



UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON, D.C. 20460

OFFICE OF THE ADMINISTRATOR
SCIENCE ADVISORY BOARD

December 20, 1999

EPA-SAB-EC-00-003

The Honorable Carol Browner
Administrator
United States Environmental Protection Agency
401 M Street, SW
Washington, DC 20460

Subject: An SAB Report on EPA's *Per Capita Water Ingestion in the United States*

Dear Ms. Browner:

This report presents the results of a review by the Drinking Water Intake Subcommittee, a special subcommittee established by the EPA Science Advisory Board (SAB) Executive Committee in response to a request from the Agency's Office of Water to review its report entitled *Estimated Per Capita Water Consumption in the United States*. The review was carried out during two meetings with representatives of the Agency during July 1999. The Subcommittee concluded that the EPA report will be an important reference with extensive utility both inside and outside the Environmental Protection Agency. In addition, the Subcommittee believes that the U.S. Department of Agriculture's (USDA) Continuing Survey of Food Intake by Individuals (CSFII) was the best available information source for the Agency to use in developing its estimates of drinking water ingestion by the U. S. population. The CSFII also provides a valid data set for estimating water ingestion for a limited number of subgroups within the population. However, the Subcommittee has concerns about the descriptive nature of the EPA report because it contains no explicit discussion of how these estimates might reasonably be used by the Agency in its scientific assessment and policy considerations.

Even though the report will be invaluable in providing information about the distribution of water consumption among the general population, it may be of limited value in providing information about the drinking water consumption of certain subpopulations that may be of interest to the Agency and to other users of the report. However, this limitation is due to the characteristics of the CSFII survey and not because of the Agency's analysis and interpretation of the data. The CSFII survey was aimed at characterizing the food intake of the general population and was not designed to gather information on specific subgroups or situations (e.g., very young children, Native Americans, individuals with diseases which impact their water consumption or workers in hot environments). As a result, although specific groups of interest are represented in the survey in proportion to their occurrence in the general population, the information needed to identify them may not be present and, even when it is, the sample sizes in the subgroups that can be identified are often too small to provide useful information on their water intake (e.g., for young children in certain ethnic or socioeconomic groups). Further, even though Native Americans are represented in the survey, the information gathered in CSFII does not allow one to differentiate which of the Native Americans who were included in the survey follow traditional Native American culture and lifestyle and which of them practice contemporary urban and suburban lifestyles.

Several approaches are possible if the Agency finds that it needs information on the distribution of water intake in subgroups, or for situations that are not adequately described by CSFII. One is to commission special surveys designed to gather the needed information about these groups. A second approach would be to rely on current understanding of the physiological need for water by individuals in different situations (e.g., developmental stages, physiological states, or environments) to characterize the likely water consumption and then to couple this information with survey information on the distribution of these developmental stages, physiological states, and environments in the population. Each approach has its strengths and weaknesses.

The draft report could be considerably strengthened, and the potential for misinterpretation of its findings could be reduced substantially, if the Agency provided information on the statistical significance of differences in water consumption between major subgroups of the population. Without such information, users of the report may be inclined to emphasize the differences in water consumption among subgroups which may in fact be artifacts of small sample sizes.

The SAB is prepared to provide additional review and assistance as EPA further develops these estimates. We look forward to the response to these comments from the Assistant Administrator for the Office of Water.

Sincerely,

/signed/

Dr. Joan M. Daisey, Chair
Science Advisory Board

/signed/

Dr. Henry Anderson, Co-Chair
Drinking Water Intake Subcommittee
Science Advisory Board

/signed/

Dr. Richard Bull, Co-Chair
Drinking Water Intake Subcommittee
Science Advisory Board

NOTICE

This report has been written as part of the activities of the Science Advisory Board, a public advisory group providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. The Board is structured to provide balanced, expert assessment of scientific matters related to problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the Environmental Protection Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names or commercial products constitute a recommendation for use.

Distribution and Availability: This Science Advisory Board report is provided to the EPA Administrator, senior Agency management, appropriate program staff, interested members of the public, and is posted on the SAB website (www.epa.gov/sab). Information on its availability is also

provided in the SAB's monthly newsletter (*Happenings at the Science Advisory Board*). Additional copies and further information are available from the SAB staff.

**U.S. Environmental Protection Agency
Science Advisory Board
Drinking Water Intake Subcommittee
Panel for Review of the EPA Report on Drinking Water Consumption
July 8 and 19-20, 1999**

Co-Chairs

Dr. Henry A. Anderson, Wisconsin Bureau of Public Health, Madison, WI

Dr. Richard Bull, Battelle Pacific Northwest Laboratories, Richland, WA

Panelists

Dr. Judy Bean, Children's Hospital Medical Center, Cincinnati, OH

Dr. Cynthia Bearer, Case Western Reserve University, Cleveland, OH

Dr. John Evans, Harvard School of Public Health, Boston, MA

Dr. Anna Fan-Cheuk, California Environmental Protection Agency, Oakland, CA

Dr. Richard Gilbert, Battelle Washington Office, Washington, DC

Dr. Barbara L. Harper, Yakama Indian Nation, Richland, WA

Dr. Michael Jayjock, Rohm and Haas Co., Spring House, PA

Dr. Kai-Shen Liu, California Department of Health Services, Berkeley, CA

Dr. Edo Pellizzari, Research Triangle Institute, Research Triangle Park, NC

Dr. Barbara Petersen, President, Novigen Sciences, Inc., Washington, D.C.

