BEFORE THE ILLINOIS POLLUTION CONTROL BOARD

| | |
|---|---|
| IN THE MATTER OF: | ) |
| | ) |
| WATER QUALITY STANDARDS AND | ) |
| EFFLUENT LIMITATIONS FOR THE | ) R08-9 |
| CHICAGO AREA WATERWAY SYSTEM | ) (Rulemaking – Water) |
| AND THE LOWER DES PLAINES RIVER: | ) |
| PROPOSED AMENDMENTS TO 35 ILL.. | ) |
| ADM. CODE PARTS 301, 302, 303 and 304 | ) |

**NOTICE OF FILING**

To:

John Therriault, Assistant Clerk
Illinois Pollution Control Board
100 West Randoph, Suite 11-500
Chicago, IL 60601-7447

Marie Tipsord, Hearing Officer
Illinois Pollution Control Board
100 West Randoph, Suite 11-500
Chicago, IL 60601-7447

Stefanie N. Diers, Assistant Counsel
Illinois Environmental Protection
1021 North Grand Avenue East
P.O. Box 19276
Springfield, IL 62794-9276

Persons on the attached service list

Please take notice that on the 20th Day of September, 2010, I filed with the Office of the Clerk of the Illinois Pollution Control Board the attached **Testimony of Marc Gorelick, MD.,** a copy of which is hereby served upon you.

Ann Alexander

By: ________________________________________
Ann Alexander, Natural Resources Defense Council

Dated: September 20th, 2010

Ann Alexander
Senior Attorney
Natural Resources Defense Council
2N. Riverside Plaza, Suite 2250
Chicago, Illinois 60606
312-651-7905
312-663-9920 (fax)
AAlexander@nrdc.org

## CERTIFICATE OF SERVICE

I, Ann Alexander, the undersigned attorney, hereby certify that I have served the attached **Testimony of Marc Gorelick, MD.** on all parties of record (Service List attached), by depositing said documents in the United States Mail, postage prepaid, from 227 W. Monroe, Chicago, IL 60606, before the hour of 5:00 p.m., on this 20th Day of September, 2010.

Ann Alexander

Ann Alexander, Natural Resources Defense Council

**SERVICE LIST**

Sep. 20, 2010

Frederick M. Feldman, Esq., Louis Kollias,
Margaret T. Conway, Ronald M. Hill
Metropolitan Water Reclamation District
100 East Erie Street
Chicago, IL 60611

Andrew Armstrong, Matthew J. Dunn – Chief,
Susan Hedman
Office of the Attorney General
Environmental Bureau North
69 West Washington Street, Suite 1800
Chicago, IL 60602

Roy M. Harsch
Drinker Biddle & Reath
191 N. Wacker Drive, Suite 3700
Chicago, IL 60606-1698

Bernard Sawyer, Thomas Granto
Metropolitan Water Reclamation District
6001 W. Pershing Rd.
Cicero, IL 60650-4112

Claire A. Manning
Brown, Hay & Stephens LLP
700 First Mercantile Bank Building
205 South Fifth St., P.O. Box 2459
Springfield, IL 62705-245 9

Lisa Frede
Chemical Industry Council of Illinois
1400 East Touhy Avenue Suite 100
Des Plaines, IL 60019-333 8

Deborah J. Williams, Stefanie N. Diers
IEPA
1021 North Grand Avenue East
P.O. Box 19276
Springfield, IL 62794-9276

Fredric P. Andes, Erika K. Powers
Barnes & Thornburg
1 North Wacker Drive Suite 4400
Chicago, IL 60606

Alec M. Davis, Katherine D. Hodge,
Matthew C. Read, Monica T. Rios,
N. LaDonna Driver
Hodge Dwyer & Driver
3150 Roland Avenue P.O. Box 5776
Springfield, IL 62705-5776

James L. Daugherty - District Manger
Thorn Creek Basin Sanitary District
700 West End Avenue
Chicago Heights, IL 60411

Ariel J. Tesher, Jeffrey C. Fort
Sonnenschein Nath & Rosenthal
233 South Wacker Driver Suite 7800
Chicago, IL 60606-6404

Tracy Elzemeyer – General Counsel
American Water Company
727 Craig Road
St. Louis, MO 63141

Jessica Dexter, Albert Ettinger
Environmental Law & Policy Center
35 East Wacker Drive, Suite 1600
Chicago, IL 60601

Keith I. Harley, Elizabeth Schenkier
Chicago Legal Clinic, Inc.
205 West Monroe Street, 4th Floor
Chicago, Il 60606

Robert VanGyseghem
City of Geneva
1800 South Street
Geneva, IL 60134-2203

Cindy Skrukrud, Jerry Paulsen
McHenry County Defenders
132 Cass Street
Woodstock, IL 60098

W.C. Blanton
Husch Blackwell Sanders LLP
4801 Main Street Suite 1000
Kansas City, MO 64112

Marie Tipsord - Hearing Officer
Illinois Pollution Control Board
100 W. Randolph St.
Suite 11-500 Chicago, IL 60601

James E. Eggen
City of Joliet,
Department of Public Works and Utilities
921 E. Washington Street
Joliet, IL 60431

