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..	)	Subdocket B
ADM. CODE PARTS 301, 302, 303 and 304	)	

NOTICE OF FILING

To:

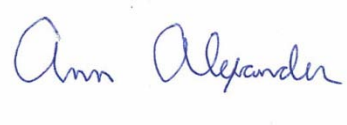
John Therriault, Assistant Clerk
Illinois Pollution Control Board
100 West Randoph, Suite 11-500
Chicago, IL 60601-7447

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 28th Day of May, 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.



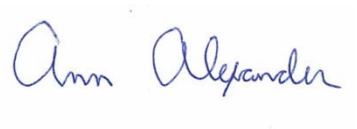
By: _____
Ann Alexander, Natural Resources Defense Council

Dated: May 28th, 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 28th Day of May, 2010.



_____

Ann Alexander, Natural Resources Defense Council

Service List

Richard J. Kissel and Roy M. Harsch
Drinker, Biddle, Gardner, Carton
191 N. Wacker Drive, Suite 3700
Chicago, IL 60606-1698

Bernard Sawyer and Thomas Granto
Metropolitan Water Reclamation District
6001 West Pershing Road
Cicero, IL 60650-4112

Deborah J. Williams and Stefanie N. Diers
Assistant Counsel, Division of Legal Counsel
Illinois Environmental Protection Agency
1021 North Grand Avenue East
P.O. Box 19276
Springfield, IL 62794-9276

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

Kevin G. Desharnais, Thomas W. Diamond
and Thomas V. Skinner
Mayer, Brown LLP
71 South Wacker Drive
Chicago, IL 60606-4637

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

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

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

Matthew J. Dunn, Chief
Office of the Attorney General
Environmental Bureau North
69 West Washington, Suite 1800
Chicago, IL 60602

Katherine D. Hodge and Monica T. Rios
Hodge Dwyer Zeman
3150 Roland Avenue
P.O. Box 5776
Springfield, IL 62705-5776

Charles W. Wesselhoft and James T. Harrington
Ross & Hardies
150 North Michigan Avenue
Suite 2500
Chicago, IL 60601-7567

Margaret P. Howard
Hedinger Law Office
2601 South Fifth Street
Springfield, IL 62703

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

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

William Richardson, Chief Legal Counsel
Illinois Department of Natural Resources
One Natural Resources Way
Springfield, IL 62702

Fred L. Hubbard
Attorney at Law
16 West Madison
P.O. Box 12
Danville, IL 61834

Lisa Frede
Chemical Industry Council of Illinois
2250 E. Devon Avenue
Suite 239
Des Plaines, IL 60018-4509

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

Sharon Neal
Commonwealth Edison Company
125 South Clark Street
Chicago, IL 60603

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

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

Georgie Vlahos
Naval Training Center
2601A Paul Jones Street
Great Lakes, IL 60088-2845

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

Dennis L. Duffield
Director of Public Works & Utilities
City of Joliet, Department of Public Works & Utilities
921 E. Washington Street
Joliet, IL 60431

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

Ann Alexander, Senior Attorney
Natural Resources Defense Council
2 N Riverside Plaza, Suite 2250
Chicago, IL 60606

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

Beth Steinhorn
2021 Timberbrook
Springfield, IL 62702

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

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

Susan M. Franzetti
Nijman Franzetti LLP
10 S. LaSalle Street, Suite 3600
Chicago, IL 60603

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

Albert Ettinger, Senior Staff Attorney, and Jessica Dexter
Environmental Law and Policy Center
35 E. Wacker Drive, Suite 1300
Chicago, IL 60601

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

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

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

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

Jeffrey C. Fort and Ariel Tescher
Sonnenschein Nath & Rosenthal LLP
7800 Sears Tower
233 S. Wacker drive
Chicago, IL 60606-6404

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

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

Susan Hedman and Andrew Armstrong, Environmental Counsel
Environnmental Bureau
Office of the Illinois Attorney General
69 West Washington, Suite 1800
Chicago, IL 60602

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

Bob Carter
Bloomington Normal Water Reclamation District
P.O. Box 3307
Bloomington, IL 61702-3307

Ronald M. Hill and Margaret T. Conway
Metropolitan Water Reclamation District of Greater Chicago
100 East Erie Street, Room 301
Chicago, IL 60611

Frederic P. Andes, Carolyn S. Hesse and David T. Ballard
Barnes & Thornburg LLP
One North Wacker Drive, Suite 4400
Chicago, IL 60606

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.	)	Subdocket B
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.