Science Advisory Board Staff

Mr. Thomas O. Miller, Designated Federal Official, U.S. Environmental Protection Agency,
Washington, DC

Ms. Dorothy Clark, Management Assistant, U.S. Environmental Protection Agency,
Washington, DC 20460

TABLE OF CONTENTS

1. EXECUTIVE SUMMARY AND CONCLUSIONS	1
1.1. EPA’s goals and objectives as stated for the Report, and the analyses it contains, are too limited	1
1.2. The report should state that EPA did not have information that would allow calculation of confidence intervals for subpopulations	3
1.3. The Agency should develop a strategy for the analysis, presentation and interpretation of the Report’s data that is consistent with the intended uses of the data	3
2. INTRODUCTION AND CHARGE	4
3. SUBCOMMITTEE COMMENTS.	6
3.1. General Comments	6
3.2. EPA’s goals and objectives as stated for this Report, and the analyses it contains, are too limited	7
3.3. The EPA Report should state that EPA did not have information that would allow calculation of confidence intervals for subpopulations.	10
3.4. The Agency needs to develop a strategy for the analysis, presentation, and interpretation of the data that is consistent with the intended uses of the data	10
Appendix A: Responses to Specific Agency Charge Questions	A-1
Charge Question 1. Statistical methodology	A-1
Charge Question 2. Confidence limits	A-1
Charge Question 3. Short-term data and long-term estimates	A-1
Charge Question 4. Data conventions	A-2
Charge Question 5. Subpopulation distributions	A-3
Charge Question 6. Subpopulations included	A-4
Charge Question 7. Indirect water	A-5
Charge Question 8. Food types not covered	A-6
Charge Question 9. Intrinsic water	A-6
Charge Question 10. Other issues	A-7
Charge Question 11. Zero values	A-8
References	R-1

1. EXECUTIVE SUMMARY AND CONCLUSIONS

The U. S. Environmental Protection Agency's (EPA) Science Advisory Board (SAB) was asked to perform a peer review of the Agency report *Estimated Per Capita Water Consumption in the United States* (hereafter referred to as the Report). The SAB Executive Committee established the Drinking Water Intake Subcommittee (DWIS) to conduct this review. The DWIS reviewed the Report during two meetings: one, a telephone conference meeting, on July 8, 1999 and the other, a face-to-face meeting, on July 19 to 20, 1999. Major Subcommittee comments on the EPA Report are contained below in this SAB report. Specific responses to the 11 charge questions are provided in Appendix A to this SAB report.

EPA is commended for seeking out databases that can be used for estimating ingestion of drinking water on a national scale. The database selected as the analytic basis for the report, the U. S. Department of Agriculture's Continuing Survey of Food Intake by Individuals (hereafter, the CSFII or the Survey), is the best available and has critical attributes that allow advancement of our understanding of ingestion of water by the general population of the United States. The Agency's efforts to develop ingestion estimates from the CSFII survey data were significant.

The committee believes that the EPA Report will be an important reference resource with extensive utility both within and outside the Agency. However, the following issues must be considered if the Report is to achieve its full potential.

1.1 EPA's goals and objectives as stated for this Report, and the analyses it contains, were too limited.

The Subcommittee is concerned that the Report is only descriptive and that it does not explicitly discuss how the estimates might be reasonably used. The Agency has both scientific and policy reasons for estimating water ingestion for the overall population, and for subpopulations, that are not discussed in the report. Some of these respond to the statutory mandate in the Safe Drinking Water Act. But there are other needs for information on drinking water intake in risk assessment and regulation which involve establishing default values for water ingestion, estimation of risks to highly exposed and/or sensitive subpopulations, and characterization of the distribution of individual risks or the impacts of specific control strategies. Important implications to these uses are not discussed in the current report.

EPA often uses default values for water ingestion levels when it develops allowable concentrations for contaminants in drinking water. The Subcommittee is encouraged that this EPA report provides information that will permit analysts to use specific data for water ingestion in many future situations where allowable concentrations must be developed. For others, the Report will provide assistance for developing information on the distribution of drinking water ingestion by individuals that includes new information and the relationship of ingestion to factors such as age, gender, and disease status.

While the report does a good job of characterizing the distribution of drinking water consumption in the entire US population, and in the major subdivisions of the US population (i.e., by age, sex, race, and geographic region), it does not provide the information that some users may want on drinking water ingestion by smaller subpopulations. Further, certain groups may have higher than normal water ingestion levels or they may be more sensitive to the effects of contaminants in drinking water. Examples of these would include very young children and workers in hot and/or dry climates.

This limitation exists because the CSFII data upon which the Agency relied for generating its estimates were collected in an effort to characterize the patterns of food consumption in the general population. They did not target certain subgroups that are now of heightened interest to EPA. Therefore the samples in certain subgroups are so small that the CSFII estimates of water consumption in these groups may be quite imprecise. Compounding this problem is the Report's omission of statistical confidence intervals for most of the ingestion estimates among subgroups of the population. If legislative mandates or regulatory analysis require information on the water consumption of these subgroups, further studies will be needed.

Many of the results presented in the report may be sensitive to assumptions made during data analysis. Examples of such data analysis conventions include the choice of regional boundaries and the assignment of a principal source for ingested water. Currently the report does not include a section analyzing the sensitivity of key results to these assumptions. The Subcommittee urges the Agency to conduct a sensitivity analysis and to add a section to the report describing the key findings from the sensitivity analysis.

Another key issue influencing the interpretation of the CSFII data is the choice of averaging time. We know that in many other settings (e.g., air pollution exposure assessment) heterogeneity tends to decrease as averaging time increases. The exact nature of the relationship between averaging time and observed heterogeneity depends on the features of the data being explored. For certain purposes (e.g., cancer risk assessment) the population distributions of long term average exposures may be of interest. The current EPA report provides information about drinking water intake averaged over TWO days. Therefore, to minimize the potential for misuse of the data in the EPA Report, users might benefit if the Agency clearly stated the averaging time on all tables and graphs in the report. Further, it may be necessary to more fully explore the sensitivity of results to alternative choices of averaging time.

Therefore, the Subcommittee recommends that the EPA Report discuss the characteristics of the EPA methods for estimating ingestion, and the USDA method for conducting the CSFII, that have important implications for those who must use the ingestion data.

1.2 The EPA Report should state that EPA did not have information that would allow calculation of confidence intervals for sub-populations.

