Frameworks for Estimating Risks Associated with Drinking Water

TCR meeting

October 17, 2007

Joseph Eisenberg, PhD

University of Michigan

Overarching Questions of Interest

- What do public health data tell us about drinking water risks?
  - Specifically those associated with distribution systems
- Can these data provide information on the extent to which the TCR is protecting public health?
- How do we proceed to improve the precisions of the answers to these questions?

Definition of Risk

- Risk: The relationship between exposure and the probability of a health outcome
  - Exposure can be defined in many ways
    - Drinking tap water
    - Ingesting pathogens
    - Living near region of distribution system with pressure drops

Definition of Risk (con't)

- Health outcomes can be defined in many ways
  - Clinical diagnosis of infection
  - Presence of antibodies to a particular pathogen
  - Self reported symptoms
- Measures of excess risk
  - Relative risk (risk difference, risk ratio, odds ratio)
  - Attributable risk (also population attributable risk)

Frameworks for Risk Estimation: Direct method

- Measure a health outcome along with one or more exposure factors
  - Data collection and analysis methods comes from the field of epidemiology
  - Data can be collected at individual, household or community level
  - Many study designs possible

Frameworks for Risk Estimation: Direct method (con't)

- Issues
  - - Methods are not particularly sensitive
  - - Studies can be expensive
  - + Provides direct estimate of an illness that may be caused by many pathogens
  - + Can find associations without the need to completely understand a complex process

Frameworks for Risk Estimation: Indirect method

- Measure exposure variables and use a dose-response relationship to predict risk
  - QMRA (quantitative microbial risk assessment)
  - Exposure variables are generally pathogen concentrations in the environment
  - Modeling techniques are used to obtain risk estimate

Frameworks for Risk Estimation: Indirect method (con't)

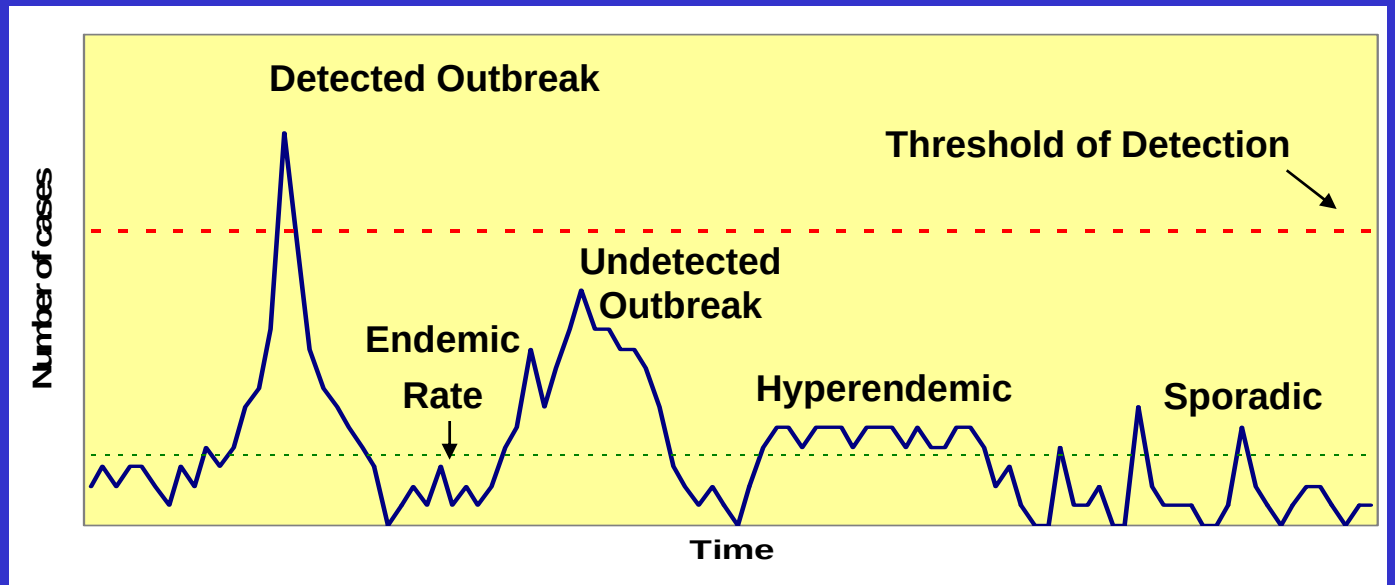
- Issues
  - - Assumes models are correct
  - - Pathogen specific
  - + Provides a means to predict risks of alternate scenarios
  - + Can be used in low risk situations

