An Approach to Evaluating Occupational and Environmental Regulatory Standards: With a Drinking Water Standard for Arsenic as an Example

How-Ran Guo

Graduate Institute of Environmental and Occupational Health, Medical College, National Cheng Kung University, Taiwan

Abstract: An Approach to Evaluating Occupational and Environmental Regulatory Standards: With a Drinking Water Standard for Arsenic as an Example: How-Ran Guo. Graduate Institute of Environmental and Occupational Health, Medical College, National Cheng Kung University, Taiwan—Occupational and environmental regulatory standards are usually determined on the basis of data from high-dose exposures, either in humans or animals. Risk assessment models are applied to identify the level below which the risks are considered as acceptable. Even when the basis is human data, the validity of extrapolation is often questionable because the number of cases actually observed in the low-dose region is usually small, if any. Validation of the risk assessment model by using data on populations with low-dose exposures is desirable, but the lack of study power is the major concern in most cases. A meta-analysis combining data from more than one study may solve the problem, especially when a model overestimates risks associated with low-dose exposures. A maximum contaminant level (MCL) of arsenic in drinking water at 0.05 mg/l published by the US Environmental Protection Agency (EPA) was used as an example to demonstrate such a scenario. Because this MCL was derived by using data from a study on skin cancer in Taiwan, a validation was conducted by using data on Taiwanese with low-dose exposures. In comparison with the number of cases observed in four studies, the model was more likely to be invalid than to be valid at exposure levels below 0.17 ppm and overestimated the number of cases (11.08 vs. 5). Whereas the EPA has published a new MCL recently on the basis of new risk assessments on urinary bladder and lung cancers, re-visiting the validity of the old standard still provides insights for validating regulatory standards in the future. (J Occup Health 2002; 44: 355–359)

Key words: Regulatory standard, Arsenic, Risk assessment, Skin cancer, Meta-analysis, Validation study

Occupational and environmental regulatory standards are usually determined on the basis of data from studies conducted on high-dose exposures, either human epidemiologic studies or animal experiment. But, the risk that is considered acceptable by government agencies is usually very small, and therefore regulatory standards are set in the low-dose region. On the other hand, the number of cases actually observed in the low-dose region is usually small, if any at all. To identify the level that is suitable for the regulatory standard, therefore, risk assessment models are often applied. Even when the model is constructed on the basis of human data, it is often questionable whether the extrapolation of risks observed at high-dose levels to low-dose exposures is valid. For some health hazards, effects of high-dose exposure may be different from those of low-dose exposure, which could be attributable to different routes of metabolism, a threshold in the dose-response relationship, and various other reasons. Validation of such risk assessment models by using data on populations with low-dose exposures is desirable, but the lack of study power due to the limited number of cases is the major concern in most cases.

A meta-analysis combining data from more than one study may solve the problem when the number of cases in each study included in the analysis is small. In particular, if a model overestimates risks associated with low-dose exposure, it is sometimes possible to validate the model by using data on a small number of cases. In the current study, a maximum contaminant level (MCL)
for arsenic in drinking water published by the US Environmental Protection Agency (EPA) was used as an example to demonstrate such a scenario. This MCL was derived by using a cancer slope factor model developed by the EPA on the basis of data from a study on skin cancer in Taiwan1, 2.

Arsenic is a human carcinogen that can be frequently found in the natural environment and food. The cancer slope factor calculated by the EPA is frequently applied to conduct risk assessments for cancers associated with ingestion of arsenic3-6. This figure was derived by using a multi-stage Weibull model with data from a study in Taiwan by Tseng et al.7 on the dose-response relationship between arsenic levels in drinking water and prevalence of skin cancer8. The model also provided the scientific basis of the MCL at 0.05 mg/l for arsenic in drinking water before a new standard was published in 20019. The study by Tseng et al.10 was an ecologic analysis in which participants with exposure levels below 0.3 ppm (about 0.3 mg/l) were put into one group, including those who were exposed to the background levels of arsenic in drinking water9. Besides the possible errors introduced by using a single value to represent the exposure status of a unit population in ecological analyses9, factors including nutritional status, possible biases in the diagnosis, etc. were not considered in the risk assessment. Therefore, the validity of the model in the low-dose region has been questioned7. Whereas the EPA has published a new drinking water standard for arsenic recently5 based on new risk assessments with data on urinary bladder and lung cancers8, evaluation of this old MCL still provides insights for evaluating regulatory standards in the future.

