Length of Follow-up for the NIH R03 Primary Analysis

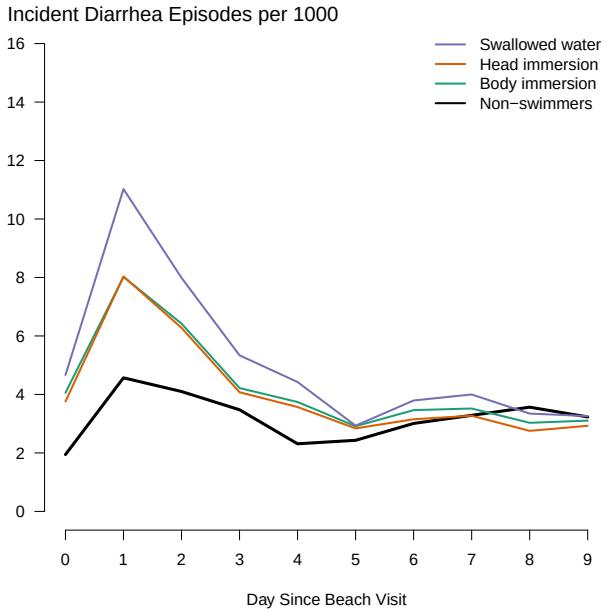
Ben Arnold (benarnold@berkeley.edu)

Updated April 9, 2015

Choice of follow-up period

In our original NIH application, we stated that we would use a 10 day follow-up period. We chose 10 days because every single person enrolled in all 13 cohorts had at least 10 days of follow-up (i.e., nobody was censored). There was additional rationale for a longer follow-up period as well: 10-days captures the incubation period for most viral, bacterial and protozoan pathogens, and the longer the follow-up period the more cases of illness there is in the cohort = more statistical power.

Daily incidence curves for diarrhea suggest that the maximum separation between swimmers and non-swimmers is in the first 3 days (these are data from 88,083 people across all 13 cohorts):



The large separation between swimmers and non-swimmers, led us to think that focusing on the first 3 days of follow-up would be the most relevant period for the analysis. For this reason, we proposed a change to the protocol to limit the primary analysis to the first 3 days (rather than 10 days), and then conduct a sensitivity analysis of the effect of follow-up period on the parameter estimates. The results of that sensitivity analysis are below and I think we should revisit this decision.

Follow-up period sensitivity analysis

We conducted a simple sensitivity analysis that varied the length of follow-up from 1 to 10 days. We re-estimated the adjusted Cumulative Incidence Ratio (CIR) associated with:

- swim exposure (comparing swimmers to non-swimmers)
- *Enterococcus* 1600 exposure among swimmers (Quartiles of concentration)
- *Enterococcus* 1600 exposure among swimmers (above and below regulatory limits)

The following pages include figures referenced in these summary results of the analysis:

Swim exposure analysis: Increasing the length of follow-up led to an attenuation in the CIR associated with swim exposure (Fig 1). The CIR with 3 days of follow-up was 1.75 (95% CI: 1.53, 2.00); with 10 days of follow-up the CIR was 1.45 (1.33, 1.59). Longer follow-up led to more precise estimates with nearly twice as many incident cases.

Entero 1600 Quartiles: Increasing the length of follow-up had no appreciable impact on the CIR associated with exposure to *Enterococcus EPA 1600* (Fig 2). However, estimates with longer follow-up periods were considerably more precise (owing to the larger number of cases).

Entero 1600 >35 CFU/100ml: Increasing the length of follow-up had no appreciable impact on the CIR associated with exposure to *Enterococcus EPA 1600* above the regulatory limit (Fig 3). Results were similar when stratified by beach conditions (point vs. non-point source) where there was significant effect modification. As with the quartile analysis, estimates with longer follow-up periods were considerably more precise (owing to the larger number of cases).