Direct Method: Epidemiology Study Designs

- Observational
  - Outbreak studies (epidemics)
  - Cross sectional (Food Net)
  - Ecological (Time series)
  - Case-control (Generally pathogen specific)
  - Cohort (Nygård, recreational swimming studies)
- Experimental
  - Intervention trials (household and community)

Outbreak Studies

- Advantages
  - Often possible to find the cause
  - Documents a problem
  - Can look at trends (need to consider reporting bias)
- Disadvantage
  - Difficult or impossible to quantify public health problem



Cross Sectional Studies

- Collect information at one point in time
  - Can associated health outcome with exposure
  - Generally data comes from surveys, but can collect clinical samples
- Advantages
  - Can get a well-represented sample of the population
  - Relatively inexpensive
- Disadvantages
  - Strength of evidence is thought to be low due to potential for confounding and difficulty to evaluate the causal temporal direction
    - Have a submitted manuscript countering this statement

Cross Sectional Studies

- Examples include Food Net
  - 0.65 episodes AGI per person per year
  - Subjects were asked about the past month (recall bias)
  - Survey can be modified to emphasize water exposures (e.g., Sonoma RDD, data collected but not analyzed)

Cross Sectional Studies

- **Hunter et al. 2005**
  - **Solid hypothesis generating result**
    - **Strong association (OR = 12.5 [3.5, 44.7])**
      - 12.5 times more likely to have diarrhea if you reported low water pressure at faucet (how was question asked?)
    - **Study was not designed for this analysis so result must be interpreted appropriately**
- **Mead estimate based on surveillance data rather than an observational study.**
  - **Weaker strength of evidence**
  - **Need good surveillance data (hard to come by)**

Ecological Studies

- Collect data at the community level
  - Generally uses secondary data sources
  - Outcome at community level (e.g., hospitalizations or emergency room visits), exposure at community level (e.g., CWS has filtration), or both
- Advantages
  - Unit of analysis may be the appropriate one
  - Relatively inexpensive
- Disadvantages
  - Potential confounding (ecologic fallacy)
  - Multiple testing

Ecological Studies

- Examples
  - Naumova et al.
    - Are temporal patterns of turbidity correlated with ER visits for GI illness?
    - Issues of confounding and multiple testing
  - Frost et al
    - Are communities that drink surface water more likely to be seropositive to *Cryptosporidium* than communities that drink ground water
    - Single pathogen result; issues of confounding

Ecological Studies

- Tinker et al
  - Are ER visits correlated with distribution system residence time by zip code?
  - Risk ratio = 1.1. Low risks. Small confounding can have important effect.

❖ All are hypothesis generating results

Case Control Studies

- Collect exposure information retrospectively
  - Enroll all cases with positive outcome and randomly selected controls with negative outcome.
- Advantages
  - An efficient study design when diseases are rare (generally used when specific pathogen is outcome)

Case Control Studies

- Disadvantages
  - Potential confounding
  - Enrollment may result in bias selection (searching for rare cases)
    - Generally no information on the population at risk

Cohort Studies

- Enrolls individuals by exposure status and cohort is followed in time
  - Compare incidence rates
- Advantages
  - Provides opportunity to get the temporal causality correct
- Disadvantages
  - Potential confounding
  - More expensive than other observational studies

Cohort Studies

- Examples
  - Nygård et al
    - When there is a main break or maintenance activity, exposed and unexposed are interviewed for one week
    - 12.7% of exposed vs. 8.0% of unexposed reported GI illness in the house (RR=1.58 [1.1, 2.3])
    - Efficient use of a cohort

Cohort Studies

- Recreational studies
  - A history of relating water quality (as measured by indicator concentrations) to GI illness
- Cohort studies can be more efficient than intervention trials
  - Cost
  - Blinding may be possible if participants are not aware of exposure

Cohort Studies: Evaluating Existing Regulations

- Recreational water studies: A meta-analysis (Wade, Pai, Eisenberg, Colford, EHP 2003)
  - To evaluate the link between specific microbial indicators of water quality and specific health outcomes.

Cohort Studies: Evaluating Existing Regulations

- **Reviewed 28 studies**
  - Prospective and intervention trials that assessed the risk of GI illness to swimming in either marine or fresh water.
- **Basic study design**
  - Enrolled participants at parking lot
  - After beach visit, participant filled out questionnaire
    - Exposure established
    - Water samples collected
  - Follow-up phone interview to report symptoms

Cohort Studies: Evaluating Existing Regulations

- Enterococci in marine water (current EPA standard = 35 cfu/100ml)
  - Review supports current EPA standards

	Below Standard	Above Standard
Relative risk	1.36 (0.91 - 2.03)	2.27 (1.74 -2.96)
Mean density (cfu/100mL)	20	139
# of studies	9	12

