Discontinuity in the cancer slope factor as it passes from high to low exposure levels – arsenic in the BFD-endemic area

Steven H. Lamm\textsuperscript{a, b, c, *}, Shayhan Robbins\textsuperscript{a}, Rusan Chen\textsuperscript{d}, Jun Lu\textsuperscript{e}, Brian Goodrich\textsuperscript{a}, Manning Feinleib\textsuperscript{b}

\textsuperscript{a} Center for Epidemiology and Global Health, CEOH, LLC, Washington, D.C., USA
\textsuperscript{b} Johns Hopkins University-Bloomberg School of Public Health, Baltimore, MD, USA
\textsuperscript{c} Georgetown University School of Medicine, Washington, D.C., USA
\textsuperscript{d} Georgetown University Graduate School, Washington, D.C., USA
\textsuperscript{e} American University Department of Mathematics, Washington, D.C., USA

**A R T I C L E   I N F O**

Article history:
Received 21 May 2014
Received in revised form 18 August 2014
Accepted 31 August 2014
Available online 18 September 2014

Keywords:
Arsenic
Dose-response
Lung cancer
Bladder cancer
Southwest Taiwan
Low level exposure

**A B S T R A C T**

Background: The ingestion of inorganic arsenic causes bladder and lung cancers demonstrably at >400–500 \(\text{ug}/\text{L}\) but questionably below 100–200 \(\text{ug}/\text{L}\). Using the standard 42-village cancer mortality dataset from the Blackfoot-disease (BFD) endemic area of southwest Taiwan (Wu et al., 1989), we examined the risk from low exposures by excluding the high exposures.

Method: Poisson regression analyses with the sequential removal of the highest exposure village have been performed using the mean, median, or maximum village well water arsenic level and demonstrated graphically.

Results: Risk estimates are positive when villages with exposures of 200–400 \(\text{ug}/\text{L}\) are included and significantly so when villages with >400 \(\text{ug}/\text{L}\) are included. Risk estimates for exposures below 100 \(\text{ug}/\text{L}\) are negative but rarely significantly so. The inflection point where the slope is no longer positive occurs in the range of 100–200 \(\text{ug}/\text{L}\), depending upon whether the exposure metric used is the mean, the median or the maximum.

Conclusion: There is a discontinuity in the cancer slope factor or risk from arsenic exposure that occurs in the range of 100–200 \(\text{ug}/\text{L}\). Above these levels, there are significantly positive risks, while below these levels there are not. The analysis reveals within this dataset an intrinsic non-linearity in the cancer risk. The literature speaks to this discontinuity, but this is the first demonstration within a single dataset that shows the discontinuity across the full exposure range and where the low-dose data are not compromised with high-dose data.

© 2014 Elsevier Ireland Ltd. All rights reserved.

**Introduction**

“Dose makes the Poison” is the classic aphorism of Paracelsus, but whether this applies to carcinogenesis has been a longstanding issue. The standard default assumption is “No”, but at times the data seem to indicate an answer of “Yes”. Such may be the case with inorganic arsenic.

It is not enough to know that a substance has the potential to cause cancer without knowing at what dose levels that potential is demonstrated. It is clear that the ingestion of drinking water containing inorganic arsenic in the hundreds to thousands of micrograms per liter is a cause of skin, bladder, and lung cancers in humans (NRC, 1999). Less clear is whether it causes those cancers in the range of 100–200 \(\text{ug}/\text{L}\) or below 100 \(\text{ug}/\text{L}\), and whether it causes other cancers in humans. Cantor (2001) pointed out that “To date, most estimates of risk at low to moderate levels of exposure (<200 \(\text{ug}/\text{L}\)) have been based on extrapolation from ecological studies of populations exposed to much higher levels.” We propose here a methodology for separating out in a single study the risk from high exposure levels in order to examine the risk at lower exposure levels.

Few studies have sufficient data to examine the risk pattern in both the exposure range in which the risk is clear and the exposure range in which it is not clear. The internal (bladder and lung) cancer mortality study (Wu et al., 1989; Morales et al., 2000) from southwest Taiwan is one that does. This study from the Blackfoot-
disease (BFD) endemic area in southwest Taiwan served as the data base for cancer risk assessments for the ingestion of inorganic arsenic in drinking water by the U.S. Environmental Protection Agency (U.S. EPA, 2001, 2005, 2010), the National Research Council/National Academy of Sciences (NRC, 1999, 2001) and others (Morales et al., 2000; Lamm and Kruse, 2005; Lamm et al., 2003, 2006, 2007, 2013). These various analyses have differed in the use of reference populations (southwest Taiwan, all Taiwan, and none), analytic model (Weibull, SMR, and Poisson), and exposure assumptions (body weight, liters/day, non-water arsenic, etc.). Previously, the US EPA (U.S. EPA, 1984, 1998) had used the skin cancer prevalence survey (Tseng et al., 1968) from the same BFD-endemic area for that purpose.

All these conditions were attributed to the use of high-arsenic artesian wells that tapped a deep aquifer rather than to the shallow wells and or the surface waters. All reported cases of BFD cases have been traced back to the consumption of artesian well water in this area (Tseng, 1989).

Blackfoot disease, first reported in the late 1950s, is a peripheral vascular disease uniquely found in a limited area of southwest Taiwan which pathologically resembles thromboangitis obliterans (ChV and Blackwell, 1968). The onset is insidious, progressing from numbness, coldness, and blanching of the toes and fingers to excruciating pain with ascending gangrene. It causes a dry gangrene, with the affected extremities turning black and becoming mummified, often with spontaneous or auto-amputation of distal limbs and digits. Blackfoot disease has only been found in this area of southwest Taiwan. It has not been found in any of the other areas of the world where arsenic disease have been investigated.