Materials and Methods

For occupational and environmental regulatory standards that aim at protecting human beings, the evaluation of a standard should be based on human data, when available. Furthermore, among epidemiologic studies around the world, the first choices should be those from the same area(s) where the study (or studies) that served as the scientific basis of the standard was conducted. Because oftentimes only a few cases, if any at all, were observed in the low-dose region in a single study, a meta-analysis combing data from qualified studies is desirable. Sometimes, the number of cases observed is too small to tell whether there is an increased risk, but if a model overestimates risks associated with low-dose exposure, it might still be possible to tell whether there is a remarkable difference between the number of cases actually observed and the number of cases projected by the model. Even when all the available data are combined, nonetheless, the number of cases might still be too small to determine whether the risk assessment model is valid or not. In that case, a power calculation may help to cast some light on the validity of the model. In summary, a general approach to evaluate regulatory standards can be laid out as follows:

1. Identify all the epidemiologic studies on the health effect(s) that the standard is intended to prevent in the same area (s) where the study (or studies) that served as the scientific basis was conducted.
2. From each study, obtain or estimate the number of participants (or person-time) in each exposure stratum for each group defined by the factors taken into account in the model (such as age and gender).
3. From each study, obtain or estimate the number of cases attributable the baseline risk in each exposure stratum for each group defined by the factors taken into account in the model.
4. For each study, calculate the number of cases projected by the model in each exposure stratum for each group defined by the factors taken into account in the model.
5. Conduct a meta-analysis combining data on each study obtained from the previous steps to determine whether there is a remarkable difference between the number of cases actually observed and the number of cases projected by the risk assessment model that was applied to derive the standard.
6. Perform a power calculation if a conclusion cannot be drawn due to the small number of case observed in the area (s).
7. If the model is determined as valid in the area (s) where the original study (or studies) was conducted, conduct similar evaluations by using data from qualified studies in the country (area) where the standard is implemented.

To evaluate the MCL of arsenic in drinking water at 0.05 mg/l determined by the EPA, a literature search was first conducted in Medline and local journals to identify studies on the association between arsenic ingestion and occurrence of skin cancer in Taiwan. Only reports mentioning specific arsenic levels in drinking water were included in the analyses, and the validation was performed for exposure levels below 0.30 ppm, the upper limit of the lowest exposure category in the original Taiwanese study10.

For the meta-analysis, a likelihood approach was applied to analyze the data. In short, because skin cancer is a relatively rare event, its occurrence in a population can be approximated to by a Poisson distribution. Therefore, the likelihood of observing x cases is

\[ P(x) = e^{-\mu} \frac{\mu^x}{x!}, \]

when \( \mu \) is the expected number of cases according to the true risk. There are two competing hypotheses: \( H_0 \), no risk was attributable to arsenic exposure (\( \mu = B \), where \( B \) is the expected number of cases from background risk), and \( H_1 \), arsenic exposure accounted for additional E cases as predicted by the EPA model (\( \mu = B + E \)). Under \( H_0 \),
(the model is inappropriate), the likelihood of observing C is
\[ P(C|B) = e^{-B} B^C / C! \]

Under \( H_0 \) (the model is appropriate), the likelihood of observing C is
\[ P(C|B+E) = e^{-B} (B+E)^C / C! \]

Therefore, observations on a single or pooled study population can be regarded as a test of the validity of the model, and the likelihood ratio
\[ P(C|B) / P(C|B+E) = e^{E} B^C / (B+E)^C \]
can be used as a measurement of validity—a ratio less than 1 is in favor of the validity of the model; the larger the ratio, the less supportive of validity.