A discussion point for the subcommittee centered on the question of whether it was appropriate to provide data without meaningful confidence intervals. The design of the CSFII survey requires use of an ultimate cluster methodology which is an aggregate of sampled persons within each primary sampling unit. Smaller subpopulations within the sample (e.g., the less than one year olds) did not meet these criteria. This prevents the calculation of confidence intervals using the ultimate cluster methodology. It would be good to clarify this point for the readers of the report.

1.3 The Agency should develop a strategy for the analysis, presentation and interpretation of the Report's data that is consistent with the intended uses of the data.

The Agency has taken a purely descriptive approach to the analysis and presentation of data. This results in numerous tables containing drinking water ingestion estimates for many conceivable combinations of attributes examined (e.g., Native American males by age group and by geographic region, etc.). While this superficially exhaustive presentation of data may seem attractive, the Subcommittee is concerned that this strategy for analysis, interpretation, and presentation of the data is inadequate and potentially misleading. We urge the Agency to develop a strategy for data analysis which, at a minimum, provides only those estimates of drinking water intake for which estimates of uncertainty can also be developed, and preferably which includes formal hypothesis tests of the significance of differences in the water consumption of various groups. Further, the number of tables presented in the report should be substantially reduced and limited to only those which support Agency needs and for which valid estimates of precision can be provided. If the Agency feels that certain tables for which valid estimates of precision can not be produced are necessary, this fact should be prominently displayed on each such table.

2. INTRODUCTION AND CHARGE

The Drinking Water Intake Subcommittee was asked to conduct a peer review of the Agency Report that provides estimates of per capita water intake in the United States. The Report contains estimates of the amount of direct and indirect water consumption. Direct water consumption is defined as plain water consumed directly as a beverage. Indirect water is that water added to foods and beverages during final home or restaurant preparation.

Empirical distributions of estimated water consumption were generated by water source and by the respondent's demographic and physical characteristics. Water sources include: a) the community water supply, b) bottled water, c) other sources including the respondent's own well, rain cistern, spring, or public spring. Physical and demographics characteristics include: age, gender, race, socioeconomic status, and geographic region. Estimates were also generated separately for pregnant and lactating women.

The distributions of estimated water ingestion include point estimates of the mean and the following percentiles: 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th, and 99th. Confidence intervals for the mean and bootstrap intervals for the upper percentiles are provided for only the larger subpopulations.

The charge to the Drinking Water Intake Subcommittee from the Office of Science and Technology, US EPA Office of Water included the following questions:

- a) The distributions of estimated water intake were generated using standard statistical methodology for surveys with complex designs such as the 1994-96 CSFII. Is the statistical methodology used to generate the estimates appropriate? Should we consider rounding?
- b) We have limited the calculation of confidence intervals about the mean and boot strap intervals for percentiles to the distributions for the larger sub-populations. The complex sample design makes the calculation and interpretation of results for smaller sub-populations virtually impossible to calculate and interpret. Is this an appropriate decision?
- c) The CSFII survey is based on short-term survey data. Upper percentile estimates may differ for short-term and long-term data because short-term survey data tends to be inherently more variable. Is it appropriate to report upper percentile estimates such as the 99th percentile?
- d) Are the data conventions used to identify direct and indirect water appropriate?

- e) Do the data support estimates of sub-population distributions?
- f) We have provided distributions of estimated water intake for numerous sub-populations. Should any additional sub-populations be added? Should any be excluded? Specify sub-populations.
- g) USDA has identified two types of indirect water in foods. They are:
 - i. The amount of water in food as consumed.
 - ii. The amount of water used to prepare food.The water intake report provides estimates of the amount of indirect water in food as consumed. If resources permit, we could expand our report as a future addendum to include estimates of the amount of indirect water used to prepare food. Would this be desirable?
- h) Additional water intake estimates associated with types of food may be useful for specific risk-exposure analyses, e.g., cold beverage intake. Such analyses are feasible using the CSFII data. We could expand our report as a future addendum if resources permit. Are any such targeted analyses of significant interest at this time?
- i) Intrinsic water is the water contained in foods and beverages at the time of market purchase. Intrinsic water includes commercial water (added to food products by food manufacturers) and biological water (found naturally in foods). Intrinsic water is not included in our current analysis. If resources permit, we could expand our report as a future addendum to include estimates of intrinsic water. Would this be desirable?
- j) What are the scientific limitations to the use of the water consumption estimates provided in this report (i.e., what other issues has the Subcommittee noted with the estimates that are not covered elsewhere)?
- k) The water intake estimates provided in this report are based on all respondents, including those who did not report consuming water during the two survey days. If resources permit, we could also generate estimates of water consumption which exclude the zero consumers of water. We noticed that for some sub-populations, especially the less than one-year-old infants, a substantial proportion consumed zero or minimal amounts of tap water per day (presumably those who were breast fed or drank undiluted formula or milk); these zero consumers of water can contribute to lower estimates. Would this be desirable?

3. SUBCOMMITTEE COMMENTS

3.1 General Comments

The Drinking Water Intake Subcommittee (DWIS) of the EPA Science Advisory Board (SAB) has reviewed the Agency's report entitled *Estimated Per Capita Water Consumption in the United States* during two meetings: one on July 8, 1999 and the other on July 19 to 20, 1999. Specific responses to the Agency's charge questions are provided in Appendix A to this SAB report.

EPA used the Continuing Survey of Food Intake by Individuals (CSFII) as its data source for use in deriving its drinking water ingestion estimates. EPA is commended for seeking out databases that can be used for estimating drinking water ingestion for the population at a national scale. This was the best data source available and it had critical attributes that allow advancement of our understanding of drinking water ingestion in the general U. S. population and the Agency made good use of the data. Several strengths of the USDA 1994-96 CSFII database are worth highlighting.

- a) The database is large, recent, and it is a population based survey.
- b) The database permits the categorization of various sources of ingested water.
- c) The convention used to estimate the fraction of water in each food as consumed was scientifically defensible.
- d) The database permits a breakdown of the US population into some major groups based on age, gender, special populations of females, regions, and broad classifications of ingested water source.