Early research (Chen and Wu, 1962; Chen et al., 1962; ChV and Blackwell, 1968) showed it to be associated with the use of artesian well water that had high levels of arsenic. The artesian well water contained high levels of arsenic (greater than 400–500 ug/L) and organic matter (i.e., humic acid, humic substances, or fluorescent substances). Studies in the BFD-endemic area (Chen et al., 1962) showed the arsenic levels in the artesian wells (median 780 ug/L; range of 350 ug/L–1140 ug/L) to be far greater than those in the shallow well waters had (median 40 ug/L; range <10 ug/L–300 ug/L). Residents of southwest Taiwan had begun to use artesian wells in 1905–1920, as the local shallow wells had become infiltrated with salt water from the South China sea, and continued to do so until the mid-1950s, when in 1955–56 they were replaced with piped tap water. Southwest Taiwan is the only place in the world where arsenic disease has been studied and Blackfoot disease has been found.

The BFD-endemic area has a long publication history in the public health literature. Yeh (1963) reported a high incidence of arsenical skin cancer and Tseng et al. (1968) demonstrated an ascending gradient for skin cancer prevalence and the well water arsenic concentration(<300 ug/L, 300–600 ug/L, and >600 ug/L as median village well water arsenic levels). Subsequent risk assessments (U.S. EPA, 1984, 1998) were based on the graphs and tables in Tseng et al. (1968) as the underlying data had been lost.

Chen et al. (1985) then reported on 1968–82 death certificates for internal cancers in 84 villages in four of the townsips in the BFD-endemic area demonstrating associations with township, by degree of BFD-endemicity, and by proportion of drinking water sources that were artesian wells. This study did not contain information on arsenic concentrations in the drinking waters. Wu et al. (1989) extended the Chen et al. (1985) to contain cancer mortality data (1973–1986) on the 42 villages (24 from Chen et al. (1985) and 15 additional) in 6 townships (4 from Chen et al. (1985) and 2 additional) for which there were individual well arsenic measurements from the Kuo (1968) BFD-endemic area well water survey of 1964–66. Wu et al. (1989) used the same three arsenic strata (cut at 300 and 600 ug/L) as did Tseng et al. (1968) and also showed dose-response relationships. Chen et al. (1988) showed that the rates in the study areas were greater than those for the general Taiwan population. Chen et al. (1992) added a strata at <100 ug/L and developed cancer potency indices of liver, lung, bladder, and kidney cancers using the Armitage–Doll multistage model. NRC (1999) [Table A10–1] published the well specific arsenic levels and village mortality data (bladder, lung, ad liver cancers), and Morales et al. (2000) incorporated them into their data and risk analyses using age-specific data released by EPA (U.S. EPA, 2004).

Arsenic has been discussed as a “threshold” carcinogen for which there is an exposure level below which either no positive dose-relationship, or no significantly positive dose-relationship, is observed. In the 1990s, arsenic was discussed as a “threshold” carcinogen based on the Tseng et al. (1968) study (Abernathy et al., 1996). Analyses of the Wu et al. (1989) study have raised similar discussion (Lamm and Kruse, 2005). They demonstrated that the Morales et al. (2000) SMR analysis of bladder cancer mortality showed no dose-relationship below 400 ug/L. Similarly, Lamm et al. (2006) demonstrated heterogeneity among the townships and examined in particular the townships whose data showed a significant dose-response with arsenic exposure (village median). They demonstrated in the linear regression of the SMRs for bladder and lung cancers combined that the no increased risk level (SMR = 100) was crossed at 151 ug/L for males and females combined. For each of the cancer subgroups – for male, female, or both; and for bladder cancer, lung cancer, or both – the regression crossed the no increased risk level at greater than 100 ug/L. These results contrast with the U.S. EPA (2010) [page F-5] conclusion that the cancer slope factor was significantly positive and robustly so across the arsenic exposure spectrum and that “no evidence was found that either a 400 ppb (ug/L) or 150 ppb represent “threshold” arsenic concentrations in drinking water below which cancer risks are not increased.”

We present here an analysis of the bladder and lung cancer mortality data for the 42 villages in the Wu et al. (1989) study in order to examine these conflicting conclusions. We use the same exposure assumptions as those in U.S. EPA (2010) and the same Poisson analytic methodology, but we limit the analytic set to the data from the 42 study villages.

We have added two analytic extensions in order to remove or diminish the influence of high exposures on the assessment of risk from lower exposures. We have developed a sequential reductive Poisson analysis where the cancer slope factor is recalculated following the sequential removal of the data from the highest exposure villages. The graphic presentation demonstrates the risk behavior across the full range as the influence of still higher levels of arsenic is eliminated.

Further, as use of the village median may hide the presence of high exposure arsenic well water sources, we have also replicated these analyses using both the mean and the maximum village well water arsenic levels as the summary exposure metric. This analytic process allows us to examine whether the carcinogenic risk estimate is continuous across all exposure levels or whether there is a discontinuity in the risk estimate and whether this is either metric dependent or analytic method dependent.

Method

Study population

The six-township BFD-endemic area in Chia-Yi and Tainan counties in southwest Taiwan was comprised of 126 villages. Well water arsenic levels were known for 42 of those villages. The adult (>20 years of age) resident population of these 42 villages comprised the study population. The years 1973–1986 comprised
the 14-year period of mortality observation. Age and gender-
specific person-years of observation data that had been released in
2004 (U.S. EPA, 2004) were used.

