Reviewer's report

Title: Water and sewage systems, socio-demographics, and duration of residence associated with endemic intestinal infectious diseases: A cohort study

Version: 1 Date: 21 March 2010

Reviewer: Karin Nygard

Reviewer's report:

The paper examine the association between water and sewage systems, sociodemographic factors, and duration of residence associated with endemic intestinal infectious diseases. The main study is a register-based retrospective cohort study conducted in a localized area in Canada, where detailed individual-level data on exposures and illness is available in different registers. Additionally, they have conducted a separate cross-sectional survey in the same area collecting additional information on private drinking water and sewage systems, and water use.

The paper is interesting and adds information to the knowledge base on the risk of intestinal infectious diseases and exposures related to water and sewage systems. The data is obtained from available registers on physician visits and hospitalization records, water and sewage systems, climatic data registers and population registers. It is a nice example on how to use available register-based data in order to gain more knowledge.

Major compulsory revisions

The manuscript is very long and need to be better structured. Several times, new analyses and results are presented in the discussion session which are not mentioned in the methods or results sections previously.

Statistical analysis

My main concern is the statistical analyses. A statistical review is needed, but I will briefly comment some of the issues I find unclear below.

The study population is an open cohort (people moving in and out of the area) with different follow up time. After an episode of illness, the person is excluded. They state that the unit of analysis is person-days, but it is a bit unclear to me how they that was done in the logistic-regression model. They have assessed the influence of repeated measures on each subject, so I assume that they use each day as ill vs not ill, and then end up with nearly 200 million subjects instead of a cohort of around 100,000 individuals. To me this seem like falsely increasing the study power, but this should be assessed by a statistician. Maybe a survival time regression model would be more appropriate (mentioned also below)

Since this is a cohort study, they should explain why they chose to use a logistic
regression model giving odds ratios as effect measures when they could have used a model with relative risk ratio estimates (in open cohort studies with unequal follow up times, Cox’s proportional hazards model is often used, however I am not sure if that would influence the result).

In result section they present relative increases and decreases in risks based on dividing the odds ratios. Odds ratios will always be further away from 1 than RR, therefore calculating the percentage increase/decrease by dividing two odds ratios will bias the effect away from zero. This bias will increase as the outcome is more frequent.

One of the main result is that chlorine disinfection is associated with lower disease, although the results only marginally show this. I.e they state that the crude incidence rate is lower for those with chlorinated water, however the crude difference is very small (1343 vs 1358 per 100,000 personyears). The same for the findings from the logistic regression (OR of 0.92 for chlorinated water, upper limit of confidence interval=1).

There is no assessment of correlation or interaction between the exposure variables, and how this could influence the results. It is reason to believe that several of the exposure variables may be correlated, as is briefly mentioned in the discussion, and the authors need to address this issue in the methods and analysis part before making conclusions about the results.

Other comments on methods:
How they account for address changes within the township during the study period should be mentioned
The description of the water systems is not very clear:
How they categorize proportion of surface water should be described, and later these categories should not be called “Water systems” but rather categories. Also, it would be clearer if they used fewer categories (why five and not limit to three?)
Was the chlorination only used for surface water, or what was the proportion of surface water chlorination vs well water chlorination?
Is chlorination the only disinfection procedure in use? No UV or membrane filtration?
Since chlorination was implemented during the study period, how was that accounted for in the analyses?
What is the difference between minicipal and community water systems – is there a difference in terms of management and control?
Is it only for those with municipal sewage systems that there is a risk of cross-contamination between water and sewage lines – or do they also run in parallell for the private systems?

Additionally, I think it would make it clearer if it is possible to present a map of the study area, showing the distribution of the water systems included in the study
Household income is aggregated data, not individual level data, and thus may be biased (ecological study bias)
They should include a mention of ethical approval for the study.

Minor Essential Revisions

Tables
Study period and area, and number of study units (N) should be included in the title
The proportion surface water in the water categories should be stated (now it says increasing, not specified what is increasing – although it can be understood by the text)

Level of interest: An article whose findings are important to those with closely related research interests

Quality of written English: Acceptable

Statistical review: Yes, but I do not feel adequately qualified to assess the statistics.

Declaration of competing interests:
I declare that I have no competing interests