I am testifying today, for the second time in this proceeding, on behalf of Natural Resources Defense Council, Environmental Law and Policy Center, Sierra Club – Illinois Chapter, Friends of the Chicago River, 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 previous testimony in April, 2009, I explained the severe limitations of epidemiological research, which I have conducted extensively myself, as a means of assessing a public health risk. I further explained the limited significance of negative epidemiological study results, *i.e.* a failure to find elevated risk, particularly in a study such as this one with many diverse variables and confounding factors (age, health, type of activity, *etc.*).

My testimony today concerns the preliminary technical reports submitted by MWRD describing raw data collected by researchers in the epidemiological study commissioned by the District, the Chicago Health, Environmental Exposure, and Recreation Study ("CHEERS"). My review of the CHEERS preliminary data indicates that the concerns I expressed in my 2009 testimony, concerning the scope of the study and inherent ambiguity of any negative result, are materializing. Equally important to recognize, however, is that this raw data does *not* represent CHEERS study results – negative or otherwise – or anything approximating them. They are merely the first step in an epidemiological study, a collection of facts and numbers obtained from testing and study subject interviews. The next critical step is evaluation of the data through statistical analysis and mathematical modeling in order to isolate the specific risk factors the study is designed to evaluate. Without that step, the data, while intriguing, are essentially meaningless. I strongly urge the Board not to consider the technical reports as a basis for its decisionmaking in this matter.

II. Qualifications

I am an expert in epidemiology and public health. A copy of my curriculum vitae is attached as Exhibit 1. A biographical sketch summarizing my work and expertise in epidemiology is attached as Exhibit 2.

My current professional positions include the following:

- Professor, Departments of Pediatrics and Public Health, Medical College of Wisconsin (2004-present).
- Chief, Section of Pediatric Emergency Medicine, Department of Pediatrics, Children's Hospital of Wisconsin (2000-present).
- Jon E. Vice Chair in Pediatric Emergency Medicine, Children's Hospital of Wisconsin.
- Associate Director, Children's Research Institute, 2007-present.

I have had numerous faculty appointments in the field of epidemiology, including the following:

- Assistant Professor, Departments of Pediatrics and Epidemiology, University of Pennsylvania School of Medicine (1994-1998).
- Senior Scholar, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania School of Medicine (1994-98).
- Adjunct Assistant Professor, Department of Epidemiology, University of Pennsylvania School of Medicine (1998-2000).
- Associate Professor, Departments of Pediatrics and Epidemiology, Medical College of Wisconsin (2000-2004).

I have conducted extensive published research in the area of epidemiology. I have co-authored more than 50 peer-reviewed original research papers publications in different areas of clinical epidemiology, including case-control and cohort studies, controlled clinical trials, and meta-analyses. Some representative publications include:

Gorelick MH, Shaw KN, Murphy KO. Validity and reliability of clinical signs in the diagnosis of dehydration in children. *Pediatrics* 1997;99(5):e6.

Gorelick MH, Shaw KN. Clinical decision rule to identify young febrile children at risk for UTI. *Archives of Pediatrics and Adolescent Medicine* 2000;154:386-390.

Gorelick MH, Brousseau DC, Stevens MW. Validity and responsiveness of a brief asthma-specific quality of life instrument in children with acute asthma. *Ann Asthma Allerg Immunol* 2004; 92:47-51.

Gorelick MH, Meurer J, Walsh-Kelly C, Brousseau DC, Cohn J, Kuhn E, Grabowski L, Kelly K. Controlled trial of two emergency department-based follow-up interventions to improve asthma outcomes in children. *Pediatrics* 2006;117:S127-S134.