Exposure

The well water arsenic data came from the 1964–1966 survey of
the BFD-endemic areas (Xuo, 1968). Well water arsenic measurements
were made using either the silver diethylidithiocarbamate method or
the mercuric bromide stain method (AWWA, 1955). These methods
had standard deviations of 10 μg/L and 60 μg/L, respectively, for
measurements of a synthetic sample of 50 μg/L (AWWA, 1955). The
well water arsenic measurements for each village were listed in NRC
(1999) [Table A10-1] along with the village medians. The understanding
is that each measurement in the table represents a different village
well. We have calculated from that listing the mean and maximum for
each village.

Outcome

These analyses are based on the bladder and lung cancer mortality
records (death certificates) of the local household registration offices
for each study village during the fourteen year period 1973–1986. These
death certificates had been coded in Wu et al. (1989) to the
Eighth Revision of the International Classification of Disease [ICD-8]
(U.S. DHHS, 1968). Person-years of observation for each village had
been developed from mid-year population data abstracted by age and
sex from the demographic reports of the local registration offices. Age
and gender-specific cancer mortality had also been released in 2004
(U.S. EPA, 2004). The analyses presented here are for the combined
count of bladder and lung cancers deaths in each village. Other analyses
(Lamm et al., 2006) have included various groupings of the four
outcome sub-groups (male bladder, female bladder, male lung, and
female lung). Here we present the analyses of the combined counts as
they are more robust than for each of the sub-groups, with an average
of 11 cancer deaths per village rather of 2–3 cancer deaths per village.
We have conducted the analyses using the smaller groupings and
found the results to be similar. The purpose of this paper is to present
the methodology and to demonstrate an application of it.

Analysis

Poisson regression analyses have been conducted to determine
the cancer slope factor for the study villages. Arsenic exposure