Kay Anderson
American Bottoms RWTF
One American Bottoms Road
Sauget, IL 62201

Jack Darin
Sierra Club
70 E. Lake Street, Suite 1500
Chicago, IL 60601-7447

Bob Carter
Bloomington Normal Water Reclamation District
PO Box 3307
Bloomington, IL 61702-3307

Tom Muth
Fox Metro Water Reclamation District
682 State Route 31
Oswego IL 60543

Frederick D. Keady, P.E. – President
Vermilion Coal Company
1979 Johns Drive
Glenview, IL 60025

Mark Schultz
Navy Facilities and Engineering Command
201 Decatur Avenue Building 1A
Great Lakes, IL 60088-2801

Irwin Polls
Ecological Monitoring and Assessment
3206 Maple Leaf Drive
Glenview, IL 60025

Dr. Thomas J. Murphy
2325 N. Clifton Street
Chicago, IL 60614

Cathy Hudzik
City of Chicago –
Mayor's Office of Intergovernmental Affairs
121 N. LaSalle Street City Hall - Room 406
Chicago, IL 60602

Stacy Meyers-Glen
Openlands
25 East Washington Street, Suite 1650
Chicago, IL 60602

Beth Steinhorn
2021 Timberbrook
Springfield, IL 62702

Lyman Welch
Alliance for the Great Lakes
17 N. State St., Suite 1390
Chicago, IL 60602

James Huff - Vice President
Huff & Huff, Inc.
915 Harger Road, Suite 330
Oak Brook IL 60523

Kenneth W. Liss
Andrews Environmental Engineering
3300 Ginger Creek Drive Springfield,
IL 62711

Susan Charles, Thomas W. Dimond
Ice Miller LLP
200 West Madison, Suite 3500
Chicago, IL 60606

Vicky McKinley
Evanston Environment Board
223 Grey Avenue
Evanston, IL 60202

Traci Barkley
Prairie Rivers Network
1902 Fox Drive Suite 6
Champaign, IL 61820

Jamie S. Caston, Marc Miller
Office of Lt. Governor Pat Quinn
Room 414 State House
Springfield, IL 62706

Kristy A. N. Bulleit
Hunton & Williams LLC
1900 K Street, NW
Washington DC 20006

BEFORE THE ILLINOIS POLLUTION CONTROL BOARD

| | | |
|---|---|---|
| IN THE MATTER OF: | ) | |
| | ) | |
| WATER QUALITY STANDARDS AND | ) | |
| EFFLUENT LIMITATIONS FOR THE | ) | R08-9 |
| CHICAGO AREA WATERWAY SYSTEM | ) | (Rulemaking – Water) |
| AND THE LOWER DES PLAINES RIVER: | ) | |
| PROPOSED AMENDMENTS TO 35 ILL. | ) | |
| ADM. CODE PARTS 301, 302, 303, AND 304 | ) | |

# TESTIMONY OF MARC GORELICK, MD

## I. Introduction

My name is Marc H. Gorelick, M.D. I am a Professor of Pediatrics and Population Health and Chief of the Section on Emergency Medicine at the Medical College of Wisconsin, and Jon E. Vice Chair in Pediatric Emergency Medicine at Children's Hospital of Wisconsin. I have extensive expertise in clinical epidemiology, and have published more than 50 peer-reviewed original research papers in that field. My qualifications are set forth more fully in my previous testimony *(see* Testimony of Marc Gorelick, MD dated August 4, 2008 and May 28, 2010).

I am testifying today, for the third time in this proceeding, on behalf of Natural Resources Defense Council, Sierra Club – Illinois Chapter, Friends of the Chicago River, Southeast Environmental Task Force, and Openlands in support of the regulation proposed by the Illinois Environmental Protection Agency ("IEPA") that would require the Metropolitan Water Reclamation District ("MWRD" or the "District") to disinfect the effluent from its three wastewater treatment plants ("WWTPs") that discharge into the Chicago Area Waterway System ("CAWS").

In my initial testimony in April, 2009 ("First Testimony"), I explained the inherent limitations of epidemiological research generally, and of the Chicago Health, Environmental Exposure and Recreation Study ("CHEERS") in particular – despite the fact that its methodology is largely sound. In my subsequent testimony concerning the CHEERS technical reports in June, 2010 ("Second Testimony"), I explained why the unanalyzed data in those reports was meaningless until confounding factors and potential sources of biases were accounted for in the final study report; and explained as well the inability of any statistical analysis, in light of those biases and factors, to render the study conclusive.

Today, my testimony concerns the final CHEERS report. That report, in sum, finds that CAWS recreators are at increased risk for one type of illness (eye symptoms) associated with water recreation. As to gastrointestinal illness, the CHEERS study found significantly elevated levels of illness on both the CAWS and general use waters ("GUW") such as Lake Michigan, even though pathogen levels are higher in the CAWS. From this set of data, the District is urging the Board to disregard decades of established medical knowledge that exposure to pathogens in sewage risks illness, and conclude that steps to disinfect its effluent are unnecessary.