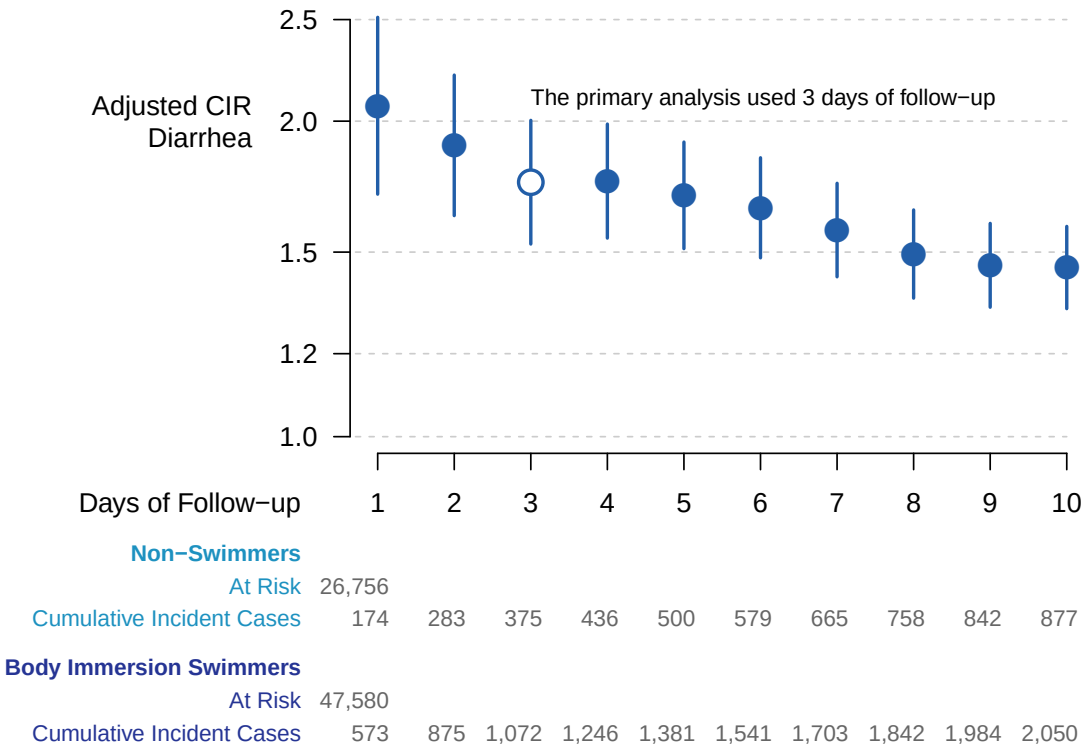


Figure 1: Sensitivity analysis of length-of-follow-up and the Cumulative Incidence Ratio (CIR) of diarrhea associated with swim exposure (comparing body immersion swimmers to non-swimmers