<p>| Table 1 |
|------------------------|------------------|-----------------|-----------------|-----------------|-----------------|
| Population, exposure, and cancer data for the 42 study villages of SW Taiwan study (NRC, 1999). |
| Person-years | Persons both | Number of wells | Arsenic median | Arsenic mean | Arsenic max |</p>
<table>
<thead>
<tr>
<th>Village</th>
<th>M</th>
<th>F</th>
<th>Number</th>
<th>Male</th>
<th>Female</th>
<th>Male</th>
<th>Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 3-H</td>
<td>4159</td>
<td>4043</td>
<td>586</td>
<td>10</td>
<td>10</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>2 2-I</td>
<td>3529</td>
<td>3194</td>
<td>480</td>
<td>11</td>
<td>11</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>3 0-G</td>
<td>5388</td>
<td>4861</td>
<td>732</td>
<td>30</td>
<td>30</td>
<td>216</td>
<td>770</td>
</tr>
<tr>
<td>4 3-S</td>
<td>7851</td>
<td>7033</td>
<td>1063</td>
<td>32</td>
<td>32</td>
<td>32</td>
<td>32</td>
</tr>
<tr>
<td>5 3-N</td>
<td>2689</td>
<td>2392</td>
<td>363</td>
<td>12</td>
<td>12</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>6 4-7</td>
<td>10,629</td>
<td>10,227</td>
<td>1490</td>
<td>42</td>
<td>42</td>
<td>42</td>
<td>42</td>
</tr>
<tr>
<td>7 6-A</td>
<td>7716</td>
<td>6820</td>
<td>1038</td>
<td>15</td>
<td>15</td>
<td>45</td>
<td>45</td>
</tr>
<tr>
<td>8 0-J</td>
<td>6501</td>
<td>5888</td>
<td>885</td>
<td>20</td>
<td>20</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>9 3-L</td>
<td>6238</td>
<td>5094</td>
<td>809</td>
<td>20</td>
<td>20</td>
<td>56</td>
<td>56</td>
</tr>
<tr>
<td>10 4-D</td>
<td>10,107</td>
<td>9227</td>
<td>1381</td>
<td>60</td>
<td>60</td>
<td>60</td>
<td>60</td>
</tr>
<tr>
<td>11 3-P</td>
<td>6724</td>
<td>5927</td>
<td>893</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>12 6-0</td>
<td>12,767</td>
<td>11,937</td>
<td>1765</td>
<td>73</td>
<td>73</td>
<td>73</td>
<td>73</td>
</tr>
<tr>
<td>13 4-8</td>
<td>11,307</td>
<td>10,332</td>
<td>1546</td>
<td>80</td>
<td>80</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>14 0-0</td>
<td>6895</td>
<td>6392</td>
<td>949</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>15 0-E</td>
<td>5753</td>
<td>5210</td>
<td>790</td>
<td>110</td>
<td>110</td>
<td>236</td>
<td>686</td>
</tr>
<tr>
<td>16 0-I</td>
<td>4930</td>
<td>3833</td>
<td>577</td>
<td>71</td>
<td>71</td>
<td>236</td>
<td>590</td>
</tr>
<tr>
<td>17 4-N</td>
<td>4709</td>
<td>4291</td>
<td>643</td>
<td>123</td>
<td>123</td>
<td>127</td>
<td>127</td>
</tr>
<tr>
<td>18 4-J</td>
<td>6508</td>
<td>5062</td>
<td>895</td>
<td>126</td>
<td>126</td>
<td>126</td>
<td>126</td>
</tr>
<tr>
<td>19 2-D</td>
<td>9702</td>
<td>8869</td>
<td>1327</td>
<td>256</td>
<td>256</td>
<td>256</td>
<td>256</td>
</tr>
<tr>
<td>20 0-D</td>
<td>3872</td>
<td>3412</td>
<td>520</td>
<td>256</td>
<td>256</td>
<td>256</td>
<td>256</td>
</tr>
<tr>
<td>21 3-Q</td>
<td>5580</td>
<td>5079</td>
<td>761</td>
<td>259</td>
<td>259</td>
<td>269</td>
<td>458</td>
</tr>
<tr>
<td>22 4-M</td>
<td>2953</td>
<td>2758</td>
<td>408</td>
<td>307</td>
<td>307</td>
<td>307</td>
<td>307</td>
</tr>
<tr>
<td>23 6-6</td>
<td>5364</td>
<td>4505</td>
<td>705</td>
<td>307</td>
<td>307</td>
<td>307</td>
<td>307</td>
</tr>
<tr>
<td>24 4-E</td>
<td>3942</td>
<td>3586</td>
<td>538</td>
<td>350</td>
<td>350</td>
<td>350</td>
<td>350</td>
</tr>
<tr>
<td>25 4-L</td>
<td>3069</td>
<td>2723</td>
<td>414</td>
<td>398</td>
<td>398</td>
<td>398</td>
<td>398</td>
</tr>
<tr>
<td>26 4-F</td>
<td>4482</td>
<td>3866</td>
<td>598</td>
<td>406</td>
<td>406</td>
<td>417</td>
<td>417</td>
</tr>
<tr>
<td>27 3-I</td>
<td>4551</td>
<td>4259</td>
<td>629</td>
<td>448</td>
<td>448</td>
<td>448</td>
<td>448</td>
</tr>
<tr>
<td>28 5-G</td>
<td>3179</td>
<td>2598</td>
<td>606</td>
<td>467</td>
<td>467</td>
<td>467</td>
<td>467</td>
</tr>
<tr>
<td>29 4-P</td>
<td>5843</td>
<td>5397</td>
<td>803</td>
<td>504</td>
<td>504</td>
<td>504</td>
<td>504</td>
</tr>
<tr>
<td>30 0-H</td>
<td>4390</td>
<td>4315</td>
<td>622</td>
<td>520</td>
<td>520</td>
<td>520</td>
<td>520</td>
</tr>
<tr>
<td>31 4-I</td>
<td>4870</td>
<td>4432</td>
<td>664</td>
<td>472</td>
<td>472</td>
<td>514</td>
<td>514</td>
</tr>
<tr>
<td>32 3-J</td>
<td>8454</td>
<td>8689</td>
<td>1225</td>
<td>529</td>
<td>529</td>
<td>529</td>
<td>529</td>
</tr>
<tr>
<td>33 3-S</td>
<td>4287</td>
<td>3667</td>
<td>568</td>
<td>538</td>
<td>538</td>
<td>538</td>
<td>538</td>
</tr>
<tr>
<td>34 3-N</td>
<td>3655</td>
<td>3413</td>
<td>505</td>
<td>544</td>
<td>544</td>
<td>544</td>
<td>544</td>
</tr>
<tr>
<td>35 2-D</td>
<td>9059</td>
<td>7977</td>
<td>1217</td>
<td>599</td>
<td>599</td>
<td>599</td>
<td>599</td>
</tr>
<tr>
<td>36 4-G</td>
<td>2425</td>
<td>2108</td>
<td>324</td>
<td>650</td>
<td>650</td>
<td>650</td>
<td>650</td>
</tr>
<tr>
<td>37 5-A</td>
<td>3155</td>
<td>2983</td>
<td>438</td>
<td>650</td>
<td>650</td>
<td>650</td>
<td>650</td>
</tr>
<tr>
<td>38 2-M</td>
<td>11,123</td>
<td>11,263</td>
<td>1599</td>
<td>693</td>
<td>693</td>
<td>693</td>
<td>693</td>
</tr>
<tr>
<td>39 0-F</td>
<td>7010</td>
<td>5720</td>
<td>909</td>
<td>694</td>
<td>694</td>
<td>694</td>
<td>694</td>
</tr>
<tr>
<td>40 3-R</td>
<td>4310</td>
<td>3576</td>
<td>563</td>
<td>698</td>
<td>698</td>
<td>698</td>
<td>698</td>
</tr>
<tr>
<td>41 3-M</td>
<td>5815</td>
<td>4877</td>
<td>764</td>
<td>717</td>
<td>717</td>
<td>717</td>
<td>717</td>
</tr>
<tr>
<td>42 2-N</td>
<td>8341</td>
<td>8342</td>
<td>1192</td>
<td>934</td>
<td>934</td>
<td>818</td>
<td>960</td>
</tr>
<tr>
<td>Total</td>
<td>253,000</td>
<td>233,959</td>
<td>34,783</td>
<td>153</td>
<td>85</td>
<td>90</td>
<td>144</td>
</tr>
</tbody>
</table>
entered the analyses as the daily dosage (ug/kg/day) for the male and female adult (20 + years) populations of each village using the same exposure assumptions used in U.S. EPA (2010)–non-water arsenic intake of 10 ug/day, drinking water consumptions of 3.5 L/day for Taiwanese males and 2.0 L/day for Taiwanese females, and body weights of 50 kg for both Taiwanese males and females. The arsenic dosage for Taiwanese males was (10 ug/day + [3.5 L/day × As (ug/L)])/50 kg/person and for Taiwanese females was (10 ug/day + [2.0 L/day × As (ug/L)])/50 kg/person. Village arsenic dosage metrics were expressed as the village median, mean, or maximum.

Our study conducted Poisson regressions using the GEN MOD procedure of SAS with village cancer mortality counts as the dependent variable. The models included arsenic dosage as a linear parameter, age as a quadratic parameter, and village population as an offset variable. Age-distributions stratified in five-year intervals varied little among the study villages. Models were fit for males, for females, and for both together. An exact method was used when the model converged, and an approximation was used when the second-order partial derivative matrix (the Hessian matrix) could not be inverted. The scale parameter was held fixed. Two-tailed p-values were reported with 95% upper and lower confidence bounds.