My testimony today sets forth the reasons why even this excellent study in no way supports the sweeping conclusion urged by the District – leaving aside the fact that the study did, in fact, identify at least one area of increased health risk. The CHEERS study itself is consistent with fundamental scientific principles, but the District's misuse of it as a conclusive definition of risk levels is not. *Any* scientific research is a cautious process of repeated study, testing of related and increasingly detailed hypotheses, and consensus building among researchers. This is doubly true of epidemiological research, which is by its very nature imprecise, reliant as it is on study subjects who are out in the world rather than in controlled laboratory conditions. For this reason, even the best epidemiologic studies unavoidably have non-trivial gaps and flaws impacting the accuracy and generalizability of their findings.

Thus, to take a single, initial epidemiologic study and treat it as definitive is flatly unscientific. That is particularly the case where, as here, the study was designed merely as a general overview of the recreating population as a whole rather than its many more vulnerable subgroups; and unavoidably rests in part on untested or uncertain analytical methods and unaccounted-for biases. And it is certainly so when there are not only negative results but positive ones as well (the eye symptoms). The sound approach, particularly when risk to human health is involved, is take steps to protect public health based on existing scientific understandings, until and unless a consensus emerges after repeated study that the risk is not material.

Specifically, and as explained in more detailed below, the CHEERS study – while itself scientifically sound – cannot support a definitive conclusion regarding CAWS recreational risk for the following reasons:

1. *Sources of bias.* Epidemiologic studies are subject to numerous sources of bias: errors in the way data are collected or recorded that can lead to misestimates of the association between exposures and outcomes. Here, the sources of bias that may impact the CHEERS study include selection bias (the study subjects may not accurately reflect the composition and behavior of the larger population), use of non-validated measures (for instance, the CHEERS methodology for quantifying an individual's level of water exposure and ingestion), recall bias (difficulty remembering illness or activity leading to exposure), and failure to account for heterogeneity among waters *(i.e.,* the fact that there are significant risk differences among various CAWS locations and general use waterbodies that were not accounted for).

2. *Lack of analysis of at-risk subgroups.* As discussed at some length in both my First and Second Testimony, a major limitation of the CHEERS study, in terms of its generalizability, is that it lacks the statistical power to draw valid conclusions concerning subgroups of users who may be at more significant risk – those with weaker immune systems than healthy adults (children, pregnant women) and those whose activities result in greater exposure (canoers, kayakers). The CHEERS study

did include some statistical analysis of subgroups, but as a peer reviewer pointed out, this analysis cannot overcome the underlying probability of too few subjects in each subgroup to draw statistically reliable conclusions.

3. *Adjustment of confounding factors.* Epidemiologic research requires adjusting for confounding factors, *i.e.*, differences among study subjects or their environment that aren't specifically being studied but that might skew the results. The CHEERS study does adjust for a set of confounding factors the researchers identified. However, as discussed in my oral testimony in June, the method by which this set of confounding factors was identified may have left out some factors that could be impacting the study results.

4. *Insuficiency of conclusions linking pathogens to GI illness.* One outcome objective of the CHEERS study was to identify specific pathogens responsible for symptoms of GI illness among recreators. The study's conclusions on this issue are unreliable for several reasons. This portion of the study lacked statistical power to draw valid conclusions; the return rate was uneven among different study groups; and it appears that a large number of people submitting samples may not even have been sick.

5. *Unaccounted-for variables.* As discussed in my earlier testimony, there are variables potentially affecting the study outcome for which data is not available. In particular, the study design is not equipped to evaluate rates of secondary transmission of illness from asymptomatic carriers; and did not gather (and could not have) sufficient data to fully analyze constantly fluctuating variables affecting pathogen levels in the subject water bodies.

6. *Insuficiency of a single study.* As summarized above, it is entirely inconsistent with scientific method to draw conclusions regarding risk from a single study – particularly in the field of epidemiology, where the study results conflict with previous research, where at least some elevated risk was identified, and where the conclusion is offered in support of lessened public health precautions.

## I. Sources of Bias

As discussed in prior testimony, epidemiologic studies are subject to numerous sources of bias, a technical term referring to errors in the way data are collected or recorded that can lead to misestimates of the association between exposures and outcomes. Bias must be addressed in the design of the study, and steps may be taken to minimize it. Even then, it is generally not possible to eliminate bias entirely, which is one of many reasons that a single epidemiologic study should never be taken as conclusive.

In reviewing the final CHEERS report, I identified the specific sources of bias discussed below. Most of these were not specifically acknowledged as biases in the CHEERS report, although several were flagged by the peer reviewers.

A. <u>Heterogeneity Bias</u>

This is perhaps the most serious potential bias, and one that I have not previously addressed. Table V-9 in the Report shows the rates of illness by location. However, there are large differences – heterogeneity – in illness rates within each of the groups depending on location within the CAWS or GUW survey area. Within the CAWS, the illness rate ranges from 38.9 illnesses per 1000 recreators at CAWS-North to 61.7 per 1000 at Main Stem – a 59% relative difference. Similarly, the rates range from 39.9 to 59.4 per 1000 in the GUW depending on location. It thus appears that some geographic areas of each of the waterways are significantly more risky than the others. Yet because the analysis treats the entire CAWS as one group, and the entire GUW as another, heterogeneity is ignored, and both waterways look more similar to each other than they truly are. This makes it harder to demonstrate differences in disease risk. Furthermore, the largest proportion of participants comes from the areas with the lowest rates of illness, suggesting that the overall risk of illness from the CAWS (compared with the unexposed group) is an underestimate.