Gorelick MH. Bias arising from missing data in predictive models. *J Clin Epidemiol* 2006;59:1115-23

Gorelick MH, Yen K. The kappa statistic was representative of empirically-observed inter-rater agreement for physical findings. *J Clin Epidemiol* 2006;59:859-861.

Gorelick MH, Alessandrini EA, Cronan K, Shults J. Revised Pediatric Emergency Assessment Tool [RePEAT]: a severity index for pediatric emergency care. *Acad Emerg Med* 2007;14;316-323.

Redman R, Nenn C, Eastwood D, **Gorelick MH**. ED visits for diarrheal illness increased after release of undertreated sewage. *Pediatrics* 2007;120:e1472-1475.

Gorelick MH, Wagner D, McLellan S. Validation of a questionnaire to evaluate water exposures in children. *Ambul Pediatr* 2008;8:388-91.

Freedman SB, Eltorky M, **Gorelick MH**, and the PERC Gastroenteritis Study Group. Evaluation of a gastroenteritis severity score for use in an outpatient setting. *Pediatrics* 2010;125: doi:10.1542/peds.2009-3270

Gorelick MH, McLellan SL, Wagner D, Klein J. Water use and acute diarrhoeal illness in children in a United States metropolitan area. *Epidemiol Infect* 2010 [accepted for publication]

Drayna P, McLellan SL, Simpson P, Li S-H, **Gorelick MH**. Association between rainfall and pediatric emergency department visits for acute gastrointestinal illness. *Env Health Persp* 2010 [accepted for publication]

I have extensive teaching experience in the area of epidemiology. Course I have taught in that area include the following:

- University of Pennsylvania: Course developer and director, Advanced Topics in Clinical Epidemiology (elective course for Master of Science in Clinical Epidemiology Program); taught in Critical Appraisal workshop for MSCE students.
- Jefferson Medical College: developed and taught course in Evidence-Based Medicine for senior pediatric residents.
- Medical College of Wisconsin: Annual Introduction to Research Design seminar for pediatric fellows; taught in Protocol Development course for MCW fellows and junior faculty; preceptor for K30 Clinical Research Scholars Program

III. Nature and Significance of the CHEERS Technical Report

I have reviewed the document entitled "CHEERS Research Update, an Interim Technical Report Prepared for Submission to the Illinois Pollution Control Board, and Appendices" ("Technical Report"). The Technical Report is a preliminary description of data gathered in the CHEERS epidemiologic study, which has been conducted by a team led by Dr. Samuel Dorevitch of the University of Illinois-Chicago School of Public Health for the MWRD, which provided the funding for this study.

This compilation of information represents completion of the initial step in conducting an epidemiological study: a population survey and gathering of related information. This initial step, reflected in the Technical Reports, is essentially a compilation and description of raw data. The next step is critical, and in many ways at the heart of sound epidemiologic research: evaluation of the data through statistical analysis and mathematical modeling in order to isolate the specific risk factors the

study is designed to evaluate. In the absence of that step, the preliminary data has very limited meaning.

As discussed in my 2009 testimony, even after this second analytical step – the "number crunching" step as it were – has been completed, there are still likely to be many factors unrelated to the risk being assessed that confuse efforts to isolate and quantify that risk. These are referred to as "confounding factors," which I address in more detail in the next section. However, without the critical step of statistical analysis to attempt to strip away some of the impact of these confounding factors, the preliminary data has little meaning at all. Indeed, knowing that there are very large differences in numerous risk factors between the groups under study, it is inappropriate and misleading to draw any conclusions about those groups without properly accounting for these other factors. Moreover, it is essential to understand that the process of disaggregating the confounding factors will in most instances reduce the statistical power of the study, and hence the significance and reliability of its results. This problem is also addressed in more detail in the section below.

As discussed in my 2009 testimony, confounding factors abound in the CHEERS study. The study's design is extraordinarily broad. Rather than zeroing in on a particular population of recreators – for example, kayakers, or canoeists, or children, or healthy adults – the study casts its net widely, gathering data on *all* CAWS recreators (and many others) participating in a broad array of activities, regardless of age or health. All of these differences among study participants, among many other factors, constitute confounding factors that must be accounted for in further analysis – and which will reduce the statistical power of the study.