As we were concerned that previous Poisson analyses of these data may have reflected the cancer risks from high levels of arsenic exposure in the assessment of risk at lower exposures, we sought a methodology whereby the assessment of risk at lower exposures was performed independent of data from higher levels of exposure. We conducted a backward stepwise sequential reductive Poisson regression analysis whereby we sequentially removed from the dataset the data from the highest exposure village. The results have been graphed with their 95% upper and lower confidence limits along with the number of villages continuing to remain in the analytic dataset. As the use of the median village well arsenic level as the exposure metric may hide the presence and influence of wells with higher arsenic levels, we have conducted the analytic models using the village medians, means, and maxima and have compared those findings. We have examined both separately and combined the risks for bladder and lung cancer mortality and for male and female lung cancer mortality. Here we present the analyses for male and female bladder and lung cancer mortality combined.

**Results**

Table 1 presents the underlying data for the analysis of the cancer mortality in the Wu et al. (1989) 42 village study. These data were first published in NRC (1999) [Table A10-1; pages 308–309] and later corrected for entry errors in Lamm et al. (2006) [Table 1; page 1778]. The 42 villages are presented in the NRC order by their median village well water arsenic level in ascending order with their village identifier. These include data on population (male and female person-years of observation and number of adult residents), exposure (number of wells and the median, mean, and maximum of the well water arsenic levels [ug/L]), and outcome (number of bladder and lung cancer deaths during 1973–1986 for males and females, and summed).

The villages have a two-item identifier that has a digit representing the township and a letter representing the village Guo (2007). The period of observation is the 14-years of 1973–1986, so the number of adult residents is the number of person-years divided by 14.

![Fig. 1. Village well water arsenic statistics for the 42 study villages of the SW Taiwan study (Wu et al., 1989).](image-url)
Well water arsenic levels ranged between 10 ug/L and 1752 ug/L with village medians of 10–934 ug/L, means of 10–818 ug/L, and maxima of 10–1752 ug/L. Arsenic levels were available for only one well in 20 of the 42 study villages (48%) and for two wells in an additional 10 study villages (24%). Only one study village had more than 11 wells. The average number of sampled wells was 3.6 per village for the 42 villages but with exclusion of the outlier (N = 47), the average number was 2.6 wells per village. The median arsenic level for the single-well villages (90 ug/L) was considerably lower than that for the multi-well villages (520 ug/L). The average number of adults per well in the 42 villages was 557 with an average of 898 for the single-well villages and 268 for the multi-well villages. (Table 1).

The village well water arsenic medians, means, and maxima are seen in Table 1 and their ratios of means to median and of maxima to median can be calculated. Table 1 and Fig. 1 contains a dashed line that separates the low (<150 ug/L) exposure villages and high (>150 ug/L) exposure villages, based on the median. The disparities between the village medians, means, and maxima are shown in Fig. 1. The high mean and maximum exposure levels in the low median exposure villages are bolded in Table 1. Two of the low exposure villages have means that are more than twice as great as their medians (villages 0-E and 0-I), and one has a mean more than seven times its median (village 0-C). The two villages have maximum arsenic levels that are 5–6 times as great as their medians, and the third low village has a maximum arsenic level that is more than 25 times as great as its median. Three additional villages (3-M, 4-F, and 0-H) have maximum levels that are about 2–3 times as great as their medians. They are also bolded in Table 1.

The bladder and lung cancer crude mortality rates (CMR) have been calculated as the number of deaths per 1000 person-years of observation for each village in Table 1 and are shown in Fig. 2. Fig. 2 is a scattergram of the CMRs by median village well water arsenic level (ug/L).

Poison analyses were conducted for the entire dataset using the median as the village arsenic metric. Then, sequential reductive analytic runs of the Poison analysis were performed with sequential truncation or elimination of the highest exposure villages. These results are shown in Fig. 3 where cancer slope factor (CSF) (change in the lifetime risk of bladder or lung cancer death per unit (ug/L) of arsenic) with its 95% confidence limits are plotted with respect to the highest arsenic level in each analysis. The cancer slope factor is statistically significantly positive with arsenic levels greater than 400 ug/L. It is positive but not statistically significantly so at levels of about 110–400 ug/L. At exposures below 100 ug/L, the cancer slope factor is negative. The cancer slope factor based on the median village well water arsenic level loses statistical significance below 400 ug/L and shows a discontinuity with a change in direction from positive to negative at 100 ug/L.

Earlier analyses have distinguished between the findings for the low exposure villages (median < 150 ug/L) and the findings for the high exposure villages (median > 150 ug/L). Fig. 3 demonstrates that the risk pattern is different between the two groups of villages. Since Table 1 and Fig. 1 showed that some of the low exposure villages have wells with high levels of arsenic, the sequential reductive Poison analyses have been replicated using the mean village well water arsenic levels (Fig. 4) and using the maximum village well water arsenic levels (Fig. 5) in order to separate out the effect of the high exposure wells.

Fig. 4 presents the findings from the sequential reductive Poison analyses using the mean as the village arsenic metric. The cancer slope factor is statistically significantly positive when exposures above 250 ug/L are included and is negative when the arsenic exposures are below 150 ug/L.

Fig. 5 presents the findings from the sequential reductive Poison analyses using the maximum as the village arsenic metric. The cancer slope factor is statistically significantly positive with arsenic levels above 400 ug/L, becomes negative below 150 ug/L, and becomes statistically significantly negative below 100 ug/L.