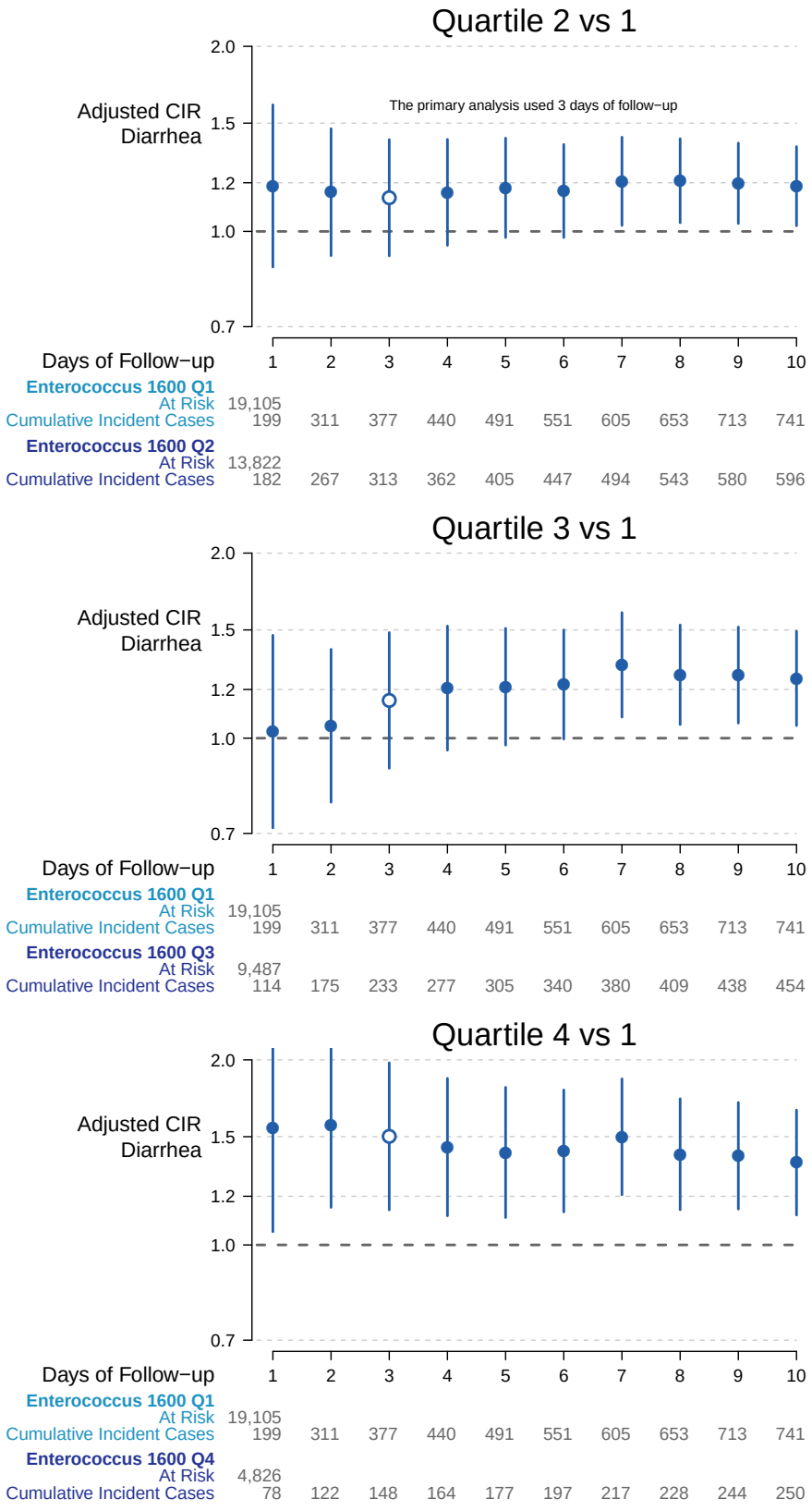


Figure 2: Sensitivity analysis of length-of-follow-up and the Cumulative Incidence Ratio (CIR) of diarrhea among body immersion swimmers associated with exposure to higher quartiles of *Enterococcus* EPA 1600 concentrations.