B. <u>Use of non-validated survey questions</u>

For much information gathered in an epidemiologic study, it is reasonable to assume that respondents will provide correct and accurate answers. For example, if we ask what year someone was born in, or their address or gender, these are fairly obvious and simple questions. There are, however, questions that may appear intuitive, but for which the answers may be inaccurate. In such instances, it is both appropriate and standard practice to validate the questions through separate research prior to using the data gathered through them in an epidemiologic study.

Not all of the substantive questions asked in the CHEERS study were validated in this manner; and resulting flaws in the method of questioning may have skewed the study results. Of particular import and concern are the questions concerning water exposure. The CHEERS researchers attempted to ascertain how wet participants got during their activities, so they could adjust for that factor in determining risk. That is, if participants were getting wetter in one body of water than another, then they might be getting sicker more often because of that rather than because of higher pathogen levels, so the study needed to adjust for that. To gather the necessary water exposure data, participants were asked whether various parts of their body got wet while recreating, and then were asked to specify whether the degree of wetness was “none, sprinkle, splash, drenched, or submerged.” While this set of questions at face value seems reasonable, it raises a set of questions, all of which may impact the accuracy and usefulness of the data:

1. *Recall issues.* How well do people recognize whether each part of their body got wet? What if someone reaches over and immerses their hand in the water, and at the same time someone splashes their head or torso – would they know it happened? Will they recall each exposure accurately at the end of a several hour trip? And do people recall the extremes – *i.e.,* not getting wet at all or getting totally submerged – better than they remember the events in the middle of the range such as getting splashed?

2. *Language issues.* How well would people agree on the meaning of the terms --"sprinkle" vs. "splash," for example? Is the level of agreement on what these terms mean similar across the entire spectrum? Or are people more or less in agreement when estimating the extremes (none vs. submerged), but less so in the middle of the range? Does the accuracy of the responses differ depending on the type of the activity, the duration of the activity, or the characteristics of the respondent?

3. *Limited range issues.* Is there a range of frequency of immersion and water ingestion that the questions do not get at? These questions are addressed in the testimony submitted in this proceeding by Sharon Bloyd-Peshkin, a kayak instructor, which I have reviewed. Ms. Bloyd-Peshkin notes that the survey does not allow participants to distinguish between a single immersion and multiple immersions, or between ingestion of a single mouthful or multiple mouthfuls; and does not allow them to specify different handwashing practices on different waterways.

The CHEERS study took the data from the water exposure questions and used it to develop a "wetness score," which was then used in the confounding factors analysis. While the concept of combining these exposures into a "wetness score" seems at first blush both creative and sensible, clinical scoring systems such as this require validation to determine their properties and ultimate accuracy. For example, is a score of 4 obtained from a sprinkle to four different body parts equivalent, in terms of its contribution to risk, as a score of 4 from having one body part submerged? And how does the scaling work? That is, is the differential between 1 and 2 the same as the difference between 5 and 6, or 15 and 16 from a risk standpoint?

None of this is to say that the water exposure information here is necessarily incorrect. However, due to the sorts of issues outlined above, when developing a novel means of measuring something relatively complex, it is generally recommended that this new measure be validated; and failure to do so calls the accuracy into question. To the extent that the water exposure questions used in this study may not provide accurate information about actual water exposure, there is information bias.

The effects of this potential bias could be clinically significant. For example, if (as Ms. Bloyd-Peshkin states in her testimony) recreators in the general use waters are immersing themselves significantly more times, and swallowing significantly more water, than their counterparts on the CAWS, they may be getting sick at increased rates as a result of that variable – which would make the risk of GUW recreation appear comparable to the risk of CAWS recreation when it really is not. More generally speaking, when information is inaccurate, but similarly inaccurate across all groups – that is, if people in both the CAWS and GUW groups provide equally inaccurate information – the effect would be to make the groups appear more similar than they really are, and would therefore tend to underestimate the association.

C. Selection bias

Selection bias occurs when the study participants are selected in a way that skews the results – either by rendering separate study groups *(e.g.,* the CAWS, GUW, and UNX study groups in the CHEERS study) non-comparable, or rendering the study participants not truly representative of all the people in the population of interest. There are many possible sources of selection bias in the CHEERS study, but the most obvious is recruiting among organized recreational groups such as rowing clubs. As a result, the study may have obtained results that are closer to accurate for the group members but do not apply to the general population that might use the CAWS. For instance, participation in an organized group may be correlated with a different skill level or different risk-taking behavior, and those differences could impact the study results.

The CHEERS investigators did attempt to assess for potential selection bias by observing all of the activities occurring in the waterways and comparing those activities with the ones reported by participants (Table III-2 1). The report states that "the distribution of activities in which CAWS participants engaged was *broadly similar* to all observed CAWS uses." (CHEERS Final Report ("Report") at II-18) (emphasis added). However, notwithstanding these "broad" similarities, there were differences cited as well – specifically fewer motor boaters and more kayakers in the study – and these differences are highly statistically significant. Moreover, it appears to be clinically important. The proportion of people engaged in motor boating – which the study suggests is the highest risk activity for GI illness (Report at V-3 8)– is more than twice as high among all users as among those enrolled in the study. If motor boating is indeed a higher risk activity, the estimate of GI disease in the study population would be an underestimate. Also, given this difference between all CAWS recreators and those who agreed to participate in the study, other important differences that create selection bias *(e.g.,* age, experience level, risk-taking behavior, etc.) are also likely to be present, but were not assessed.