Thus, there is no sound scientific basis for drawing conclusions from the Technical Report regarding risks associated with CAWS recreation, and it would be a serious error to attempt to do so. Indeed, the authors of the Technical Report acknowledge this on page 33 of the report, where they state, "It must be emphasized that these comparisons do not account for differences in the demographic and other characteristics of the three groups highlighted in Chapter 4…. Thus, firm conclusions can not be drawn from these data regarding differences in AGI across groups or recreational water exposure as a cause of AGI." For this reason, I was surprised and dismayed to see that the District has claimed publicly that the Technical Reports represent a finding that "there are no increased health risks for recreational users in the inland Chicago Area Waterways System (CAWS) compared to swimmers in Lake Michigan."[1] There is absolutely no basis in the preliminary Technical Report data to reach that conclusion. Moreover, as described in the next section, it appears that the confounding factors I identified in my 2009 testimony – plus a number of additional factors and potential statistical biases that have emerged – will preclude a sweeping conclusion of that nature even when the study analysis is complete.

[1] *See* http://www.mwrd.org/irj/go/km/docs/documents/MWRD/internet/News&Media/Newsroom/Media/Press%20Releases/May/2010/CHEERS_study_filing.pdf (last accessed May 24, 2010).

IV. Confounding Factors Reflected in the Technical Report Raw Data

The CHEERS study represents an admirably ambitious attempt to conduct a broad survey of health impacts of all CAWS recreational uses and all categories of users. While these study results may ultimately yield an interesting first look at the health of CAWS recreators, as I explained in my 2009 testimony, the study is inherently not capable of providing a statistically meaningful assessment of health risk to the many sub-populations subsumed in the study.

A review of the preliminary Technical Report data indicates that many of the issues I identified in 2009 regarding the scope of the study and its statistical significance are materializing. Specifically, the data reflects numerous very significant confounding factors that have not yet been addressed through statistical analysis, which when sorted out will significantly reduce the statistical power of the study – to the extent they can be sorted out at all, which is not always the case. In addition, I have identified potential statistical biases that have not been accounted for, and that will weaken the reliability of the study results.

Those confounding factors, biases, and impacts on the statistical power of the CHEERS study (and on the significance of the raw Technical Report data) are discussed in the sections below.

A. Many Confounding Factors Are Reflected in the Technical Report Data

All observational epidemiologic studies, where subjects do what they would normally do rather than being assigned to those activities, are subject to confounding factors. This refers to the fact that individuals who differ with respect to the factor of interest (in this case, CAWS exposure) may also systematically differ with respect to other factors. It may appear that the factor of interest is associated with the outcome (in this case, GI illness), when in fact it is the other factors that are really the cause. For example, cigarette smoking is known to cause lung cancer. If one compares drinkers with non-drinkers, one might find a higher rate of lung cancer in those who drink, and conclude that drinking also causes lung cancer. However, if people who drink are also more likely to smoke, then it might *appear* that drinking causes cancer when in fact it is the smoking that does so. We say that the association between drinking and cancer is confounded by smoking.

Confounding can thus occur any time there are differences in factors other than the one under study, and those factors can themselves be associated with the risk of the outcome of interest. Confounding may also work to obscure an association. For example, if adults who use the CAWS are generally younger and hence statistically healthier than those who use Lake Michigan, then a simple comparison of those who use the CAWS would reveal that illness is less likely – not because the water is safer, but because those who go on it are less prone to get sick. One could falsely conclude that the risk from the CAWS is lower than it actually is unless you account for the difference in age as it relates to overall health.

In the CHEERS study, the researchers hope to compare people who recreate on the CAWS, those who recreate on General Use Waterways (GUW), and those who recreate but not on water, to draw conclusions about the relative risk of illness from CAWS and GUW compared with no water exposure. However, the characteristics of the study participants on pages 24-30 show that there are large and important differences in many other factors that are very likely to affect the risk of acute illness. These include:

Year of enrollment. For example, only 30% of the CAWS users were enrolled in 2009, compared with 40% of the GUW users. Water quality varies by year, and risk of illness is therefore likely to vary by year as well.