The committee believes that the EPA Report will be an important source of information on drinking water ingestion. The report will enjoy extensive use as a reference resource for those within and outside of the Agency.

Even though this report will allow EPA to better understand contaminant exposures associated with drinking water ingestion it does not, nor was it intended to, provide insight into exposure to drinking water contaminants associated with dermal exposure (e.g., during bathing or showering). Further, even though it provides estimates for some combinations of attributes, many such attribute combinations are possible. Most of these are not included nor could all possible combinations of potential interest be covered (e.g., infants who live in hot climates and have health conditions which affect water intake). Therefore, to fully estimate contaminant exposures associated with drinking water, EPA will need to go beyond projections that are based solely on information contained in this specific Report.

Notwithstanding the strengths of the Survey and the EPA Report noted above, the Subcommittee does have a number of concerns with the Report. If the Report is to achieve its full potential there are a number of issues that require further attention. These are discussed in Sections 3.2 through 3.4 that follow.

3.2 EPA's goals and objectives as stated for this Report, and the analyses it contains, are too limited.

The report was constructed only as a descriptive report without an explicit discussion of how the estimates in the Report might reasonably be applied by users. The subcommittee has several recommendations for revision to address this current shortcoming:

- a) The Report needs a prominent and early explanation of the logic used in the survey design and in the analyses used to develop the Agency's estimates. This explanation should be understandable by the educated layperson. This is not a criticism of the technical logic used in the analysis, rather, it simply recognizes that most users of these estimates will not have the specialized knowledge of statistics needed to understand fully the approach used.
- b) The report must provide a much clearer indication of which estimates are reliable and which ones are not as reliable. The extensive tables of statistics that appear to break down the population to several subgroups provide potential users of the data with a false sense of security about the precision of the estimates. This practically guarantees that the results will be applied in ways not supportable by the database.

The Agency has both scientific and policy reasons for estimating water ingestion in the overall population and in subpopulations of interest. Some of these come from the statutory mandates of the Safe Drinking Water Act (SDWA), but others come from the broader environmental health community, such as: a) risk assessment; b) development of default values; and c) sensitive subpopulations. The implications of the survey characteristics and the analyses supporting EPA's estimates on these uses of the ingestion estimates are not sufficiently discussed in the Report.

Risk assessments are scientifically-based efforts to estimate the impact that exposure to a contaminant, or groups of contaminants in water, may have on human health. For waterborne risk scenarios, it is important to construct as complete a picture of water ingestion as is possible. Some of the distinctions in the present estimates limit that capability. For example, as the Agency rightly points out, direct and indirect water represent only part of potential tap water ingestion (and therefore exposure to waterborne contamination). Commercial water (that water added by the manufacturer prior to marketing—not now included in the EPA estimates) is frequently taken from tap sources, although these are frequently far removed from the point of consumption. This does not mean that the estimates obtained from the present study cannot be used in developing risk assessments, but part of

the exposure assessment may have to obtain broader categories of water source than are identified in the present analysis of the data. Such limitations in the tabular data need to be clearly stated in the report.

In its current configuration, the report provides estimates that are composites of both those who reported drinking water during both survey days and those who reported drinking none on those days. As EPA noted in its charge to the Subcommittee, this could result in underestimates of drinking water ingestion. EPA traditionally uses a default value for water ingestion when converting a “safe dose” (mg/kg/day) to enforceable concentration limits in drinking water. The Report permits EPA to use improved data in developing such limits. However, the Subcommittee believes that such analyses should focus on those portions of the population that actually ingest drinking water. The estimates needed in this circumstance should not be diluted by including large numbers of individuals that reported no water ingestion during the survey (see Question 11 in Appendix A). In its current configuration, the report provides only the diluted estimates. Ingestion estimates should be developed by EPA to reflect only those who actually reported water ingestion as well as the current composite situation. When sufficient data are available to estimate confidence intervals, these Survey data can be used to develop default values.

Some subpopulations of interest are adequately represented in the report (e.g., pregnant women) but others identified included too few representatives (e.g., children of Native Americans). For this reason, the Subcommittee strongly recommends that the Report make explicit the limitations of the estimates. The breakout of pregnant and lactating women provides at least a starting point for defining the amount of water that is consumed by populations that may have special sensitivities. There are also some data that can be used to estimate water consumption by individuals of varying age. However, it is important to recognize and identify the limitations of these data for smaller populations (e.g., children of Native Americans/ Alaskans). In addition, other populations could be identified that consume higher amounts of water (e.g., diabetics and individuals with kidney disease) that, while not rare in the overall population, are well below the statistical power of the Survey to detect. If there are not sufficient data to support development of relatively robust measures of confidence, the use of the data to describe water ingestion by these smaller subgroups would be misleading and do a disservice to these groups. If these groups are to be a source of particular concern in the Agency’s regulatory agenda, surveys should be conducted that are adequate to support such estimates. Some other data sources might be superior for such purposes (e.g., NHANES).

The report provided insight into the 1.0 liter/10 kilogram default value for ingestion of drinking water by children that is currently used by EPA. The analysis presented in the EPA report shows that water consumption per unit body weight is very high at birth and falls off sharply with age. The Subcommittee is encouraged that the EPA Report now provides information that will permit analysts to use specific data for water ingestion in many future instances where allowable concentrations must be developed. For others, the Report will provide a better basis for developing reasonable defaults.

In the Agency's derivation of maximum contaminant limit goals (MCLGs) the mathematical operation essentially converts consumption to ml/kg/day, the Subcommittee believes that there is significant value to be gained from expressing estimates in these units as well as volume ingested. When shown in such units, the real differences in water consumption by age become much more apparent than when given as volume measures alone. ML/kg/day figures are best used until ingestion stabilizes and then the daily volume becomes equally appropriate.