The power of a single or pooled study population to detect the difference between the number of cases predicted by the model (B+E) and the number of cases observed (C) at a significance level of \( \alpha = 0.05 \), \( \rho(R) \), can be obtained by
\[ \rho(R) = 1 - \Phi[1.645 - 2 (R^{1/2} - 1)C^{1/2}] \]
where R, the relative risk, is the ratio of the number of cases predicted by the model (B+E) to the number of cases observed in the study population (C), and \( \Phi[x] \) is the standard cumulative normal distribution function

In addition, the EPA model takes into account the effects of age and gender. Therefore, for each study, the expected number of cases and the projected number of cases were first calculated for each group as defined by age and gender, and then the total number of expected and the projected numbers of cases were obtained by adding the data from all the groups.

**Results**

In the literature search, four studies fit the selection criteria (studies on the association between arsenic ingestion and occurrence of skin cancer in Taiwan that reported specific arsenic levels in drinking water below 0.30 ppm)\(^{11-13}\) (Table 2). One of the studies had data on two exposure levels below 0.30 ppm\(^9\), and altogether those four studies could provide data on four exposure levels, among which the highest was 0.17 ppm. Nonetheless, studies on exposure levels of 0.00117 ppm and 0.00485 ppm could not provide useful information, because the expected numbers of cases introduced by the baseline risks were very close to zero, and the study failed

**Table 1. Characteristics of the Taiwanese Studies Applied to Validate EPA’s Risk Assessment Model for Arsenic in Low Dose Region**

<table>
<thead>
<tr>
<th>Study</th>
<th>Design</th>
<th>Study Period</th>
<th>Outcome</th>
<th>Study Area</th>
<th>Cases</th>
<th>Arsenic Level (mg/l)</th>
<th>Inclusion in the Current Study</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tseng et al.,</td>
<td>cross-</td>
<td>1958–1965</td>
<td>prevalence</td>
<td>2 villages near</td>
<td>all</td>
<td>0.00117</td>
<td>included</td>
</tr>
<tr>
<td>1965</td>
<td>sectional</td>
<td>1965</td>
<td></td>
<td>the BFD area</td>
<td>both</td>
<td>0</td>
<td>included</td>
</tr>
<tr>
<td>Tseng et al.,</td>
<td>cross-</td>
<td>1958–1968</td>
<td>prevalence</td>
<td>37 villages in</td>
<td>both</td>
<td>0.001 – 0.29</td>
<td>excluded</td>
</tr>
<tr>
<td>1968</td>
<td>sectional</td>
<td>1965</td>
<td></td>
<td>the BFD area</td>
<td>428</td>
<td>0.30 – 0.59, 0.60 – 1.82</td>
<td>excluded (because data were adopted by EPA risk assessment)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5 villages</td>
<td>0</td>
<td>0.001 – 0.017</td>
<td>included</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>near the BFD area</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Matsu Island</td>
<td>0</td>
<td>not detectable</td>
<td>excluded (because no actual exposure level reported)</td>
</tr>
<tr>
<td>Tseng et al.,</td>
<td>prospective</td>
<td>1958–1983</td>
<td>mortality</td>
<td>612 BFD patients</td>
<td>both</td>
<td>0.001 – 0.29</td>
<td>included</td>
</tr>
<tr>
<td>1983</td>
<td>cohort</td>
<td>1982</td>
<td></td>
<td>all both 1</td>
<td>4</td>
<td>0.30 – 0.59</td>
<td>excluded</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td>0.60 – 1.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wu et al.,</td>
<td>retrospective</td>
<td>1973–1989</td>
<td>mortality</td>
<td>42 villages in</td>
<td>both</td>
<td>0.001 – 0.29</td>
<td>included</td>
</tr>
<tr>
<td>1989</td>
<td>cohort</td>
<td>1986</td>
<td></td>
<td>the BFD area</td>
<td>&gt; 20</td>
<td>0.30 – 0.59</td>
<td>excluded</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>men</td>
<td>2</td>
<td>0.30 – 0.59</td>
<td>excluded</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>8</td>
<td>0.60 – 1.82</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>women</td>
<td>2</td>
<td>0.001 – 0.29</td>
<td>included</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
<td>0.30 – 0.59</td>
<td></td>
<td>excluded</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
<td>0.60 – 1.82</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\(^{1}\)blackfoot disease
to identify any cases. Data for the other two exposure levels suggested the model might not be able to provide accurate risk estimates for exposure levels below 0.17 ppm. When all three studies with useful data were taken into account, the model is about nine times more likely to be invalid than to be valid (Table 2).