Intervention Trials (Experimental Studies)

- Exposure is assigned by research team
  - Main difference compared to observational studies
- "Gold standard" in Epidemiology
  - Is this true for environmental epidemiology? (IJE submitted)
- Advantages
  - Controls for unknown confounders
  - Blinding and randomization increase weight of evidence

Intervention Trials (Experimental Studies)

- Disadvantages
  - Costly
  - Sensitivity (logistics may limit sample size)
  - Misclassification
  - Generalization is always an issue (one intervention trial is a sample size of 1)

Intervention Trials (Experimental Studies)

- Examples (risks of those exposed to tap water compared with those not exposed)
  - Laval 1: AR = 100 * ((0.76 - 0.5) / 0.76) = 34%
  - Laval 2: AR = 100 * ((0.66 - 0.58) / 0.66) = 12%
  - Walnut Creek: IRR = 1.32 [0.75, 2.33]
  - Australia: IRR = 0.99 [0.85, 1.15] => AR < 15%
  - Davenport: IRR = 0.98 [0.87, 1.1] => AR < 11%
  - Sonoma: Manuscript in preparation (< 55 yrs)

Intervention Trials (Experimental Studies)

- **Differences can be explained by**
  - Secular trends (temporal variability)
  - Differences in water quality (source water, treatment efficacy, and distribution system)

Intervention Trials (Experimental Studies)

- Examples (continued)
  - Payment 2 also has information on effects of distribution
    - AR = 100 * ((0.60 - 0.58) / 0.60) = 3%
    - This is the risk attributable to the distribution system

Intervention Trials (Experimental Studies)

- Tap Water Intervention Trial (Colford et al. AJE, 2005)
  - To evaluate the risks of GI illness associated with drinking tap water.

Intervention Trials (Experimental Studies)

- **Basic study design**
  - **Enroll 456 households**
    - 227 receive an active home water treatment device (and carry drinking water to work, etc. when practical)
    - 229 receive a “placebo” home water drinking device (does nothing to the water)
  - **Follow the subjects for one year with daily logs of GI illness**
  - **Alternative design: Each household changes device type after 6 months.**

Intervention Trials (Experimental Studies)

- Risks attributable to drinking tap water (per 10,000 population)
  - –365 (95 percent confidence interval: –2555, 1825)
    - A cohort of 1296 persons reported 1223 episodes of HCGI in the active group and 1348 in the sham group
- Trial conclusion
  - No evidence of a significantly elevated drinking water risk
- Regulatory significance
  - As many as 1825 cases per 10,000 people per year attributable to drinking water

The Indirect Approach: Risk Assessment

- One way to get around the lack of sensitivity is to use QMRA
  - Originally adopted from the chemical risk assessment paradigm (Four step process)
    - Hazard identification
    - Dose-response assessment
    - Exposure assessment
    - Risk characterization

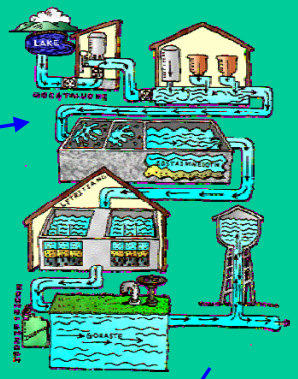
The Indirect Approach: Risk Assessment

- Issues unique to assessing risks associated with pathogens are
  - Secondary spread of infection and immunity
  - Risks effects are manifested at a population level
- ILSI/EPA framework addresses the unique aspects of the infectious disease process in a risk framework
  - Starts with a problem formulation phase
  - Next need to characterize exposure, how exposure affects health, and then estimate risk

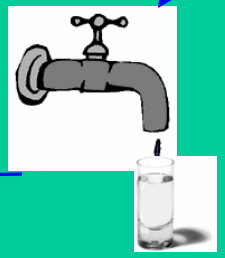
A Site-specific Risk Assessment: Davenport Iowa (Eisenberg et al. EHP, 2007)

- If exposure pathways can be characterized, the risk can be estimated

Source Water
Log normal distribution of source water concentrations estimated from local data.



Treatment Efficiency
Filtration, sedimentation was modeled using a Weilbull based on surrogate markers (*B. subtilis* for *Cryptosporidium* and *Giardia*, and somatic cholophage for enteric viruses). Chlorination was assumed to have no effect on *Cryptosporidium* and local data was used to model the chlorination effects on *Giardia* and the viruses.



Pathogens in Tap Water
Number of pathogens in water is Poisson with mean determined by source water and treatment efficiency.

Tap Water Consumption
Lognormal distribution estimated from local survey data.

Dose-Response
Used dosing trials of healthy volunteers.