Clearly, the EPA Report is not intended to answer questions about other critical subpopulations (e.g., workers that consume very large quantities of water because of the exertion involved in their work or because of working in hot and/or dry climates). This points to an opportunity for future work in this area. Some of this information may already be available in the literature. If not such efforts could involve designing a relatively simple hypothesis and model of the determinants of water ingestion. Some independent variables for such a model of water ingestion could include: a) level of effort or metabolic rate; b) average ambient air temperature; c) average ambient relative humidity; d) body weight; and e) age.

Describing and capturing data for these predictor variables, and subsequent water ingestion for subpopulations that share common (and relatively narrow) ranges of these variables, could lead to the identification of the subpopulations of greatest concern for contaminant exposures through drinking water. It might also lead to the development and validation of a comprehensive model for the prediction of water ingestion from such parameters. The resulting simple hypothesis and model of the determinants of water ingestion could be used generically because it would reflect water needs of individuals. In some individuals most, if not all, of that water requirement might come from tap water. Those are the persons that the SDWA is intended to protect. If more accurate estimates of actual drinking water ingestion are needed, appropriate data could be collected by targeted surveys. The results could always be benchmarked against the basic water needs of individuals under different physiological conditions.

The value of some of the tabular distributions provided in the analysis is not clear. For example, water ingestion was provided by region. The Subcommittee's agrees with the need for regional estimates; however, the political regions identified in the Agency Report were probably too large. The within region variability of ingestion is probably much larger than that between regions.

It is important to emphasize that risk is a function of both exposure and sensitivity. Sensitivity is determined by genetics, developmental stage (old as well as young), lifestyle, and preexisting disease conditions that are not addressed in the Report. The Agency should simply point out that these other determinants of sensitivity are not addressed in the report.

3.3 The EPA Report should state that EPA did not have information that would allow

calculation of confidence intervals for sub-populations.

A discussion point for the subcommittee centered on the question of whether it was appropriate to provide data without meaningful confidence intervals. The design of the CSFII survey requires use of an ultimate cluster methodology which is an aggregate of sampled persons within each primary sampling unit. Smaller subpopulations within the sample (e.g., the less than one year olds) did not meet these criteria. This prevents the calculation of confidence intervals using the ultimate cluster methodology. It would be good to clarify this point for the readers of the report.

3.4 The Agency needs to develop a strategy for the analysis, presentation, and interpretation of data that is consistent with the intended uses of the data.

The report should contain a description of the methodology used for analyzing the data. This would better explain the approach employed for those who are not experts in the sophisticated statistical techniques. In addition, the report should contain a strategy for future analyses of the data including some hypothesis testing.

Data validation and quality assurance procedures used in the development of the report should be prominently documented, with especial attention to conventions that were developed to handle some of the data.

The presentation of numerous tables containing estimates developed in the Agency analysis are clearly not appropriate for many of the applications the Agency will have for this information. Tables should be substantially reduced. Instead of numerous tables with estimates having unknown confidence levels the report should be limited to tables with estimates that support agency needs and for which valid estimates of reliability can be provided. These tables should be displayed in a useful way with significant figures appropriate to the level of precision in the estimates. The text surrounding these fewer tables should make clear the limitations of the estimates and whether they can be applied with confidence to evaluations of the subpopulations with which they are identified.

For example, the Subcommittee had very little confidence that the data reported for Native Americans reflected a Native American lifestyle (see Question 6 in Appendix A). There is a difference between “race” and “lifestyle”. The reasons for different intake rates primarily reflects lifestyle (secondarily SES), and probably not race *per se*. If the Agency is convinced that this data reflects such a lifestyle, it should explain the rationale supporting the conclusion. A contrary conclusion should also be clearly explained.

Similarly, separate tables should be provided reflecting ingestion estimates for those respondents reporting water ingestion during the two days captured in the CSFII. This should be in addition to tables that reflect estimates based on a composite of respondents reporting tapwater ingestion and those who did not report such ingestion. Both sets of analyses provide important

perspectives depending upon the use that the data will be applied to by the Agency. There are also good reasons to display data in both in terms of ml/kg/day as well as liters consumed. In all cases these data should include some measure of the precision of the estimate.

It is extremely important to segregate estimates for children by age for the reasons stated earlier. However, it is much less important to separate estimates for adults by age because the differences observed are much smaller. In adults the future analytical focus should be on identifying subpopulations that consume more water for other reasons, such as preexisting disease (e.g., diabetes mellitus), occupational conditions, or effects due to climate.

Appendix A

Responses to Specific Agency Charge Questions

1. Statistical Methodology

Charge Question 1: *The distributions of estimated water intake were generated using standard statistical methodology for surveys with complex designs such as the 1994-96 CSFII. Is the statistical methodology used to generate the estimates appropriate?*

The methodology described in the document is an appropriate technique to produce estimates from a multi-stage, stratified, clustered sample. The Agency, however, did not clearly state that the estimates were generated from a summary tape containing only final weights assigned to individuals. This means that the Agency was limited in what it could do with the data. References to the documents describing estimating equations for the US Department of Agriculture's Continuing Survey of Food Intakes by Individuals are needed.

2. Confidence Limits

Charge Question 2: *We have limited the calculation of confidence intervals about the mean and boot strap intervals for percentiles to the distributions for the larger sub-populations. The complex sample design makes the calculation and interpretation of results for smaller sub-populations virtually impossible to calculate and interpret. Is this an appropriate decision?*

Yes. However, the rationale for this is buried in the narrative. The Subcommittee recommends that the Agency state more clearly, and in a prominent place, its reasoning for not calculating such intervals throughout the report. Also, the convention of placing "zeros" as entries in the tables for place-holders where no estimates have been generated is confusing. The Subcommittee recommends inserting "dashes" in place of such zeros. This convention is used by others reporting results from such efforts.

