

Supplementary Information for

Simulated gastrointestinal risk from recreational exposure to Southern California stormwater and relationship to human-associated Bacteroidales marker HF183

Sarah A. Lowry¹, Joshua A. Steele², John F. Griffith², Kenneth C. Schiff², and Alexandria B. Boehm^{1*}

¹ Department of Civil and Environmental Engineering, Stanford University, 473 Via Ortega, Stanford, CA 94305, USA

² Southern California Coastal Water Research Project, Costa Mesa, CA, USA

Details on hypergeometric function computation

Hypergeometric functions were computed in MATLAB (The MathWorks Inc., 2023), and a lookup table was generated and imported into R to use in risk characterization. For adenovirus, the hypergeometric function predicted that risk was equal to 1 at a dose of 4565.4 gene copies. For norovirus and *Campylobacter*, at high doses that were outside of the range of the lookup table (upper limit of 5,000), the hypergeometric function was linearized and that linear equation was used to compute the probability of infection. R^2 values of this linearization were 0.998 for both norovirus and *Campylobacter*. The dose at which the probability of illness was equal to 1 was found so and added as a limit so that the probability of illness could not exceed 1 for any dose of pathogen.

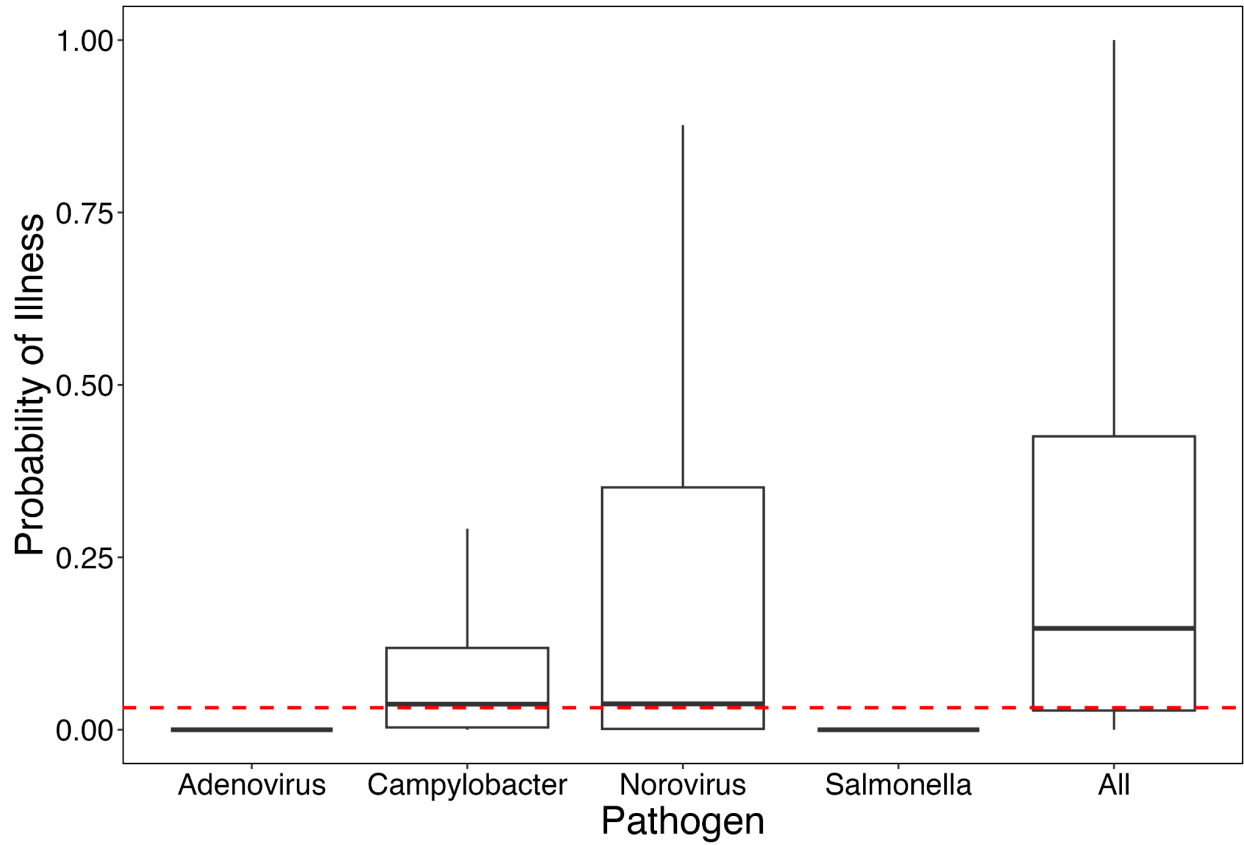


Figure S1. Estimated risk from norovirus and *Campylobacter*, setting concentrations of adenovirus and *Salmonella* equal to 0. The median P_{iii} is 0.15, or 15%, as compared to the median P_{iii} including adenovirus and *Salmonella* of 0.19, or 19%.

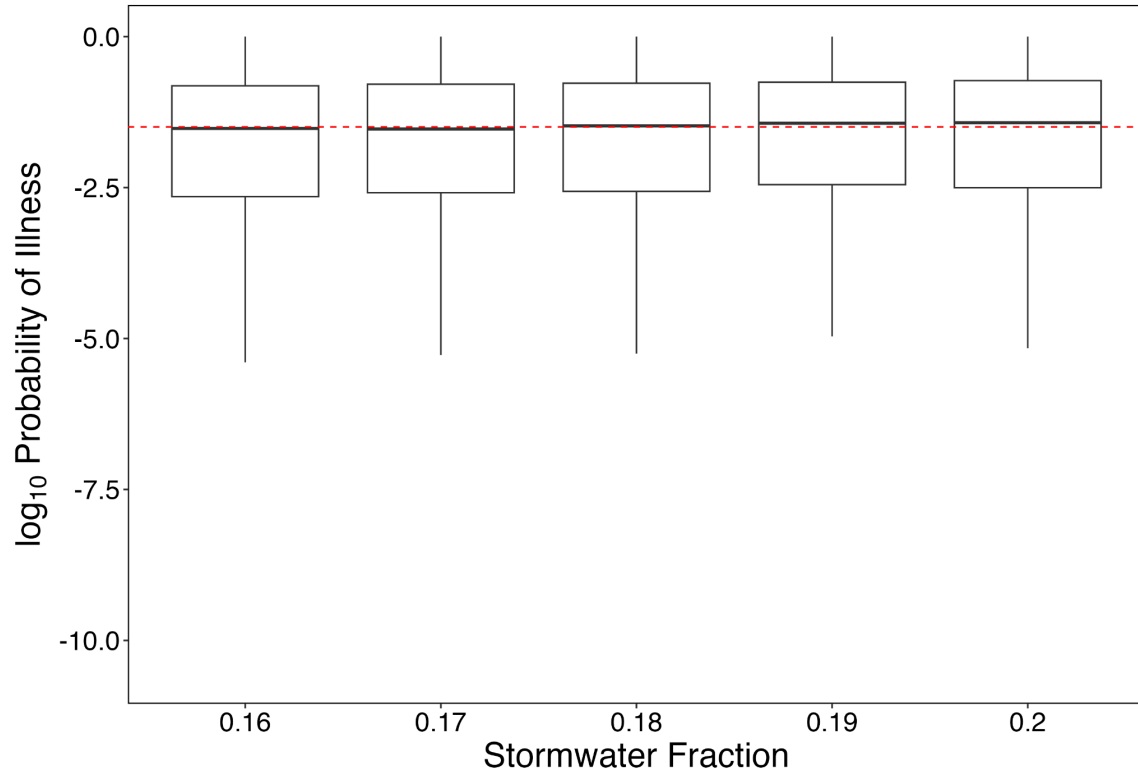


Figure S2. Box plots of \log_{10} -transformed probability of illness versus \log_{10} -transformed stormwater fraction zoomed into stormwater fractions between 0.16 and 0.2. The horizontal line in the box represents the median, the bottom and top of the box represent the 25th and 75th percentile, and the top and bottom whiskers represent $Q3 + 1.5 \cdot IQR$ and $Q1 - 1.5 \cdot IQR$, respectively. Each box plot represents 10,000 model runs from the Monte Carlo simulation. The red dashed line represents the \log_{10} -transformed risk threshold of 32/1000 illnesses.