Another source of selection bias is location. According to Table III-1, approximately 51% of all users observed on the CAWS were recreating on CAWS-North, while 67.9% of study participants in the CAWS group were enrolled at those locations (Table V-9). Again, this difference is highly statistically significant. This means that the section of the CAWS with the lowest risk of disease is substantially over-represented in the study group compared with the population of actual users, leading to biased estimates of risk.

D. Self-reporting bias

The problem of self-reporting bias is addressed in my Second Testimony, so I will only summarize it here. People may not report accurately for numerous reasons. They may simply forget; they may remember but fail to report; they may remember and either consciously or unconsciously report incorrectly. While it is true that a prospective design is less of a problem in this regard compared to a retrospective one, it does not "prevent" recall bias as asserted in the report (Report at V-27-28). It merely reduces differential recall bias concerning questions about whether or not people have symptoms (differential recall bias refers to the fact that people with symptoms might report exposures differently than those without). It does not prevent non-differential recall bias *(e.g.,* they simply can't remember whether they dripped water on their hand, or ate undercooked meat yesterday, or kayaked 4 times last year or 5), or differential bias based on which waterway they use or what type of activity they engage in.

## II. Failure to Address Risk to Subgroups

As discussed fairly extensively in my prior testimony, risks of waterborne illness are unlikely to be uniform for everyone. Some people will be at lower or higher risk based on their personal characteristics (age, gender, underlying health status, etc.), or by the nature of the activities they engage in and the way in which they do so.

This study demonstrates many such differences. Some of the group-specific risks are surprising and contrary to existing literature *(e.g.,* the failure to find higher risk to the oldest and youngest recreators). But there are clearly a number of lower- and higher-risk subgroups among the participants. There is no question, as I have said many times previously and Dr. Dorevich has acknowledged, that this study lacks the statistical power to fully evaluate the risk to potentially more vulnerable CAWS subgroups: members of biomedically more sensitive populations; and participants in particular recreational activities that may for whatever reason be associated with higher risk (in the CHEERS study, anglers and power boaters). The sample size power calculation for this study – i.e., calculation of the number of participants necessary to obtain a statistically meaningful result – was 9,330 participants. While more than that number of people participated in the study, far less than that number fell into any of the important subgroups. Thus, the study on its face lacks the statistical power to draw any meaningful conclusions concerning those subgroups.

The study does take steps to focus on risk to specific subgroups, but these measures do not obviate the fundamental problem of lack of statistical power.

In the first instance, it is important to differentiate between analysis of *confounding* and analysis of *efect modification.* Confounders are factors that confuse efforts to isolate and quantify the risk being studied. Effect modification, on the other hand, is an actual substantive connection between a group characteristic (age, gender, activity, *etc.)* and the risk being studied, e.g., a greater risk to such group. Although multivariable analysis is used to address both confounding and effect modification, the concepts are critically different, and require different steps in the analysis. Moreover, many studies are underpowered to assess effect modification adequately Multivariable analysis of confounding does not actually isolate and assess directly the actual differences among subgroups in the particular risk being studied (in this case, exposure to the CAWS). Rather, it seeks merely to ensure that the statistical analysis is not rendered inaccurate by any differences that may exist between groups – whether relevant to the risk being studied or not -- that could potentially skew the comparisons. By way of example, if a researcher has reason to suspect either that risk of illness may actually vary with gender, or simply that the two sexes tend to answer the questions about risk differently, the researcher would want to make sure that neither gender was overrepresented in any one study group; and that the gender differences didn't somehow skew the broader results. This type of multivariable analysis thus ensures that one is comparing apples to apples, as it were. However, in adjusting for gender and its associated differences, the researcher is not actually studying whether a particular gender is *actually* more susceptible to the risk factor being studied.

However, when an effect modification actually exists, confounding analysis will merely ensure that this difference does not skew a broader statistical analysis; and may actually mask the existence of the effect modification if the risk over several subgroups is averaged when in fact that risk is different for each. For example, in the CHEERS study, it is possible that exposure to contaminated water actually increases the risk of GI illness, but only if the level of exposure involves at least submerging the head in water, which is more likely to happen when kayaking than with other activities. That would be an effect modification of kayaking. (In fact, the CHEERS data did show an elevated risk for kayaking in the CAWS vs. GUW, but that risk was determined, questionably, to be statistically insignificant – see below). In that scenario, if your exposure is any less than dunking your head, it doesn't matter how dirty the water is. Thus, simply comparing the *average* overall risk of CAWS vs. CUW recreation for all activities together, as the CHEERS study has done, would overlook the fact that specifically for the groups of recreators who occasionally dunk their heads, but only for those groups, the CAWS would be riskier.

The process of actually determining whether a particular subgroup is *in fact* at higher risk specifically due to the risk factor being studied is *interaction* analysis, which requires additional steps in the multivariable analysis: the addition of interaction terms to the model. These terms are designed to evaluate whether the risks of one factor are heterogeneous (differing) among groups associated with another factor. For example, they might include an interaction between waterway and age to see if the risks of recreating on the CAWS differ in older versus younger people, or between waterway and type of activity to see if the risks of recreating on the CAWS might only apply to certain types of activity.

