 (referent)	10					
			19 to < 33	10		1.09	0.4	3.1	0.08
			33 to < 53	7		0.68	0.2	2.3	0.06
			≥ 53	4	0.53	0.1	1.9	0.05	
Karagas et al. 2004 (Table 2)	U.S.	Case-control	Toenail concentrations (mcg/g):		OR				
			0.009 to 0.059 (referent)	15					
			0.060 to 0.086	20		0.85	0.38	1.91	0.13
			0.087 to 0.126	22		1.18	0.53	2.66	0.13
			0.127 to 0.193	11		1.10	0.42	2.90	0.09
			0.194 to 0.277	3		0.49	0.12	2.05	0.04
			0.278 to 0.330	0	--	--	--		
			0.331 to 2.484	0	--	--	--		
Lewis et al. 1999 ^a (Table 4)	U.S.	Cohort	All exposure groups:		SMR				
			< 1,000 ppb-years to						
			≥ 5,000 ppb-years:						
			Men	3	0.42	0.08	1.22	0.08	
			Women	2	0.81	0.10	2.93	0.05	
Steinmaus et al. 2003 ^b (Table 4)	U.S.	Case-control	Cumulative dose (mg):		OR				
			< 6.4 (referent)	23					
			6.4 to 82.8	3		2.65	0.49	14.24	0.03
			> 82.8	3	0.50	0.12	2.05	0.04	
Bates et al. 2004 (Table 4)	Argentina	Case-control	Fluid intake adjusted exposure index (μg/L):		OR				
			0 to 1.0 (referent)	9					
			1.1 to 17	4		0.36	0.1	1.7	0.04
			18 to 80	10		0.95	0.2	3.9	0.04
			> 80	6		0.59	0.1	2.9	0.03

Table 1. (cont.)

Study (Data Source)	Study Country	Study Design	Study-Specific Arsenic Exposure Range	Observed Cases or Deaths Among "Exposed"	Type of Relative Risk Estimate	Estimate	Lower Bound	Upper Bound	Relative Weight of Study
Kurttio et al. 1999 ^c (Table 7)	Finland	Case cohort	Concentration of arsenic in water ($\mu\text{g/L}$):		RR				
			< 0.1 (referent)	8					
			0.1 to 0.5	4	0.95	0.25	3.64	0.05	
			≥ 0.5	5	0.87	0.25	3.02	0.06	
Summary Relative Risk = 0.808						95% CI: 0.603–1.083			
<i>p</i> -value for Heterogeneity =						0.937			

Note: OR - odds ratio
 RR - relative risk
 SMR - standardized mortality ratio

^a Cohort presumed to be nonsmokers.

^b Analysis based on 40-year lag.

^c Never or ex-smokers included in the analysis.

Table 2. Arsenic exposure and risk of bladder cancer among NEVER SMOKERS: Study-specific collapsed exposure categories

Study	Study Country	Study Design	Study-Specific Arsenic Exposure Range	Observed Cases or Deaths Among "Exposed"	Type of Relative Risk Estimate	Estimate ^a	Lower Bound ^b	Upper Bound ^b	Relative Weight of Study
Bates et al. 1995	U.S.	Case-control	Cumulative dose: 19 to \geq 53 mg vs. < 19 mg	21	OR	0.9375	0.3861	2.3577	0.17
Karagas et al. 2004	U.S.	Case-control	Toenail concentrations: 0.06 to 2.284 mcg/g vs. < 0.059 mcg/g	56	OR	0.9688	0.5020	1.9297	0.31
Lewis et al. 1999 ^c	U.S.	Cohort	All exposure groups: < 1,000 ppb-years to \geq 5,000 ppb-years	5	SMR	0.5155	0.1889	1.143	0.23
Steinmaus et al. 2003 ^d	U.S.	Case-control	Cumulative dose: 6.4 to > 82.8 mg vs. < 6.4 mg	6	OR	0.8889	0.3031	2.3520	0.16
Bates et al. 2004	Argentina	Case-control	Fluid intake adjusted exposure index: 1.1 to > 80 μ g/L vs. \leq 1.0 μ g/L	20	OR	0.5420	0.1873	1.5992	0.13
Summary Relative Risk = 0.763 95% CI: 0.519–1.120									
<i>p</i> -value for Heterogeneity =						0.724			

Note: OR - odds ratio
SMR - standardized mortality ratio

^a Individual crude estimates based on re-calculated collapsed exposure category data; calculated using *Episheet*.

^b Mid-*p* confidence limits.

^c Cohort presumed to be nonsmokers, combined men and women and recalculated SMR.

^d Analysis based on 40-year lag.

Table 3. Arsenic exposure and risk of bladder cancer among EVER and NEVER SMOKERS: Categories re-defined for analyses restricted to studies of arsenic concentration in drinking water

Study	Study-Specific Arsenic Exposure Range	Observed Cases or Deaths among "Exposed"	Relative Risk Estimate	Lower Bound	Upper Bound	Relative Weight of Study
Karagas et al. (2004) ^a	Estimated drinking water concentrations based on toenail concentrations					
	< 36 µg/L (referent)	366	0.884	0.474	1.610	0.17
	≥ 36 µg/L	17				
Steinmaus et al. (2003)	Arsenic concentrations					
	< 20 µg/L (referent)	160	1.460	0.790	2.670	0.17
	≥ 20 µg/L	21				
Bates et al. (2004)	Fluid intake adjusted exposure index					
	< 18 µg/L (referent)	55	1.150	0.683	1.940	0.23
	≥ 18 µg/L	59				
Michaud et al. (2004)	Estimated drinking water concentrations based on toenail concentrations					
	< 10 µg/L (referent)	209	1.020	0.701	1.495	0.43
	≥ 10 µg/L	71				

Note: Summary relative risk = 1.087

95 percent confidence interval: 0.847–1.396

p-value for heterogeneity = 0.676

^a Estimates of drinking water concentrations in Karagas et al. (2004) based on correlations with toenail concentrations (Karagas et al. 2001)

Table 4. Post-hoc evaluation of statistical power given results of meta-analysis for never smokers (mRR = 0.8; 95 percent confidence interval: 0.52–1.12)

Hypothesis: trueRR	Z	P-value
1.0	-1.24	0.107
1.1	-1.72	0.043
1.3	-2.55	0.005