Season. For example, only 14.5% of CAWS users were enrolled in spring (March-May), compared with 30% of GUW users and almost 45% of unexposed recreators. Water quality varies by season, and risk of illness is therefore likely to vary by season as well. Additionally, other causes of illness are more prevalent in the community at certain times of year. For example, rotavirus, the most common cause of gastroenteritis, is usually present at high frequency in the community in the spring. If more GUW and unexposed recreators are enrolled in spring, their risk of illness will be overstated, diminishing the apparent relative risk to CAWS users.

Gender. 50% of CAWS users were male, compared with almost 60% of GUW users. It is well known among epidemiologists that gender may affect not only risk of illness but reporting of illness.

Age. As shown in Table IV-4, the age distribution differs significantly among the three groups. Age is known to be associated with risk of illness. People 18-44 are on average least susceptible to infectious diseases, but they are overrepresented in the CAWS group. This could serve to underestimate the association between CAWS use and illness in an unadjusted analysis.

Race/ethnicity. As shown in Table IV-5, the racial and ethnic characteristics of the three groups are different. Race and ethnicity are often associated with economic status, which may in turn affect the risk of GI illness.

Water activity. The types of activities performed by CAWS and GUW users differ significantly, and these also vary by year. Type of activity affects the likelihood of water exposure, and would therefore affect the risk of waterborne illness.

These are merely examples. Other confounders that must be addressed include the duration of activity (*i.e.* people on the water for a longer time will have greater exposure), and post-activity behavior (*i.e.*, people who eat or drink immediately after recreation will have greater exposure, and people who wash up immediately after recreation will have less exposure).

Given the large number of actual and potential confounders, and the very large differences between the exposure groups with regard to these factors, the simple, unadjusted analysis presented in this preliminary report, which in no way accounts for this confounding, is essentially meaningless. Even when the adjustments are made to the extent they can be, as discussed below, these adjustments will negatively impact the statistical power of the study.

B. Numerous Potential Sources of Bias

In addition to the problem of confounding, epidemiologic studies generally – and the CHEERS study is no exception -- are prone to various sources of bias, or systematic error. An important source of bias is called information bias, which arises when the accuracy of information is compromised. If the accuracy of information differs between groups, then a researcher may falsely conclude there is a difference between groups when in truth there is none, or to falsely conclude that the two groups are similar when in fact they are different. For example, if CAWS users perceive

their exposure to be riskier, they may be more likely to report symptoms they believe to be related to the water than unexposed participants. This would tend to inflate the apparent risk of illness from the CAWS. Alternatively, if CAWS users are concerned about losing access to the waterway if they report illness to the research team, they may hold back on reporting, thereby making the CAWS seem safer than it truly is. Similarly, if CAWS users are more concerned about potential illness from contact with water, they may be more aware of such contact and hence more likely to report it. Such bias, known as recall bias, is a well-known problem in epidemiologic studies. The fact that CAWS users reported much higher rates of exposure to water suggests that CAWS users may be reporting differently than GUW users.

Additionally, in any study such as the CHEERS study that is based on after-the-fact participant self-reporting, the quality of information may suffer. For example, the longer the period of time that passes between an illness and being questioned about it, the less accurate the information is likely to be. Some people may have had diarrhea but forgotten about it, or may not be able to recall the exact day they became sick when asked 3 weeks later. When parents are asked about illness in their children, they may not know whether the child had loose stools or not. When this happens – some people over-report while others under-report - then groups of people will seem more similar to each other than they really are (the extreme example would be if everyone just flipped a coin when answering), and this always produces a bias such that any association between the exposure and the outcome will be underestimated. Another important example of potential bias in this study is the averaging of microbe counts. Even at the lowest level of aggregation, what is presented is an average of daily averages at a given sampling site. This would tend to obscure important peaks (for example, if microbe counts are highest when people are actually on the river), leading to an underestimate of the association between microbe count and illness.