A Site-specific Risk Assessment: Davenport Iowa

- Risk estimates based on site-specific water quality, water treatment, and water consumption patterns
  - 13.9 cases/10,000 (2.5 and 97.5 percentile: 1.6, 37.7)

Cases of Illness (per 10,000)		
	Mean	Percentile (2.5, 97.5)
Cryptosporidium	2.1	(0.8, 3.5)
Giardia	3.4	(0.6, 15.5)
Enteric viruses (disinfection = 4 log removal)	8.4	(0.2, 18.7)
Enteric viruses (disinfection = 6 log removal)	0	(0, 0.2)

Intervention Trial vs. A Site-Specific Risk Assessment

	Risk Assessment	Intervention Trials
Sensitivity	Not relevant	Low
Causal evidence	Indirect	Direct
Pathogen inclusion	Few	Many
Model Specification	Adds uncertainty	Not relevant
Transmission processes	Can be included[a]	In a limited way
Distribution System effects	Can be included[a]	Was included
Examining alternative control strategies	Yes	No
Expense	Low	High
Time	Fast	Slow

[a]Was not included in this study

Generalization

- QMRA can estimate risks under different scenarios
  - Modeling may be needed to characterize exposure pathways
  - To date we are not aware of a QMRA of exposures due to the distribution system
    - Characterizing distribution system exposures is more challenging than source water exposures
- Another approach is to generalize epidemiology studies using models
  - Messner et al. (2006)
  - Colford et al. (2006)

Estimating AR at the National Level

- Basic framework of both the Messner and Colford approaches are similar
  - Characteristics of a community's water quality can explain the AR estimates obtained from intervention trials
    - And these characteristics can be quantified; i.e., we can collect the appropriate data
  - Assumptions are made to enable the generalization of intervention trial results to other communities
- Model based approaches
  - (-) Model assumptions are assumed correct
  - (+) Assumptions are explicit and can be challenged
    - Sensitivity studies can quantify importance of assumptions

Estimating AR at the National Level (Messner Model)

1. Each CWS has a certain, but unknown, mean AGI incidence rate due to DW
   - Incidence comes from both source water/treatment and distribution system causes
   - Laval studies provide a means to disaggregate source/treatment and distribution system effects
   - Improvement in treatment between Laval 1 and 2 assumed to be due to source/treatment improvements
     - 0.26 (Laval 1) -> 0.08 (Laval 2)
     - Assume half of 0.08 due to distribution system (0.22 source/treatment, 0.04 distribution system)

❖ Ideally would like more intervention trials
   - Need to plan out studies carefully to maximize utility of results

Estimating AR at the National Level (Messner Model)

2. The type and spread of the distribution of mean AGI incidence rates among populations served by CWS can be estimated
   - Assume a log normal distribution
     - Asymmetric, non zero, commonly used
   - Assumes a 2 – 5 log spread
     - Based on information from surface water data (ICR)

❖ Ideally would want to have a comprehensive database of risk factor data by CWS

Estimating AR at the National Level (Messner Model)

3. A percentile AGI incidence rate for the distribution above can be estimated
   - Linking AGI estimates from epidemiology studies to microbial risks
     - Ranking of Laval 1 source treatment: 0.90 – 0.995
     - Ranking of Laval 1 distribution system: 0.50 – 0.99
4. The statistical distribution derived from 2 and 3 can be integrated to inform a mean national estimate of AGI due to DW
   - Monte Carlo simulations provide national estimate

Estimating AR at the National Level

- The key to this approach is estimating the appropriate risk factors associated with the communities where the intervention trial was conducted
  - Source water quality
  - Treatment characteristics
  - Distribution system characteristics
- This is an important research agenda item

Estimating AR at the National Level

- Summary of results
  - Messner
    - AGI attributable to drinking water: 0.06 cases per year [0.02, 0.16]
    - Or an estimated in 16.4 million cases of AGI per yr
  - Colford
    - 4.26 – 11.69 million cases of AGI per yr

Estimating AR at the National Level

- The value is less important than the process
- Need a set of independent estimates using the same framework
- Need more data
  - National estimate provides a structure from which to prioritize research needs

Conclusions and Recommendations

- **Short term**
  - **Independent groups re-evaluate national estimate using a common framework (Messner et al.?)**
    - **Would require a wide variety of experts**
    - **Can address alternate ways to use existing data**
      - E.g., Can we use the Davenport result?

Conclusions and Recommendations

- Medium term
  - Improve risk characterization of CWS by collecting data and creating an extensive database
  - Explore a QMRA approach
- Long term
  - Carefully plan out more epidemiology studies (what community characteristics and what study design)
    - Current experience with national estimate should enable a more strategic selection process.