Acute gastrointestinal illness and public health impacts following recreational exposure to water and fecal indicator bacteria at beaches in the United States

Supplementary Appendices

Compiled June 12, 2015

Supplementary Appendix Section

1 Beach and Water Quality Details

Figure S1 maps the 13 study beaches and summarizes some of their key characteristics. All four of the freshwater beaches in the study were in the great lakes region – Huntington, Silver, West, and Washington Park beaches.

Tables xxx and xxx summarize the number of water samples collected at each beach and the summary statistics for the *Enterococcus* EPA 1600 and EPA 1611 qPCR assays. In general, there was not a strong relationship between the two fecal indicator bacteria in the studies (Figure S2).

XXX Include: Water quality averaging methods

XXX Placeholder for tables that include water quality summaries from each beach

Figure S1: Placeholder for Beach Map. Will include locations, dates, enrollment, and type (fresh/marine, point/non-point).

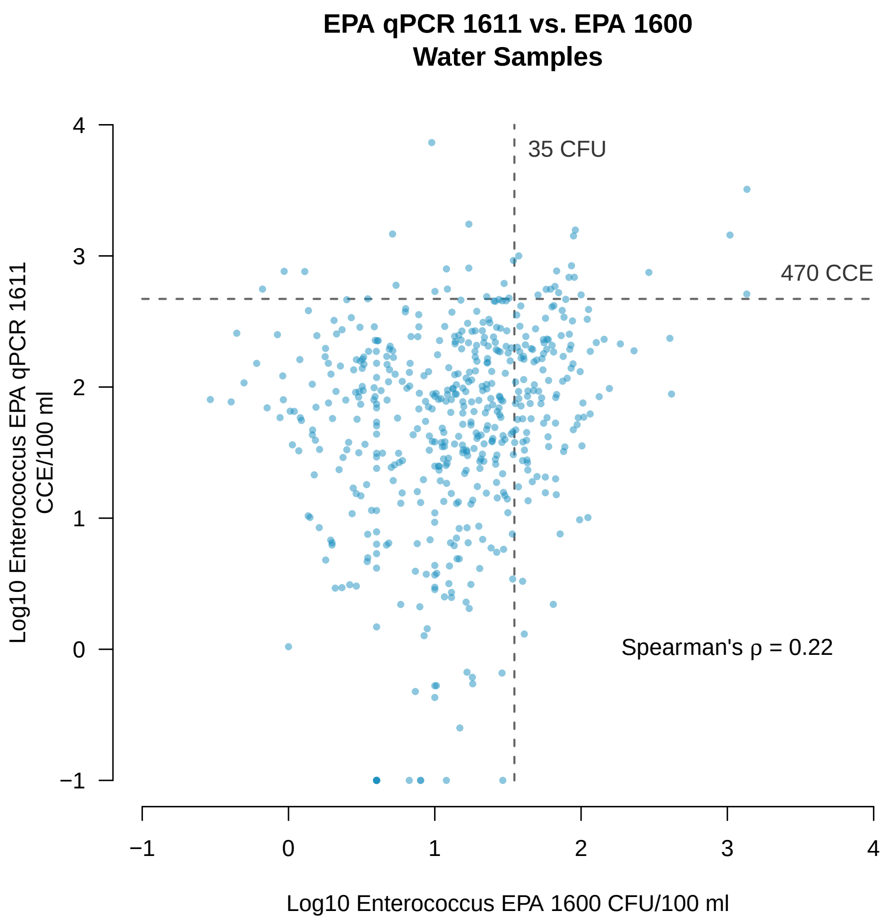

Figure S2: Scatter plot of *Enterococcus* EPA qPCR 1611 values versus *Enterococcus* EPA 1600 values in water samples. The figure includes USEPA regulatory guidelines for each indicator. Spearman's rank correlation between the indicator values is 0.22.

2 Study Participant Details

The study included 88,083 individuals across the 13 cohorts. The studies tended to enroll families with children – for this reason the age distribution of the study population is approximately bi-modal (Figure S3). Teenagers were under-represented in the cohorts due to the inclusion criteria that required an adult present from the household.