Another important source of bias that may need to be recognized in the CHEERS study is selection bias, when the participants are selected in a way that makes the groups non-comparable, or when the participants are not truly representative of all the people in the population of interest. For example, by recruiting among organized groups such as rowing clubs, the study may obtain results that do not apply to the general population that might use the CAWS. Unlike confounding, it is difficult to know how much recall or selection bias there may be in a study, and virtually impossible to account for it in the analysis. It simply needs to be recognized in the study analysis as a potential limiting factor.

To the extent that potential biases exist, they call into question the strength of the CHEERS study's conclusions and generalizability of its results. For this reason, it is important in the research context to identify all such biases in evaluating data, so that its strength and significance can be better understood. Certainly, the statistical data in the Technical Report should not form the basis for any conclusions whatsoever until the potential epidemiologic biases are identified and discussed in a final study report.

C. Diminished Statistical Power

Critical to the predictive value of an epidemiological study is the size of the study sample. This is because epidemiology is, by its essence, a statistical endeavor. Much like a political poll, one surveys a large group of people to determine whether any patterns emerge that may be predictive for the larger population. And like a political poll, since one is reviewing only a sample and not the whole population, it is necessary to interpret the results with a "margin of error." That is, if one finds that out of 1,000 people surveyed that 50 of them will get sick, one cannot then make a straightforward extrapolation that in a population of 1,000,000, 50,000 people will get sick. The proper way to understand the result is that 50,000 people plus or minus X percent (the margin for error) will get sick.

The margin for error – X – is inversely correlated with the size of the sample. That is, the more people involved in the study, the more precise your results will be, and the smaller X will be. But if you do not have enough people in your study, your results will have a much larger margin of error. Thus, if you survey only 100 people and find that 5 of them got sick, this five percent positive finding is less reliable, and needs to be understood as a broad range of possible illness rates, ranging far above and far below 5 percent. X, the percentage margin for error, is necessarily very large. If you survey only 10 people, your results are essentially meaningless.

For this same reason, very little can reliably be concluded from negative results based on a small sample. There may be a small but significant percentage of the population that is becoming ill from the risk being screened for, but too small a sample may well miss all such people merely by chance. In other words, if approximately 50 out of every 1,000 people are getting sick, but you survey only 100 of those 1,000 people, there is a substantial possibility that you will not find among those random 100 even one of the 50 in 1,000 who is actually getting sick.

The question of sample size is largely determinative of the "statistical power" of a study. Statistical power refers to the probability that a study would conclude that there is a difference between groups if such a difference truly exists. Statistical power is most strongly related to sample size: the more subjects included in a study, the smaller the margin of error in the estimate for each group, and the easier it is to identify even a relatively small but important difference between groups.

The CHEERS study is a very large, ambitious epidemiologic study, and the investigators are to be congratulated for their successful enrollment and follow-up over several years. However, the ambitious size and scope of the study is not reflective of meaningful statistical power. In my 2009 testimony, I set forth my preliminary concern that the study as designed, despite its large overall sample size, lacks sufficient statistical power to identify clinically important differences in risk, in particular risk to the various sub-populations of CAWS users. That concern is borne out by the Technical Report data. That data reflects both an inherent inability of the study to assess health impacts to CAWS subgroups, as well as the prospect of reduced statistical power when the confounding factors are accounted for through statistical analysis. The Technical Report data also reflects two other limitations on the statistical power of the study: clustering and missing data.

1. Small Sample Size of Important Subgroups

As noted in my 2009 testimony, it is likely that subgroups of users have different risks, and may be of particular interest. For example, children may be at higher risk of illness due to differences in how much water they are exposed to, and underlying differences in immunity to infectious agents; those engaged in certain activities such as kayaking may be at particular risk as opposed to those who are fishing. While the margin of error for estimating illness among all the participants in the study may be adequately narrow to draw meaningful conclusions, that margin of error will be far greater for these important subgroups. To give a single example: if the risk of illness among all CAWS users is 4.3% (Table V-2), then the margin of error for this estimate is +/- 0.7%. However, for the group of CAWS users under 10 years of age, the margin of error is +/- 3.3% - five times higher. Because of these greater margins of error, this study will be unable to identify risks to important subgroups of people even if those subgroups are truly at higher risk.

2. *Effect of Adjustment for Confounding*.