The patterns of the sequential reductive Poison regression analyses are similar, whether using the village well water arsenic median, mean, or maximum as the exposure metric. Fig. 6 shows the three together. In each case, the cancer slope is statistically significantly positive if the data for villages with greater than 400 ug/L arsenic are included in the analytic set. Statistical significance for the positive cancer slope factor is observed at >400 ug/L in two models and at >200 ug/L in the third (mean). Positivity is lost for the cancer slope factor below 150 ug/L for two models (mean and maximum) and below 100 ug/L in the third (median). These findings contrast with the U.S. EPA (2010) that “no evidence was found that either 400 ppb (ug/L) or 150 ppb (ug/L) represent “threshold” arsenic levels in drinking water below which
cancer risks are not increased (page F-7).” Whether one chooses to call it a “threshold” or not, Fig. 6 demonstrates that there is a discontinuity in the cancer slope factor in the range of 100–200 ug/L in the analysis of the bladder and lung cancer mortality data of the 42 study villages of Wu et al. (1989).

The cancer slope factor showed a marked discontinuity below 100 ug/L with the cross-over from positive to negative in the range of 100–200 ug/L, depending upon the choice of exposure metric. Each of the three models (using the median, mean, or maximum) showed the same pattern (Fig. 6) – i.e., positive above 200 ug/L, and negative below 100 ug/L. The cross-point of a zero CSF ranged from about 100 ug/L to about 200 ug/L, possibly evidence of a “threshold.”

**Discussion**

Epidemiological studies from three continents and from multiple countries have demonstrated the carcinogenic risk of
ingested inorganic arsenic from drinking water exposure levels in the hundreds to thousands of micrograms per liter (ppb). The estimated risk at lower exposure levels, i.e., exposure to levels of 10–100 ug/L that are the levels of interest in the United States, has generally been back-extrapolated from the risks estimated at high levels of exposure. Few studies have data that reasonably distribute across the exposure spectrum with multiple data points. The southwest Taiwan study of Wu et al. (1989) from the BFD-endemic area that has been used as the basis for recent quantitative risk analyses does.

The Wu et al. (1989) study has the advantage that its 42 study villages are spread widely across the arsenic exposure spectrum with about 1/3 of the villages having a median below 150 ug/L, 1/3 between 200 and 600 ug/L, and 1/3 above 600 ug/L. Publicly available analyses have generally analyzed the full dataset without giving consideration as to whether the risk from high arsenic exposures might overwhelm the assessment of risk at low exposures. We have sought to assess whether the risk exists consistently at all exposure levels or whether there is an exposure level at which there is a discontinuity in the risk estimate. We have
presented a sequential reductive analysis in which Poisson regression analyses were sequentially replicated with the removal of the villages with the highest level of arsenic exposure in order to assess the risks from still lower levels of exposure.

**Data analysis**

In this paper, we deal only with the issue as to whether it is reasonable to use the risk at high exposures as the measure of risk at low exposure, and we conclude that it is not. The analyses above consistently demonstrate a positive cancer slope factor when the analytic dataset includes the exposures that are greater than 200 ug/L and a statistically significant positive cancer slope factor when the analytic data include the exposures that are greater than 400 ug/L. The analyses above also consistently demonstrate a negative cancer slope factor when the levels are restricted to those below around 150 (±50) ug/L (i.e., 100–200 ug/L). Whether this slope factor is statistically significant or not may reflect the Poisson regression becoming less powerful with fewer villages in the analysis. Nonetheless, the discontinuity in these results contrasts with the U.S. EPA (2010) conclusion that “No evidence was found that either 400 ppb (ug/L) or 150 ppb (ug/L) represent “threshold” arsenic concentrations in drinking water below which cancer risks are not increased” (page F-7).

The use of the mean and of the maximum more effectively assigns to high exposure levels the villages with high exposure wells. The median inadequately represents the village exposures with respect to high levels of arsenic in the villages, since in villages with more low arsenic wells than high arsenic wells the median (but not the mean or the maximum) hides the presence of the high arsenic wells. For example, three of the villages (villages 0-G, 0-E, and 0-J) that had median village well water arsenic levels below 150 ug/L had one or more wells with arsenic levels >500 ug/L (range 580–770 ug/L).

While the EPA analyses were limited to the use of the median as the village exposure metric, our analyses both replicated their analyses and extended them to the village mean and the village maximum. The use of the additional two exposure metrics reveals the influence of the high exposures (>500 ug/L) exposures that are hidden within the village medians. We have examined for a threshold effect (discontinuity in the risk estimate) across the exposure range rather than only at 150 and 400 ug/L. As the iteration of the village exposure metrics removes the higher exposure levels, the exposure point at which the calculated cancer slope factor loses positivity increases from above about 100 ug/L to above about 200 ug/L.

The EPA analyses had included the southwest Taiwan regional data (a population more than 2000 times the size of the average village) as if it were an additional village in the study population and had assigned it an exposure level of zero ug/L. This transformed those data into a hyperinfluential data point that had the effect of forcing the direction of the slope. Further, those analyses made the assumption that all differences between the BFD-endemic area villages and the regional urban and rural data were solely attributable to the arsenic exposure. Analyses of the data for the low exposure villages have demonstrated significant risk differences between the regional data and the village data and have shown that these risk differences were attributable to factors other than the drinking water arsenic levels (Lamm et al., 2013). The analytic conclusion in this paper is that the risk above 200 ug/L did not predict the risk pattern below 200 ug/L. This analytic conclusion was not method-dependent as it was demonstrated both in the Poisson analyses (Fig. 6) and in the SMR analysis (Fig. 7).

A variant of Fig. 7 with village medians was previously published in Lamm et al. (2006). In spite of the more sophisticated analyses, the regression lines for the SMR analyses of the village means (Fig. 7) are quite similar to what would be seen with the crude mortality rate analyses of the village medians (Fig. 2).