3. Short-term Data and Long-term Estimates

Charge Question 3: *The CSFII survey is based on short-term survey data. Upper percentile estimates may differ for short-term and long-term data because short-term survey data tends to be inherently more variable. Is it appropriate to report upper percentile estimates such as the 99th percentile?*

The decision whether to report upper percentile estimates depends in part on whether the

quality of these estimates is sufficient for their intended use. Quality may be judged by the number of individuals interviewed, the fulfillment of underlying assumptions, and the computed statistical precision, bias and confidence in the percentile estimates. Uses of the drinking water ingestion estimates may be very broad and could include risk assessment, rule-development for microbial contaminants of drinking water and disinfection by-products, as well as other uses not now anticipated. Some uses of drinking water ingestion estimates may require the short-term survey data available from the present CSFII survey data (i.e., estimates of daily averages based on only two non-consecutive days of data), while other uses may need long-term survey data (i.e., estimates based on more than 2 days of data). For example, short term data and a knowledge of the variability of such data can be useful for risk assessments of acute health effects such as diarrhea due to microbiological contamination, whereas long term data and a knowledge of its variability are needed for risk assessments of long-term health effects such as cancer. As the short-term data available from the current CSFII survey are not ideally suited for all uses, it is particularly important that the report adequately describe the quality of the estimates so that users can judge if the results of the current survey are of sufficient quality. As indicated above, this quality can be described in various ways such as by providing variances and confidence limits for estimated percentiles, by carefully stating and explaining all assumptions used in obtaining those estimates, and by the number of individuals interviewed in the various subcategories.

The number of individuals interviewed in subcategories is sometimes very small in the CSFII data. This point is illustrated by reference to Table A-3b in Section 11e of the EPA Report. In this table, there is only one individual in the <0.5 year age category and only three individuals in the 0.5-0.9 age category. Clearly, upper percentiles should not be reported for categories for which the number of persons interviewed is so small. The National Center for Health Statistics has issued guidelines on minimum sample sizes required to obtain credible estimates. These guidelines should be considered by EPA as a way to decide when drinking water estimates should be flagged as being of lower than acceptable quality.

Taking these considerations into account, this Subcommittee believes it is appropriate that the lower and upper percentile estimates obtained from the CSFII survey be reported, but that additional guidance on their quality and when they should and should not be used should be provided.

4. Data Conventions

Charge Question 4: *Are the data conventions used to identify direct and indirect water appropriate?*

A series of conventions was established to allow the estimation of water intake as a result of water consumed as a component of foods. The procedure is described in detail and is essentially the same as that used previously by Ershow and Cantor (1989) and by the Office of Pesticides Program (Tolerance Assessment System, 1985). The procedures as described are appropriate and will allow EPA to account for moisture gained and lost during cooking and allow the estimation of the proportion

of water from home supplies versus from commercial water sources. A quick check of the results of applying the conventions to the CSFII food codes indicates that the procedures worked well. The results appear to be in the anticipated ranges. The data should be rounded to reflect the appropriate level of precision. It would also be useful to note in the text and on any files containing the factors that these represent a factor that is a composite of factors, e.g., that different types of rice, rice cooked different lengths of time and by different consumers will have different amounts of moisture and therefore different factors.

The Agency did not conduct a quality assurance check on the data. Given the multitude of uses for this information, the Subcommittee recommends that a formal QA/QC audit be conducted to ensure that the conventions were actually applied to each code as described in the methodology.

Where indirect water and intrinsic water are lost during cooking, it is necessary to determine how much is lost from each source. This is an arbitrary decision and the proposed approach seems reasonable. Validation of the estimates should be undertaken to verify the results.

5. Subpopulation Distributions

Charge Question 5: *Do the data support estimates of subpopulation distributions?*

The CSFII data were used to generate point and interval estimates of daily average *per capita* water ingestion in the manner presented in Section 8b of the EPA Report. Point estimates presented include the mean, 1st, 5th, 10th, 25th, 50th, 75th, 90th, 95th and 99th percentiles. Subpopulations defined are gender, age, region, race, economic status, residential status and certain specific female subpopulations of pregnant and lactating women of childbearing age. The results are presented in section 11 by water source and by nine sociodemographic categories.

Examination of the tables on pages 11-3 through 11-326 easily reveals many subcategories without sufficient observations to support the point estimates. For example, Table A-3b on pages 11-15 and 11-16 shows point estimates of community water intake by race and fine age category. Between the American Indians and Native Alaskans, there is only one individual under 6 months and there are only three individuals in each of three other age categories. Presenting point estimates this way will likely mislead readers. Potential users should be cautioned about the uncertainty of point estimates having small sample sizes.

Whenever possible, point estimates should be presented with confidence intervals. But due to small sample size of some subpopulations, not all confidence intervals can be computed from the data. It is not clear how many interval estimates cannot be derived from the data available to the Agency. Only Tables 1, 2, and Figure 9-20 in Section 9 include 90% confidence intervals. A survey of over 15,000 individuals should allow more confidence intervals to be calculated and presented.

Although parameter estimation, hypothesis testing, and modeling are difficult because of the complex nature of this survey, the valuable information collected deserves further exploration. A strategy should be developed to analyze, interpret, and present data on sub-populations in a systematic and meaningful way. The first set of tables presented should be for major subpopulations such as gender (male vs. female), age (infants, children, youth, adults), race (white, black, Asian/Pacific Islander, American Indian/Native Alaskan), and region (northeast, Midwest, south, west) without further subdivision. Both point and interval estimates should be provided for each category of these major subpopulations. Hypothesis testing should be carried out to see if the differences among categories are statistically significant.

In Section 11, ingestion estimates for nine sociodemographic subpopulations are presented by water source. No rationale is given for why, among all the possible combinations of major subpopulations that could have been selected, these nine combinations of sociodemographic variables were chosen for presentation. Further, without understanding the meaning and limitations of the data, over 200 pages of tables are of limited usefulness to readers. If the relative importance of various sociodemographic variables can be evaluated by modeling and hypothesis testing, cross-tabulation can be focused on a limited number of significant variables.