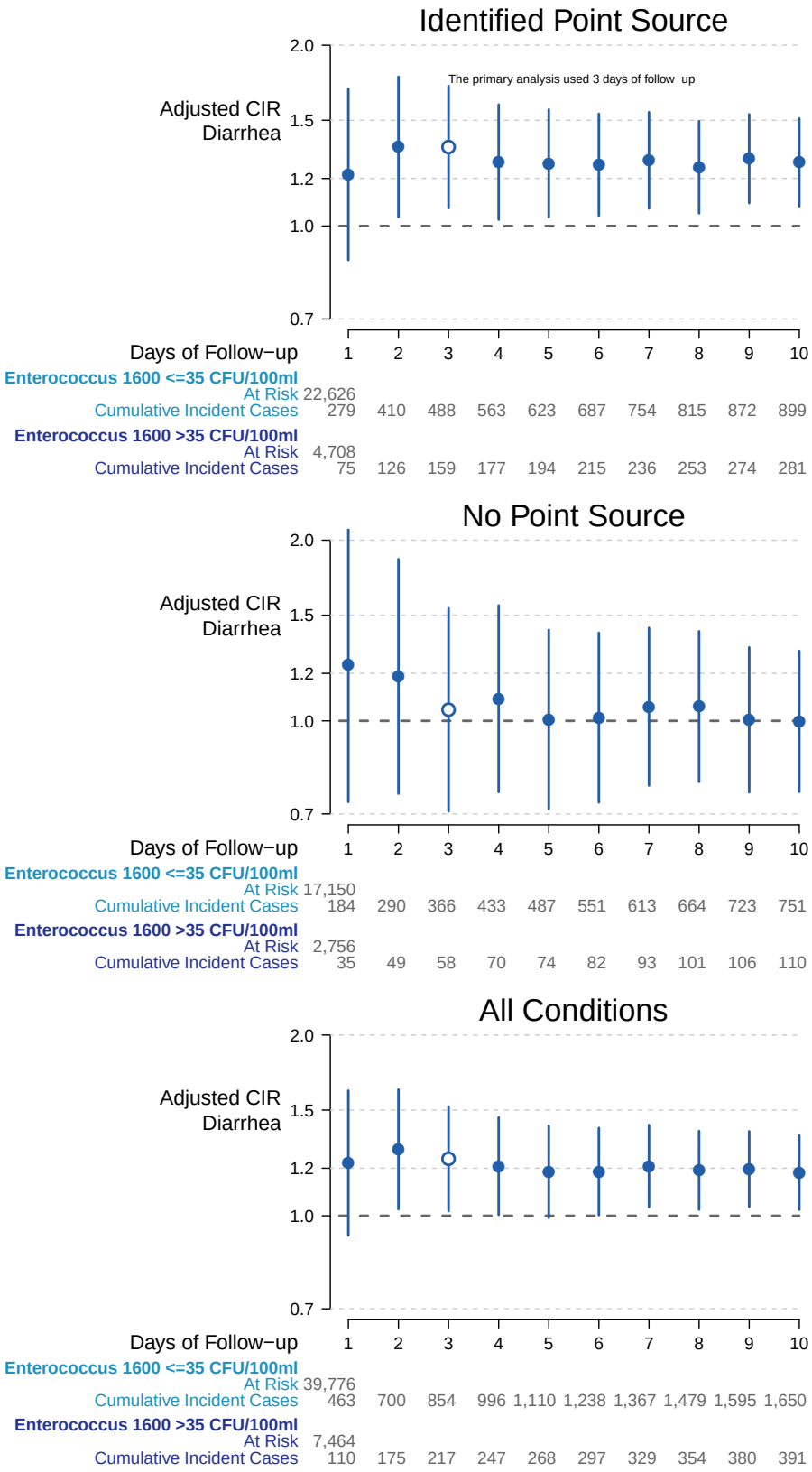


Figure 3: Sensitivity analysis of length-of-follow-up and the Cumulative Incidence Ratio (CIR) of diarrhea among body immersion swimmers associated with exposure *Enterococcus* EPA 1600 concentrations >35 CFU/100ml, stratified by beach conditions.

Reflections and Points For Discussion

Based on the results of this sensitivity analysis, I think it would be useful to reconsider a 10-day follow-up period. These are the main considerations from my perspective that would support a 10 day period compared to a 3 day period:

- 10 days was the length of follow-up we specified in our original protocol, for reasons outlined on page 1 (more cases, capture relevant pathogen incubation periods)
- 10 days is closer to all of the previously published analyses (which used 10-12 days)
- 10 days clearly wins out in a bias-variance tradeoff for the water quality indicators, because there is little "bias" (difference between estimates) and clearly lower variance [note: we cannot do a formal bias-variance tradeoff without making some assumption of what the true CIR is to estimate bias]
- 10 days leads to an attenuated CIR for the swim exposure analyses compared to 3 days. But, it's important to note that observation is on the relative scale. The Risk Difference actually increases with follow-up (despite a smaller relative risk) because of the consistent accumulation of excess cases over follow-up (see figure on page 1). The RD at 3 days is 8.5 cases per 1000, and at 10 days is 10.3 cases per 1000.
- For the outcomes of GI-related hospital visits, and missed days of work/school, we will need to look at a 10 day follow-up period because of how rare the outcomes are. Using a 10-day follow-up period throughout would be most consistent.