The CHEERS study did perform interaction analysis on various factors, and concluded that they were not statistically significant. But this effort, while interesting, is fundamentally limited by the lack of statistical power discussed above, which is necessary to actually obtain a meaningful result regarding these differences. The researchers found that none of the interactions was statistically significant. However, a peer reviewer appropriately observed, "Keep in mind these [tests for heterogeneity] have low statistical power and some authors advocate $p<0.2$ to describe heterogeneity." (Appendix D, page D-10) In lay terms, the reviewer was pointing out that investigators have set the bar too high for including these interactions and establishing separate risks for different subgroups due to the lack of statistical power for these subgroups. Another reviewer noted that there appeared to be elevated risk for CAWS users among some groups, but that these differences were ignored because they did not rise to the level of statistical significance. *(Id.)* This problem is, of course, particularly pronounced when the subgroups are small, in which case the statistical power is comparably reduced. For these reasons, experts sometimes recommend including interactions between factors for which there is strong underlying scientific evidence even when the interaction term analysis conducted in the study (limited as it may be by lack of statistical power) concludes that the interaction is not statistically significant.

Thus, it is unlikely that the CHEERS study can provide a satisfactory answer to the question of risks to subgroups. As large as it is overall, the study was simply not powered to address this question.

## III. Incomplete Adjustment for Confounding Factors

I addressed the issue of confounding extensively in my Second Testimony. As I anticipated, and one would expect in any credible epidemiologic study, most of these factors were adjusted for in the final Report. (I note that the actual analysis did in the end adjust for both year and season. I had referenced the need to adjust for these factors in my Second Testimony, and they were not included on the list of confounders found in Table V-2.)

The methods used by the investigators to control for confounding are sound and sophisticated. However, aside from the possibility of residual confounding due to factors not considered, I note in addition that the method for excluding some confounders from further analysis is open to question. As discussed earlier in oral testimony, the study team proposed a long list of potential confounders that was narrowed down through a screening method. When choosing which of these potential confounders to include in the multivariate analysis, the researchers screened by examining the association between each of the confounders and the outcome on a one-by-one (bivariate) basis. Although this is often done, it is potentially risky, as the very nature of confounding means that the association among variables is skewed. That is, because multiple confounders may act in concert, an association between one variable and another may not appear significant unless you account for the other confounders. Such bivariate screening risks missing potential confounders and failing to account for them in the multivariate analysis. In addition, any variable for which there is a strong biological reason to consider confounding should be included in the final analysis regardless of the results of the one-by-one screening. I raise this issue not to suggest that the results of the confounding analysis would necessarily be different if bivariate screening were not used. Rather, it illustrates the point that no study is perfect, epidemiology is not an exact science, and hence no single study should be treated as conclusive.

I note, in addition, that while the researchers did decide to address year and season as confounders, as I had recommended in my Second Testimony, it does not appear that they addressed several other confounders that I identified, including socioeconomic status, hand washing behavior, and duration of activity.

## IV. Analytical Problems with Pathogen/GI Illness Correlation

The primary objective of the CHEERS study, objective 1, was to determine the rates of acute GI and non-GI illness attributable to CAWS recreation. A secondary outcome was to identify pathogens responsible for symptoms of GI illness among recreators. I recognize that this secondary objective might be of less direct relevance than objective 1 to assessing CAWS risk. However, for purposes of further illustrating the inevitable imperfections in even the best epidemiologic studies, I note that the data and analysis associated with this secondary outcome is problematic for a number of reasons.

First, the study was not powered for this secondary objective. Any failure to find a difference in pathogens between study groups in such an underpowered study would be preliminary at best. Second, the rate of return of stool specimens, while good by the standards of such large epidemiologic studies, is still inadequate to draw firm conclusions. The fact that the return rates differed between study groups leads to possible bias. Finally, I note that, as a general matter, it is notoriously difficult to isolate viral pathogens from stool samples, depending on the pathogen. Although I am not a clinical microbiologist or virologist, in my clinical experience a very large proportion of individuals with clinical gastroenteritis never have a pathogen identified. This is reported even from studies of disease outbreaks where extensive efforts are made to find the responsible organisms (Reynolds, 2008). Finally, the selection of subjects to submit stool samples was not consistent with the study aims. All participants who developed any new GI symptoms were asked to submit a stool sample. This was 2467 people, or 22.4% of all study participants. However, only 431 people (4.1%) met the study definition of acute GI illness. Presumably, the other 2036 people had GI symptoms that were not severe enough to meet the definition of AGI. In other words, most of the testing for pathogens was in people who may not, in truth, have been ill. It should not be surprising that many of these tests were negative. It would be preferable to limit the stool testing to those who actually had acute GI illness as they defined it, but of course then the numbers would truly be too small to be meaningful. Many of the peer review comments were highly critical of this portion of the study.

## V. Additional Unaccounted-for Variables

In my First Testimony, I described several variables that are not directly addressed in the CHEERS study, but may nonetheless influence its outcome. I will not repeat that testimony in its entirety but will summarize it here as it pertains to the final Report.