The most common racial categories (self-reported) were white/caucasian (55.4%) and non-white, hispanic (29.5%). Table S1 summarizes the study cohort demographics and water exposure, stratified by age.

Figure S4 summarizes the distribution of *Enterococcus* sample concentrations and swimmer exposure concentrations in the study cohort for the EPA 1600 assay and the EPA 1611 qPCR assay. We excluded from the distribution plots 73 water samples from freshwater lagoons at Dohney and Malibu beaches because although our team collected the water samples, very few individuals were actually exposed to that water – only 54 (Doheny) and 68 (Malibu) participants were exposed to lagoon water.

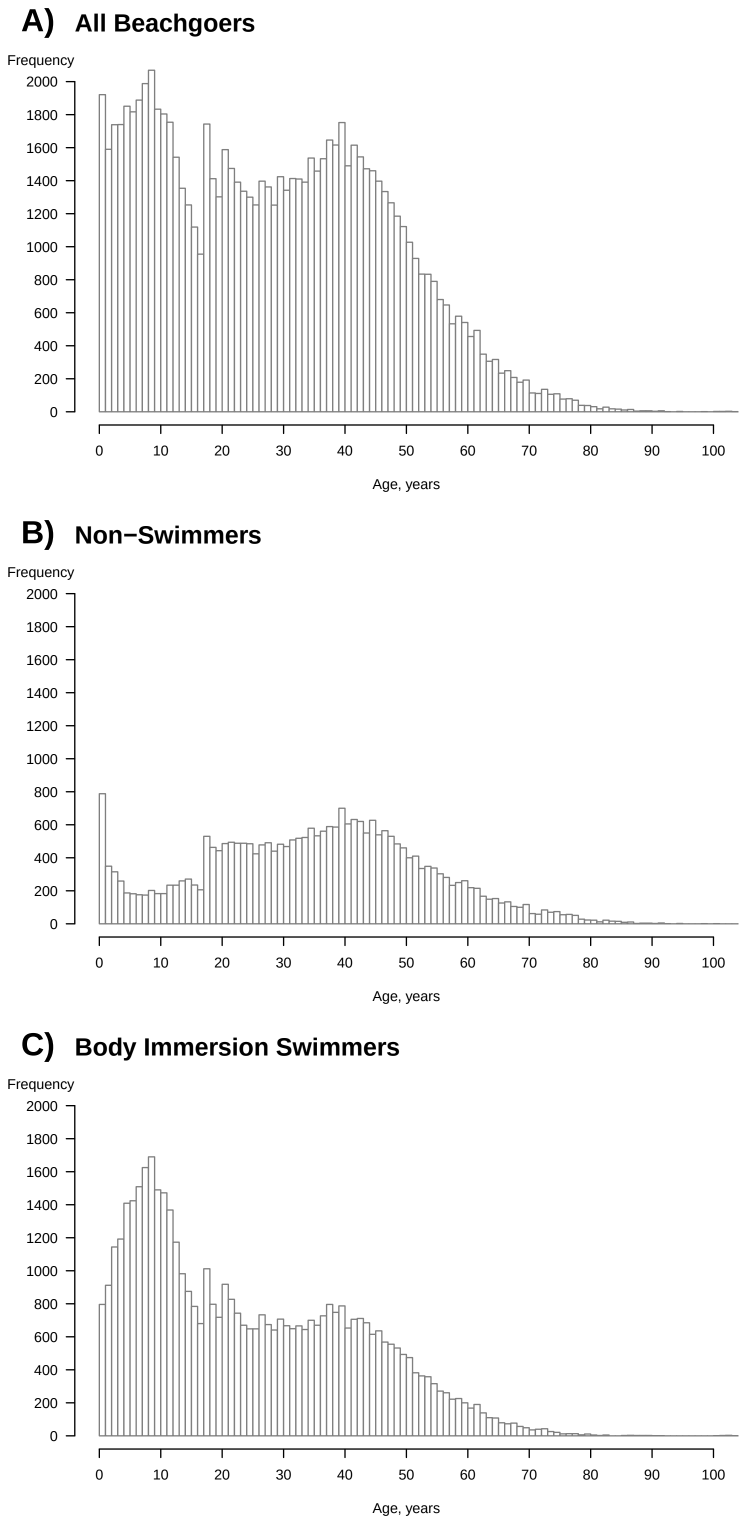

Figure S3: Age distribution of the study population; bin width is 1 year. **A)** All participants; **B)** Non-swimmers (no water contact); **C)** Body immersion swimmers (entered the water to waist depth or more).