I have already discussed the problem of confounding. It is possible to adjust for confounding when analyzing the data, but this comes at the cost of decreasing the power. Confounding has traditionally been accounted for by doing a type of analysis, called a stratified analysis, which divides the total study population into subgroups based on the factors of interest and the possible confounders. This is perhaps best understood using the drinking and smoking example I alluded to above. We know that drinkers are more likely to smoke, and we think that the association between drinking and cancer could be explained by the smoking, rather than the drinking. To know whether drinking itself truly causes cancer, we could look at people who smoke and those who do not smoke separately. We would then compare the rate of cancer among smokers who drink and smokers who do not drink. If, when we look at the smokers in isolation, we see a higher rate of cancer in those who drink, we would conclude that drinking has an independent effect on cancer. On the other hand, if drinking is *not* a cause of cancer, than the rates of cancer would be similar in both the drinking smokers and the non-drinking smokers; the same would be true if we look at the isolated subgroup of non-smokers and compare drinkers and non-drinkers.

In a stratified analysis, instead of comparing 2 groups (drinkers and non-drinkers), one compares 4 groups (drinking smokers, non-drinking smokers, drinking non-smokers, and non-drinking non-smokers). Now say there is also the possibility that people who drink are also less likely to get enough cancer-preventing vitamins; we might also have to account for diet as a confounder. Then we would have at least 8 groups to compare, based on drinking, smoking, and diet. (This assumes adequate diet is a yes/no question. If it is, say, low, medium, and high levels of vitamins, then there would be 16 groups.) With each additional confounder, the size of the subgroups gets smaller, and the margin of error for those subgroups gets wider – and the power to find a difference goes down.

Modern techniques to adjust for confounding use mathematical models called logistic regression models, rather than stratified analysis. With modeling, the same effect of adjusting for confounding is achieved at less loss of power, but the margin of error (also known as the confidence

interval) does get wider and there is a loss of power; the greater the number of confounders, the greater the loss of power.

3. Effect of Clustering.

The type of analysis noted above – logistic regression – is based on the idea that each participant in a study is completely independent of all the other participants. Whenever participants are somehow clustered, or linked in a way that makes them more similar than would be expected if participants were sampled completely at random, this must be accounted for. This is a technical statistical issue, but the end effect is that the margin of error (confidence interval) is wider than would be expected based solely on sample size, leading to further loss of power. In this study, participants were recruited at least in part in clusters. For example, a family of four sharing a boat would be counted as four individual participants. A more important example would be recruitment at an organized event, or from group activities (*e.g.*, a high school rowing team) where many individuals who may share important unmeasured characteristics (such as skill level, which may impact exposure and other factors) are considered independent when they are not. Assuming the investigators account for such clustering in the analysis, there would be some loss of study power.

4. Effect of Missing Data on Statistical Power

It is almost inevitable in an epidemiologic study that some participants will be missing some information. For example, someone may forget to check the box for male/female, or may fail to return a stool specimen. Although the investigators have done a thorough job in data collection, the magnitude of missing data is unclear from this report, and will need to be addressed in the next phase of the analysis. While the amount of data missing for any given variable may be very small, they can add up when many variables must be accounted for, since only those subjects with complete information for all of the variables in the analysis can be included in that analysis. For example, if there are 8 confounding variables to be considered in the model, and data are missing for only 1% for each variable, then the cumulative effect would be to drop 8% of the subjects from the analysis, with a corresponding loss of power.

Conclusion

I again congratulate the CHEERS study team on their successful completion of this ambitious study, which is a good overview and first step toward better understanding of risks associated with water exposure in the Chicago region. However, it is clear that the Technical Report can in no way serve as a sufficient basis for any conclusions whatsoever regarding health risks associated with the CAWS. Additionally, the Technical Report data bears out my initial concern with the lack of statistical power of the study to provide a meaningful evaluation of risk to CAWS recreators, and suggests additional study limitations as well.

I therefore urge the Board to draw no conclusions from the Technical Reports; and to be extraordinarily cautious and skeptical of any ultimate claims regarding the significance of the final CHEERS study once it is available.

_____

Marc H. Gorelick, M.D.