The point of these observations is not to argue that the study should have been performed differently in order to account for these variable, as such would likely not have been possible. I merely cite them in support of my larger point that even the best epidemiologic studies are imperfect and inconclusive.

### A. Asymptomatic Illness

The CHEERS study is based on self-reporting by participants of symptoms. However, many of the types of pathogens associated with sewage contamination are frequently asymptomatic – that is, a water recreators can become infected with that pathogen, and can pass it along to other people (who might or might not develop symptoms). That secondary illness would not be reflected at all the survey data concerning illness rates. It is impossible to even estimate the effect of this information gap on the ultimate results, as there is no basis to know whether asymptomatic types of illness may be more prevalent in one study group or another.

While the CHEERS study follow-up survey did ask about illness in certain people living in close proximity, this information is of limited usefulness in bridging this information gap. The study obviously could not delve any further into the source of illness in non-study participants; and the survey did not ask at all about people who may have come into close contact with infected recreators but do not live in the same household – *e.g.,* children in the care of an infected day care provider.

B. Varying water conditions

The level of sewage-related waterborne pathogens inherently varies widely over time and distance. The levels of any given pathogen will vary with such ever-changing factors as water temperature, sunlight, and distance from the source. Each of these factors may affect different pathogens in a different manner. For instance, some pathogens survive much longer and hence can be harmful further downstream than others. Some pathogens may be more sensitive to temperature than others.

The CHEERS study partially, but incompletely, addresses this issue. The researchers did attempt to sample the water reasonably closely in time and place to recreational use. However, the 6-hour pathogen sampling frequency – while entirely reasonable for a study such as this one – was not nearly sufficient to get at the kinds of constantly changing differences in pathogen levels that can occur in these waterbodies. Thus, it is entirely possible that there is a set of conditions in which the pathogen levels will be very dangerous – e.g., a hot time of day coinciding with no sunlight near the pathogen source – but the study does not reflect how many, if any, participants were actually exposed to those conditions. Once again, this factor has the potential to create false negative results. The recreators who encountered the more hazardous conditions may have reported higher levels of illness, while those who encountered the less hazardous conditions would have reported lower levels of illness; and the results from the two groups would effectively have cancelled each other out.

## VI. Overall Insufficiency of a Single Study

In the previous sections, I have explained why the CHEERS study, although very good, is not perfect. However, it is essential to bear in mind that even if there were a "perfect" epidemiologic study, it could not provide conclusive evidence to guide a policy decision. As noted by Rothman and Greenland in their book Modern Epidemiology, "All the fruits of scientific work, in epidemiology or other disciplines, are at best only tentative formulations of a description of nature, even when the work itself is carried out without mistakes." This is even more true for epidemiologic research, which is carried out not under laboratory conditions but in the real world, where it is subject to the types of knowledge gaps, bias and confounding detailed in this testimony. This is not to say that epidemiologic evidence is not valuable or even crucial in public health decision making. However, the inherent limitations of epidemiologic research make it critical that the results of an observational study be interpreted and understood only in the broader context of what is already known to science about a particular problem.

There are two particular limits to epidemiologic research, independent of any errors it may contain, that I would emphasize here: the need to interpret it in the context of existing biomedical knowledge, particularly when it conflicts with that knowledge; and its limited generalizability.

A. Need to Interpret Conflicting Research in Context

When interpreting epidemiologic results, not only must any prior epidemiologic studies be considered, but underlying biomedical information must also be considered, regardless of whether there are other epidemiologic studies. This is all the more important where the results a particular epidemiology study appear to contradict either established biomedical knowledge or prior epidemiologic research.

Because of the frequency of conflict in epidemiologic results, it is unusual for there to be only a single epidemiologic study related to a given issue. The results of these studies often differ in important ways, and even the most carefully done studies may be followed by an equally well-done study with contrasting results. Who has not read of a study showing some food to be associated with a particular health risk, only to find weeks later that another study has found the opposite result? Reconciling conflicting studies requires repeated study and careful consideration of the ways in which the studies may have differed. Although one can speculate on the reasons why the findings may have differed, in such situations it is almost always necessary to conduct yet more research to resolve the question.

In this case, it is medically well known that pathogens in water can cause disease, and that exposure to more pathogens (either from a greater concentration of pathogens in the water or greater exposure to the contaminated water) generally produces an increased risk of illness. The CHEERS study found indicators of sewage contamination to be higher in the CAWS than in the GUW, such that one would expect more illness in CAWS users. Yet the rates of illness were similar among recreators in both types of water. Even the peer reviewers found this surprising. One can speculate on a variety of reasons, some of them quite plausible, but it s still speculation. The CHEERS study does not provide the data necessary to decide which of these or any of a number of other hypotheses is correct. While the peer reviewers noted that these findings are contrary to what would be expected, they comment not that these findings overturn what we know about pathogen exposure and illness, but rather that these findings make one “begin to question” certain assumptions. (Report Appendix D-2)