Table S1: Participant Characteristics by Age Category

	All Ages			Age 0 to 4 Years			Age 5 to 10 Years			Age >10 Years		
	N	%	Median (IQR)	N	%	Median (IQR)	N	%	Median (IQR)	N	%	Median (IQR)
Number of Participants	88,083			6,990			11,446			68,428		
GI illness at enrollment	2,025	2.3		196	2.8		195	1.7		1,620	2.4	
Individuals at risk of GI illness	86,058			6,794			11,251			66,808		
Incident diarrhea within 3 days	2,357	1.9		248	2.8		350	1.9		1,732	1.8	
Incident diarrhea within 10 days	3,409	4.0		398	5.9		393	3.5		2,585	3.9	
Age in years			28 (12,43)			2 (1,3)			8 (6,9)			35 (22,46)
Female	47,204	53.6		3,413	48.8		5,655	49.4		37,566	54.9	
Race												
White	48,829	55.4		3,429	49.1		5,843	51.0		39,026	57.0	
Non-White, Hispanic	26,014	29.5		2,109	30.2		3,340	29.2		20,146	29.4	
White, Hispanic	1,262	1.4		170	2.4		238	2.1		846	1.2	
African American	2,600	3.0		204	2.9		386	3.4		1,960	2.9	
Asian	2,018	2.3		173	2.5		257	2.2		1,564	2.3	
American Indian	240	0.3		13	0.2		23	0.2		202	0.3	
Multiple Races	1,753	2.0		315	4.5		492	4.3		924	1.4	
Other	1,008	1.1		115	1.6		149	1.3		717	1.0	
Missing	4,359	4.9		462	6.6		718	6.3		3,043	4.4	
No water contact	27,460	31.2		1,711	24.5		1,104	9.6		24,325	35.5	
Any water contact	60,623	68.8		5,279	75.5		10,342	90.4		44,103	64.5	
Body immersion	48,573	55.1		4,044	57.9		9,147	79.9		34,656	50.6	
Head immersion	37,999	43.1		2,901	41.5		7,988	69.8		26,515	38.7	
Swallowed water	11,208	12.7		1,679	24.0		3,166	27.7		6,209	9.1	
Hours spent in the water			1.0 (0.2,3.0)			1.0 (0.3,2.5)			2.0 (1.0,4.0)			1.0 (0.2,2.5)
Hours spent in the water (cat)												
0 – 1	17,715	29.2		1,708	32.4		2,168	21.0		13,691	31.0	
1.1 – 2	9,677	16.0		744	14.1		1,678	16.2		7,142	16.2	
2.1 – 3	3,966	6.5		281	5.3		955	9.2		2,705	6.1	
3.1 – 4	4,553	7.5		342	6.5		904	8.7		3,250	7.4	
4.1 – 5	1,249	2.1		101	1.9		337	3.3		801	1.8	
>5	5,128	8.5		335	6.3		1,078	10.4		3,616	8.2	
Missing	18,335	30.2		1,768	33.5		3,222	31.2		12,898	29.2	