6. Subpopulations Included

Charge Question 6: *We have provided distributions of estimated water intake for numerous subpopulations. Should any additional subpopulations be added? Should any be excluded? Specify subpopulations.*

The Report provided distributions of estimated water intake for a relatively large number of subpopulations. As discussed earlier, the available data do not support reporting of some of the values that are placed in the tables. This does not negate the need to lay out water ingestion rates for subpopulations that might be at greater risk from drinking water contaminants. There are clearly examples that are at least as important as those reported upon. These are pointed out by the Subcommittee with the recognition that the CSFII database will not provide the needed data for such analyses. Nevertheless, the Agency is encouraged to seek better estimates of the distributions for two broad categories:

- a) Sub-populations with different lifestyles, occupations, or activities.
 - i) Infants and toddlers are not a homogeneous group. There is a population of infants in the 0-3 months of age group that receive constituted powdered formula exclusively. These infants could be consuming as much as 180-200 ml/kg/day from the same source of tap water.
 - ii) Dietary survey misses lifestyles of specific cultural groups (e.g., Native American, recent immigrants) that are still practiced

- iii) People who live in hot climate areas.
- iv) People who consume large amounts of water because of physical activity (can consume as much as 300-500 ml/kg/day)

b) Health conditions that affect water intake:

- i) Diabetes
- ii) Conditions requiring rapid rehydration needs (GI upsets, food poisoning)
- iii) Disorders of water and sodium metabolism.

The subcommittee also noted that there are aspects of water ingestion that might be better addressed by taking a physiological approach. If total water ingestion is first thought of in terms of the needs that are defined by physiological state, developmental stage, levels of activity (reflected in metabolic rates), and environmental settings a general model could be constructed. This approach will always capture the upper limit, as one can assume that all of the water that is not intrinsic to food could be derived from the tap. Then more accurate estimates of sources of the actual water consumed could be constructed from survey information that is targeted to the sub-populations of interest. This could be a more efficient way of addressing drinking water ingestion by subpopulations of interest to EPA, in particular those noted in 'a' above.

7. Indirect Water

Charge Question 7: *USDA has identified two types of indirect water in foods: a) the amount of water in food as consumed; and b) the amount of water used to prepare food. The water intake report provides estimates of the amount of indirect water in food as consumed. If resources permit, we could expand our report as a future addendum to include estimates of the amount of indirect water used to prepare food. Would this be desirable?*

The current ingestion report provides estimates of the amount of indirect water in food as consumed. The amount of water used to prepare food may be greater, owing to evaporative loss during preparation. This loss can result in a concentration of non-volatile contaminants. Such increases are chemical specific. To be able to calculate the amount of residue concentration, both the amount of indirect water in food as consumed, and the amount of indirect water used to prepare food must be known. This analysis should be limited to only those foods where the amount of water added to prepare the food is known. The amount of water which is first boiled, then added to food such as that used to prepare infant formula, is not known.

The critical question is whether preparation leads to large changes in the distribution or ingestion of water contaminants in the population. Certainly in some cases the losses of water volume could be large, but are they consistent within individual consumers. In addition, it is not clear how common a practice unattended boiling or extensive boiling might be. There are many other more important

variables that remain unaddressed with respect to sensitive populations. Consequently, pursuit of this issue should reflect programmatic priorities with respect to sensitive subpopulations.

An omission in consideration of indirect water that could be significant appears to be soft drinks prepared from syrup in restaurants, fast-food establishments, and bars. Again, the pursuit of this detail has to set within the priorities of the program. However, some initial evaluations might be made by contacting the appropriate industry representatives to obtain information on the ratio of syrup to canned/bottled soda sold.

8. Food Types Not Covered

Charge Question 8: *Additional water intake estimates associated with types of food may be useful for specific risk-exposure analyses, e.g., cold beverage intake. Such analyses are feasible using the CSFII data. We could expand our report as a future addendum if resources permit. Are any such targeted analyses of significant interest at this time?*

This question was withdrawn by the Agency during the discussions at the July 19-20, 1999 Drinking Water Intake Subcommittee meeting.

9. Intrinsic Water

Charge Question 9: *Intrinsic water is the water contained in foods and beverages at the time of market purchase. Intrinsic water includes commercial water (added to food products by food manufacturers) and biological water (found naturally in foods). Intrinsic water is not included in our current analysis. If resources permit, we could expand our report as a future addendum to include estimates of intrinsic water. Would this be desirable?*

Yes, this would be desirable, but the Subcommittee would like to point out that the Agency's use of the term "intrinsic water" is unusual. In most instances intrinsic water is that in the raw food product, not water added by processors. In some cases (e.g., NASA) the term includes both free water and metabolic water that is derived from a food. There would be some value of using another term (e.g., commercial water) to describe this category. Care would have to be taken that it is not confused with bottled water, however.

The Subcommittee felt that one advantage of including intrinsic water (as the term is used in the Agency report) in the analysis would be to enable the derivation of a fluid requirement distribution by recognizing this additional source of water. This combined direct/indirect water ingestion distribution will be less variable than direct use only, as it is closer to a biological/physiological measure than one of lifestyle. However, this is only one of a number of other sources of water ingestion that would have to be known to construct the physiological need for water for individuals under different conditions. Knowing intrinsic water does capture another tap water source, even though it may be removed from

the consumer's own tap water. The relative component of commercial water could then be calculated by examining only those products with a major contribution to one or more subpopulations. Examples are soda in cans, iced tea in cans, bottled soda, beer, milk, prepared infant formulas.

10. Other Issues

Charge Question 10: *What are the scientific limitations to the use of the water consumption estimates provided in this report (i.e., what other issues has the Subcommittee noted with the estimates that are not covered elsewhere)?*

This report needs to be viewed as a key reference for population based information on water consumption. The following are examples of the many potential uses of the information.

- a) It will be valuable to programs where consumption of water estimates are needed.
- b) It will be useful to support Agency rule making.
- c) It can be used to evaluate existing default water consumption rates and to provide new defaults for subpopulations.
- d) It can serve as a reference to compare to other data sources containing similar information.