. In addition to the conflict with existing biomedical knowledge concerning pathogen risk, the CHEERS study authors acknowledge that the findings of similar risk in sewage impacted (CAWS) versus non-impacted (GUW) waters are contrary to a study by Fewtrell et al. (1992). The authors speculate that this may be due to differences between whitewater and calm water. While plausible, this is pure speculation, and cannot be addressed by the data presented. They also note that their findings of a lower risk of illness in the youngest and oldest subjects is contrary to a prior study of swimming (Wade et al., 2008). Indeed, it is generally known that children and the elderly are at increased risk of infectious diseases including waterborne illness, albeit not specifically in the context of secondary contact recreation on the CAWS (Gerber et al., 1996). Are these other studies wrong, or is there something different about this particular scenario? Either is certainly possible, but it would be exceedingly difficult to conclude the former based on this study alone. Most good studies raise as many questions as they answer when put in the context of existing knowledge, and the CHEERS study is no exception. In my role as a peer reviewer and editor for several journals, I have frequently encountered this situation of a study that is contrary to the existing literature, and in those cases if the paper is accepted, an editor will commonly solicit an accompanying editorial to discuss the controversy.

It bears noting, in this regard, that the CHEERS study did in fact confirm existing medical knowledge regarding the overall relationship between recreational water exposure and GI illness identified in previous studies. The anomaly between this demonstrated correlation on the one hand, and the failure to identify a relationship specifically between wastewater contaminated water exposure and GI illness on the other, makes it all the more important that the anomaly be further studied rather than simply accepted as established fact. This is particularly so given that the rates of GI illness identified in the study as being associated with recreational water use were significantly higher than EPA's benchmark for acceptable recreational risk.

Specifically, in the CHEERS study, rates of GI illness were higher for users of both the CAWS and GUW for water recreation than for those who did not recreate on water (unexposed). After adjusting for many differences between the groups, the odds of developing acute GI illness were found to be 41% higher for CAWS users and 44% higher for GUW users. The authors estimate that recreating on the CAWS would lead to an estimated 12.5 cases of acute GI illness per 1000 people on the CAWS, and 13.4 in GUW. Both numbers are higher than the EPA threshold of 8 cases per 1000. These findings of increased illness following water recreation are consistent with other studies showing increased GI illness related to kayaking (Fewtrell, 1992) and swimming (Wade, 2008). In addition, as noted previously rates of eye infection were higher among CAWS users than both unexposed participants and those exposed to GUW waters.

I note as well that with respect to disease severity, the illnesses in this study are comparable to, or greater than, those reported from waterborne disease in other studies. For example, in the 1993 Cryptosporidium outbreak in Milwaukee, WI, considered the worst waterborne disease outbreak in US history, 6.5% of victims sought healthcare for their symptoms, compared with approximately 14% of the CAWS users who developed GI illness (MacKenzie et al., 1994). More than half required an over-the-counter medication for symptoms, compared with 31% of victims in a study by Frosst *et al.* (2006).

B. Limits to Generalizability

Epidemiologists are always careful about generalizing their results to the broader population. One limitation to observational studies that may lead to conflicting results is that the population in one study may differ from that in another in important ways. Indeed, the population included in a study may not even be representative of the setting in which it was conducted – the so-called target population. In that case, the ability to generalize the results from an otherwise valid study is particularly threatened.

Here, as noted above, the CAWS users who actually agreed to participate in the CHEERS study differ in the types of water activities they engage in than the overall population of CAWS users. There may be many other ways in which they differ as well. Again, I do not single out this study for being especially unrepresentative – it may even be better in this regard than other studies. But the identified differences underscore the need to be cautious about generalizing from a single study, and the need for replication that is essential in science.

## Conclusion

The CHEERS study provides some interesting new data concerning the risks of illness associated with water recreation on the CAWS and other waterways. However, although the study is well-designed and well-conducted, its findings – or indeed those of any epidemiologic study – cannot be considered conclusive, or in this instance even firmly negative. The study findings confirm that water recreation activities on the CAWS are associated with an increased risk of illness, in one case specific to the CAWS (eye symptoms). The study's other findings concerning comparable GI illness rates in the CAWS and GUW are contrary to established medical knowledge and the findings of other epidemiologic studies. Further research will therefore be necessary to explain these differences, as well as to establish risks to important subgroups. In the meantime, existing well-established biomedical knowledge concerning the risks associated with sewage-related pathogens should govern any decision concerning protection of public health. The results of the CHEERS study do not tip the balance of evidence away from the need to disinfect effluent flowing into the CAWS to reduce human health risks.

Marc H. Gorelick, M.D.

## References

Fewtrell L, Godfree AF, Jones F, Kay D, Salmon RL, Wyer MD. 1992. Health effects of whitewater canoeing. Lancet 339(8809): 1587-1589.

Frosst GO, Majowicz SE, Edge VL. Factors associated with the use of over-the-coutner medications in cases of acute gastroenteritis in Hamilton, Ontario. *Can J Pub Health* 2006;97:489-93.

MacKenzie WR, Hoxie NJ, Proctor ME et al. A massive outbreak in Milwaukee of Cryptosporidium infection transmitted through the public water supply. *New Eng J Med* 1994;331:161-167.

Reynolds KA, Mena KD, Gerba CP. Risk of waterborne illness via drinking water in the United States. *Rev Environ Contam Toxicol* 2008; 192:117-158.

Wade TJ, Calderon RL, Brenner KP, et al. High sensitivity of children to swimming-associated gastrointestinal illness: results using a rapid assay of recreational water quality. *Epidemiology* 2008;193: 375-383.