The Lamm et al. (2006, 2007) analysis used the same 42-village dataset as does this one and reached similar results. The analyses differ in that the prior analysis presented linear regression analysis for the villages in the three townships that showed a

---

Fig. 7. SMRs for bladder and lung cancers for low and high arsenic villages by mean village well water arsenic level (ug/L).
significant dose-response relationship for bladder and lung cancer SMRs with respect to the median village well arsenic levels and assumed no uncertainty in the individual point estimates. It also presented separate SMR analyses for each of the outcome groupings so one could examine the differences whether outcomes are split or combined. This analysis presents a Poisson regression analysis for the villages of all six townships, assumes Poisson uncertainty in the number of cases for the individual point estimates, and, importantly, replicates the EPA methodology. The Poisson analysis has the advantage that it takes into consideration the uncertainties intrinsic to the variation in both case counts and village population and, critically, it is a non-linear analytic methodology applied to a data set that is demonstrably non-linear. The method avoids the situation of using a method that assumes linearity in order to examine for non-linearity.

The Poisson regression methodology is the preferred methodology as, unlike simple linear or multiple variable regression, it does not presume linearity. Poisson regression was chosen both because that is the model that EPA used and because it is of particular utility where the “n” (number of cases) is small. Here, 97% of the cells have a case count of fewer than two.

The most important difference is that this analysis focuses on the uncertainty that the median exposure represents the exposure of risk. It does that both by presenting identical analyses based on the mean and the maximum in addition to the median and by sequentially peaking back the high exposure data from calculations for the low exposure villages. The analyses presented here have been on the village bladder and lung cancer case counts combined rather than on either separately the bladder and lung cancer cases or the male and female cases. The analysis of the combined cancer cases showed that with respect to the mean village well arsenic levels that the cancer slope factor changed from positive to negative between 126 and 216 ug/L. Similarly, we have found that for male bladder, male lung, and female lung cancers the change occurred between 126 and 216 ug/L and for female bladder between 216 and 236 ug/L (not shown). The exposure level at which the cancer slope factor becomes zero is 157 ug/L for the cancers combined. The exposure levels at which the cancer slope factor becomes zero are lower for lung cancers (~30 ug/L) and higher for bladder cancers (~300 ug/L).

### Data quality

The data used in these analyses are the same data used by NRC, EPA, and WHO in their risk analyses. Issues that have been raised include (1) that the study villages comprise only 1/3 of the villages in the study area (42/124 = 34%), (2) that presumably the data are single measurements rather than well means or medians, (3) that the village median may not be the best summary statistic for the arsenic levels, and (4) that possibly not all wells used historically by the villagers were included in the 1964–1966 survey. The exposure assessment is limited to arsenic levels and does not include as relevant exposures either cigarette smoking or levels of humic or fluorescent substance in the village well waters. The issue of humic substances may be unique to this study. On the positive side, it might be noted that the exposure data are directly relevant to the study villages and contain no proxy exposures from “similar” or “neighboring” villages.

The first issue was the fact that the selection criteria for village inclusion was the availability of exposure data, not the availability of cancer data. The second is an observation for which we have no independent observation but which probably would have little effect on the analyses. The third has been dealt with in our approach to the data, and its implications have been demonstrated. The fourth, exposure misclassification due to missing wells, is a reasonable concern raised by the circumstances of the survey. Tap water supply came to the villages from about 1959–1967, with the proportion of residents on tap water at the time of the survey varying by township (Tsaï et al., 1998). The well sampling had been conducted in 1964–1966. Our on-going investigations have so far confirmed one-third of the NRC data points from currently available information including the published Kuo (1968) survey report. Future review of contemporaneous documents that are not now in the public domain may confirm the remainder of the data points and identify the magnitude and nature of missing well data. While there are issues related to the quality of the exposure data and its completeness or its representativeness of the village’s drinking water exposure, they are the only currently available data and the data upon which current water quality standards have been based and assessments considered.

### Global perspective

This geographical area came to public health notice and investigation in the late 1950s because of the presence of Black-foot Disease (BFD), the peripheral vascular disease uniquely found in this area. BFD has been found to be attributable to the high-arsenic artesian waters and possibly to associated fluorescent or humic substances. BFD is a known risk factor for the skin cancers (Tseng et al., 1988) and for the bladder cancers (Chiang et al., 1988). The bladder and lung cancer mortality pattern has been shown to be consistent with a threshold model at about 150 ug/L (Lamm et al., 2006, 2007), and bladder cancer incidence has been shown to be directly and significantly related to the level of fluorescent intensity, a measure of humic substances (Lu et al., 1986).

The dose-response relationship between bladder and lung cancer mortality and low levels of arsenic is unclear. The results of these Poisson analyses of the data are unique in terms of finding an occasional statistically significant negative slope for low-level arsenic exposure. However, a similar negative slope was also seen for white male bladder cancer mortality in the USA (1950–1979) and low levels of arsenic where the median ground water arsenic levels were in the range of 3–60 ug/L (Lamm et al., 2004). These results are consistent with the rest of the literature in not finding significant positive slopes in the exposure range below 100–200 ug/L for bladder or for lung cancers. Multiple bladder cancer studies (Bates et al., 1995; Kurttrio et al., 1999; Lewis et al., 1999; Steinmaus et al., 2003; Bates et al., 2004; Karagas et al., 2004; Michaud et al., 2004; Bastrup et al., 2008) showed no significant increased bladder cancer risk at exposures below 100 ug/L. In a meta-analysis of epidemiological studies of bladder cancers and low arsenic levels (<200 ug/L), Mink et al. (2008) demonstrated no association for non-smokers and only a non-statistically significant finding for “ever” smokers. Tsuji et al. (2014) concluded that the bladder cancer results for low-exposure studies, particularly for never-smokers, were statistically inconsistent with predicted risk based on high-dose extrapolation. Similarly, Celik et al. (2008) presented a systematic review and meta-analysis of the literature on lung cancers and arsenic in drinking water and found significant elevations only for those populations with mid-range exposures >150 ug/L.