The document needs to keep these uses in mind and the text and tables should be designed to be user friendly for these purposes. Thus many users will prefer the data summarized in a ml/kg body weight format while others will need the ml/day summary. Both formats should be provided. Keeping the uses in mind, it becomes especially important that the limitations of the Survey database and the ingestion estimates based upon it be clearly spelled out in the introduction and that the report contain only statistically valid estimates. It should be noted that some sensitive subpopulations are not in the database or cannot be identified in the database. These are identified under other charge question responses. It should be explicitly stated when data are sufficient (and give the criteria used) and when they are not. Where it is not obvious why estimates are not provided, it needs to be explained.

A use-restricting limitation is the survey design that precluded estimating water ingestion in subpopulations that either by choice, or access, utilize only one source of water for ingestion. The survey data identify and provide descriptive tables for three significant sources of ingested water; community tap water, bottled water and other (private wells, cisterns, etc). While the report provides detailed ingestion distributions for each water source within defined demographic groups, "sole source" subpopulations of water ingestion limit the utility of the report for local risk assessments. Such "sole source" ingestion distributions would be especially valuable to assessing health risks from ingestion. The overall national mean water ingestion finds community tap water contributing 75% of the water

ingested. The tables provided show that the 75% contribution is not evenly distributed over the population. A valuable statistic not provided is the percent of individuals obtaining virtually all their water from community taps or all from the other two sources and their estimated ingestion rates. Such individuals may be consuming nearly 1/3 more tap water than the national estimate provides. If the size of this population is substantial, using the national ingestion estimate to characterize contaminant exposure to this group could significantly underestimate tap water contaminant risks. This underestimate may partially be seen in Section 9, figure 2 which shows that over 47 million US residents are estimated to consume no tap water. Since these individuals require fluid to survive, they probably represent those ingesting only “other” water from private wells or bottled water only. The inclusion of these “unexposed” individuals in the ingestion estimates leads to underestimates of ingestion among those with access to the water source. The potential for underestimating ingestion is even more pronounced for infants where Section 9 figure 3 shows nearly half of the infants drank no tap water. This probably reflects the high percentage of infants being breast fed or using bottled water to mix formula. This seriously reduces the utility of the information provided on this vulnerable population. Whenever possible it would be useful to many users to have confidence intervals around the estimates.

11. Zero-Values

Charge Question 11: *The water intake estimates provided in this report are based on all respondents, including those who did not report consuming water during the two survey days. If resources permit, we could also generate estimates of water consumption which exclude the zero consumers of water. We noticed that for some sub-populations, especially the less than one-year-old infants, a substantial proportion consumed zero or minimal amounts of tap water per day (presumably those who were breast fed or drank undiluted formula or milk); these zero consumers of water can contribute to lower estimates. Would this be desirable?*

Yes, it is desirable, probably necessary, to eliminate the non-consumers of community tap water from the survey statistics for purposes of developing a set of consumption estimates for use in predicting exposure to drinking water contaminants. The DWIS suggests that those data, for which there are adequate numbers of individuals, should be displayed both ways. In other words inclusive of the population and a second display of only those individuals that are consumers of tapwater.

Based on the projections in Section 9, Figure 3, approximately 50% of the children under 1 year of age do not ingest community tap water. The mean and upper confidence limits generated from data from which these projections were made will greatly reduce the estimated ingestion rates in some groups. A rough arithmetic estimate can be made of how important this would be by recognizing that removing half of the population that does not consume water will increase the mean consumption of water in the under 1 year of age group to approximately 90 ml/kg body weight. This is roughly six times that of an adult. Thus, the differential between adults and children is at least twice that which is derived from currently utilized defaults. The subpopulation of children representing the highest tap water intake will be those fed reconstituted powdered formula. This will result in the greatest dose (per

kg) of water contaminants. There is less impact in the general population, where only about 8% of the total population does not ingest community tap water. Nevertheless, the principle is the same.

REFERENCES

- Damiano, J. and J. R. Mulhausen (1999) *A Strategy for Assessing and Managing Occupational Exposures*. (Second Edition). AIHA Exposure Assessment Strategies Committee Edited by, American Industrial Hygiene Association, Fairfax, VA.
- Ershow, A. G. and K. P. Cantor. (1989) *Total Water and Tapwater Intake in the United States: Population-Based Estimates of quantities and Sources*. National Cancer Institute. Number 263-Md-810264.
- National Center for Health Statistics Analytic Working Group. (1993) Guidelines developed by the NCHS. As cited in Life Sciences Research Office (LSRO) (1994). Third Report on Nutrition Monitoring in the United States. Interagency Board for Nutrition Monitoring and Related Research.
- Saunders, S. and B. Petersen. (1986) Introduction to the Tolerance Assessment System. U.S. Environmental Protection Agency. May.
- US EPA (1999) *Estimated Per Capita Water Consumption in the United States*. US EPA/Office of Water. Prepared under EPA Contract No. 68-C4-0046. June 1999.



AN SAB REPORT: EPA's *Per Capita* Water Ingestion Estimates for the United States

**A Review by the Drinking Water
Ingestion Subcommittee of the
Science Advisory Board**

ABSTRACT

The Drinking Water Intake Subcommittee (DWIS) of the Science Advisory Board's (SAB) Executive Committee reviewed a report on the *Estimated Per Capita Water Consumption in the United States*. The document presents estimates of drinking water ingestion for the total U.S. population and a number of subgroups of interest. Estimates are given for many age, gender, and other descriptors. The Subcommittee was pleased with the report's use of a substantial existing data base to improve upon the current EPA estimates for drinking water ingestion. The current Report is largely descriptive and contains little discussion of factors embedded within the original survey and the Agency's analytical method for deriving estimates that inform the reader of important factors that should guide use of the estimates. The Subcommittee noted its desire to see a greater level of discussion on these elements so that unintended misuse of the data can be minimized.

Keywords: Drinking water ingestion, exposure factors, drinking water consumption, drinking water intake.