The study power to detect the difference between the number of cases projected by the model and the number actually observed in the four epidemiologic studies combined was quite close to 80%, which is generally regarded as sufficient. The results indicated that even though the model was developed on the basis of a study in Taiwan, it cannot be validated by epidemiological data from Taiwan. Therefore, it is not suitable to conduct the second stage of evaluation—evaluation with data from studies in the country (area) where the standard is implemented.

**Discussion**

Many occupational and environmental regulatory standards were developed by extrapolating data from studies on high-dose exposures by using risk assessment models, which might not be valid in the low-dose region. With the advancement of measurement technology and accumulation of epidemiological data, it is now possible to assess the applicability of some of these models for generating risk estimates for low-dose exposures. In particular, with modern technology, trace amounts of many toxic substances can be detected in most food items and environmental samples. The validity of these models in the low-dose region should be critically evaluated to avoid unnecessary panic and loss, and the approaches used in this study can be easily applied for this purpose.

As shown in this study, the previous EPA cancer risk assessment model for ingested arsenic cannot even be applicable to the population that provided the data for the development of that model.

One of the major problems with risk assessment for low-dose exposure is the study power. It is not uncommon for the risk to be too small to be detected by a single study, or even all the existing studies pooled together. The case presented in this paper is a good example: the expected number of cases introduced by the baseline risks (5.66) was quite close to the number of cases actually observed (5), and therefore it is impossible to determine whether there was an increased risk in the pooled population. Nonetheless, with the likelihood ratio approach, we can still evaluate the validity of the model, as long as the expected number of cases attributable to baseline risks is not too close to zero. Therefore, this approach can be regarded as an alternative meta-analysis method that can be applied to validate risk assessment models.

This validation study showed that the previous EPA cancer risk assessment model for ingested arsenic cannot generate accurate risk estimates for low-dose exposure, even in the population that provided the basis for its development. Differences between effects of high-dose exposure and those of low-dose exposure could be attributable to various factors such as different routes of metabolism, a threshold in the dose-relationship, etc. These factors were ignored in the development of the standard under validation. Whereas the EPA has lowered the MCL for arsenic in drinking from 0.05 mg/l to 0.01 mg/l (similar to 0.01 ppm)\(^5\), the present study found that the previous risk assessment model tended to overestimate
risks in the low-dose region. Therefore, validation of the new model should also be conducted for low-dose exposure to ensure revisions of the standard are on a sound basis. In fact, the report leading to the revision of MCL claimed that studies in the US population could not detect any increase in the risk of bladder cancer because the lack of statistical power but did not provide data to support the claim, and some researchers argued that actual calculation of power be conducted. In addition to conventional power calculation methods, the approach applied in the current study may be used to determine whether data from previous studies in the US have the power to invalidate the study, especially because the current risk assessment model is likely to overestimate the risk in the low-dose region. Furthermore, the re-visit of the previous risk assessment model for arsenic as presented in this paper demonstrated the need to reevaluate published regulatory standards whenever new data become available.

Acknowledgments: This study was supported by the National Science Council of Taiwan, R.O.C. (Grant NSC-89-2314-B-006-171).

References
5) EPA. National primary drinking water regulations; Arsenic and clarifications to compliance and new source contaminants monitoring; Final rule. Fed Reg 66: 6976 (2001)
11) W-P Tseng, S-W How and S Yeh: Skin changes due to chronic arsenic poisoning on 186 residents emigrated from the blackfoot disease endemic area. Repts Inst Pathol Natl Taiwan Univ 16, 27–32 (1965)