The more recent studies in northeastern Taiwan found no significant increase in urinary cancer risk below 100 ug/L (Chen et al., 2010a [CPBM]) (Chen et al., 2010b [Environ. Res.]). Other recent findings include both those by Mostafa et al. (2008) who found in Bangladesh a significant association with lung cancer only for male smokers with arsenic exposures at >100 ug/L and found no association between lung cancer and arsenic at any level for non-smoking males and those by Steinmaus et al. (2013) who found in Chile no increased risk of lung cancer in those with arsenic exposures <200 ug/L.
Thus, these studies show that there is a great consistency in the published literature of high risks of cancers with arsenic exposures in the hundreds of micrograms per liter and not in areas with arsenic exposures below that level, i.e., <100 ug/L. The data that suggest associations between arsenic levels <100 ug/L and either bladder cancers or lung cancers are limited to smokers.

Risk analysis model

The predominant models for risk analysis include a linear dose-response at low levels and presume that this is an extension of the risk at high levels. The data from the Wu et al. (1989) study do not conveniently support this model unless all sources of uncertainty are removed from the analysis and an external reference point is added. Some analyses demonstrate a fit to a threshold model (Lamm et al., 2006, 2007), and the analyses above conform to a non-linear model with a "J" or "U" shape. Snow et al. (2005) have described toxicological modes of action that could explain such biologically plausible patterns. Cohen et al. (2006) have demonstrated the non-linear dose-response for arsenicals in rodents, and Suzuki et al. (2010) have demonstrated in rats fed sodium arsenite a dose-response with cytotoxicity, cell proliferation, and hyperplasia of the urothelium at 50 ppm but not at 1 ppm (1000 ppb). Arnold et al. (2013) concluded that their data on rats and mice orally administered arsenite supported the hypothesis that the sequence of events in the mode of action for urachelial effects involves superficial cytotoxicity with consequent regenerative increased cell proliferation similar to the findings associated with the administration of dimethylarsinic acid (DMA\textsuperscript{a}) in rats. The default to a linear model appears to fit neither the epidemiologic data nor the toxicologic findings.

Summary

The major finding demonstrated by the above analyses is that the cancer risk calculated from the high arsenic exposure data does not predict the risks observed among the villages with only low arsenic exposure data. The slope for the low exposure villages consistently presents here as a negative slope that is only rarely statistically significantly different from zero. These data do not show a positive slope, and most certainly not a statistically significant positive slope. In whatever manner the data for the low exposure villages are collapsed or aggregated or truncated, the risk in the low exposure villages is consistently seen to not be predicted by the risks calculated from the high exposure villages, i.e., the risk estimates based on the data for the high exposure villages over-predict the number of such cancers observed among the low exposure villages. These data do not fit a linear no-threshold model but do fit a model with discontinuity in the carcinogenic risk occurring in the range of 100–200 ug/L.

The southwest Taiwan dataset (Wu et al., 1989) has served as the basis for the development of the carcinogenic risk from the ingestion of inorganic arsenic from the NRC (1999, 2001) reports through the EPA reports of 2001, 2005, 2008, and 2010 draft and the SAB reports of 2007 and 2012.

In January 2013, the EPA announced that it was not promulgating its draft U.S. EPA (2010) document and would begin anew. EPA announced that its next toxicological assessment of the carcinogenic risk from the ingestion of inorganic arsenic would be based on the entire collection of relevant and reliable studies from the world’s literature and would not be limited to a single study, such as the southwest Taiwan study. In the next rendition of a toxicological assessment, all epidemiological studies, including the southwest Taiwan study, should include their dose-response patterns along with their strengths and difficulties and up-to-date re-analyses. In 2013, the National Research Council advised US EPA to separately look at disease occurrence at levels below 100 ug/L from that at greater levels and to include studies that had exposures at both above and below 100 ug/L (NRC, 2013). Our analytic approach would comport with that recommendation.

Conclusion

Poison analysis of cancer risk with arsenic exposure (NRC, 1999) using the data for the 42-village study from the BFJ-endemic area of southwest Taiwan shows a positive cancer slope factor with drinking water arsenic exposure levels above 200 ug/L that is statistically significant at levels of 400 ug/L or greater. In contrast, analyses limited to exposures below 150 ug/L show a negative cancer slope factor that is usually not statistically significant.

The use of a sequential reductive Poison regression analysis has revealed in the SW Taiwan study that the dose-response pattern of low arsenic exposures is different from the dose-response pattern of high arsenic exposures. There is no significant increase in risk across the low exposure spectrum (10–100–200 ug/L), and there is a significant increase in risk at higher levels. This difference is consistent with the published literature that has generally not found increased risks for studies at less than 100 ug/L and has generally found significantly increased risks at levels above that, particularly at 400–500 ug/L and above. Our analyses of this dataset demonstrate that the presumption of an arsenic-exposure cancer slope factor (i.e., straight line) consistent across the full exposure range is erroneous. Here we demonstrate in a single dataset the exposure range at which a significant cancer slope factor is observed and the exposure range at which that cancer slope factor does not apply.

Conflict of interest

The authors declare no conflict of interest.

Transparency document

The Transparency document associated with this article can be found in the online version.

Acknowledgments

This paper is dedicated to the late Dr. Arnold Engel for his persistence over twenty years in developing the refinement of risk analyses of arsenic and cancers with regards to the drinking waters of Taiwan, Inner Mongolia, and the United States. This paper was presented in part at the 2010 annual meeting of the Society for Risk Analysis (Salt Lake City, Utah). We thank Hamid Ferdosi and Nana Afari-Dwamena for their assistance in the development of the final manuscript.

References