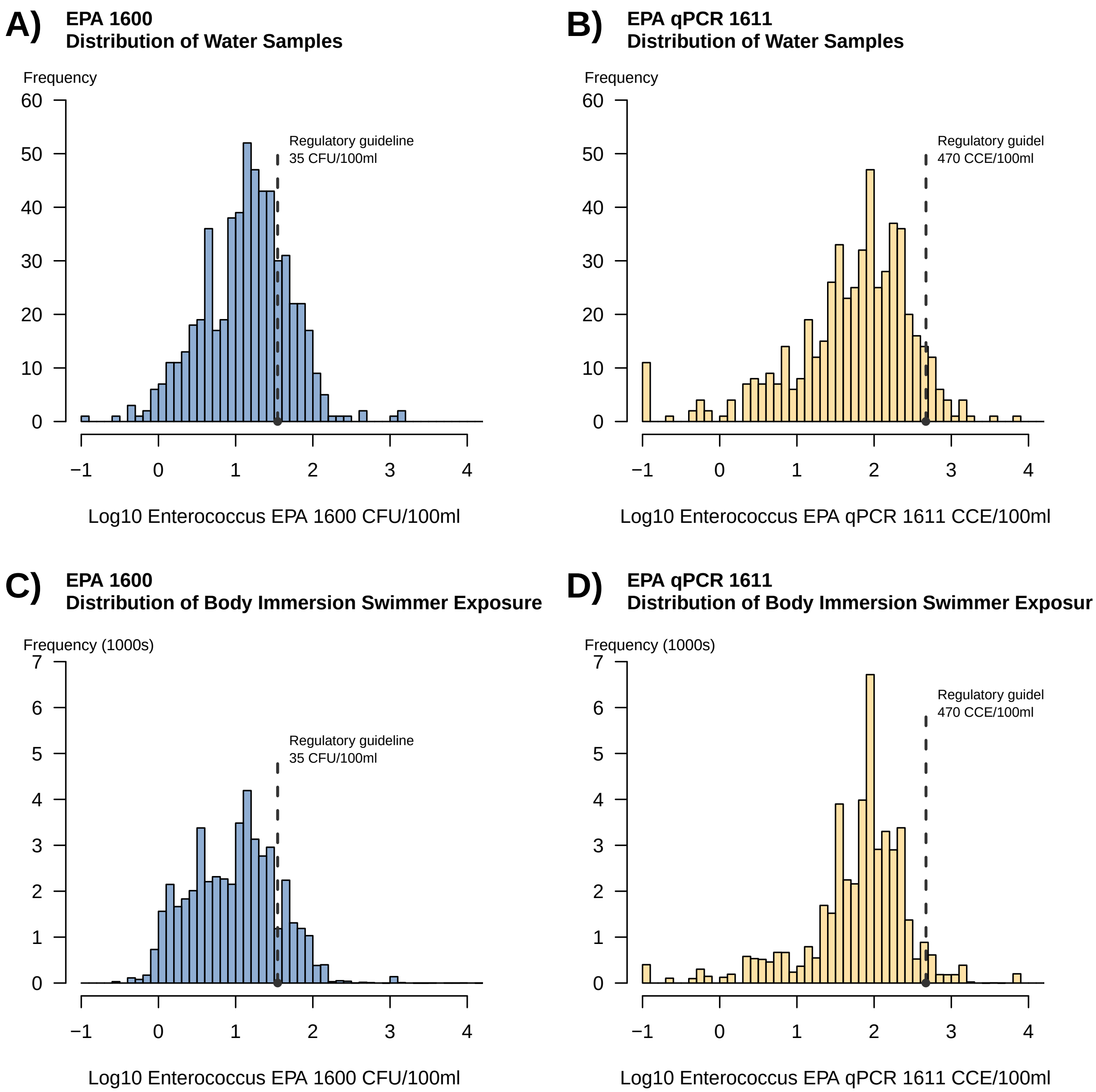

Figure S4: Distribution of Enterococcus Water Samples and Enterococcus Exposure Among Body Immersion Swimmers Matched to Water Samples. Bin width is 0.1. **A)** Enterococcus EPA 1600 water sample distribution. **B)** Enterococcus EPA qPCR 1611 water sample distribution. **C)** Enterococcus EPA 1600 body immersion swimmer distribution. **D)** Enterococcus EPA qPCR 1611 body immersion swimmer distribution. All figures exclude 73 water samples from Malibu and Doheny beaches freshwater lagoons.

3 *Enterococcus* EPA 1611 qPCR Results

In this section we summarize the association between *Enterococcus* EPA 1611 qPCR fecal indicator bacteria levels and incident diarrhea among body immersion swimmers. Similar to the results for the culture-based method EPA 1600, we observed no consistent relationship between *Enterococcus* EPA 1611 qPCR levels and incident diarrhea among children ages 0-4 years (Figure S5). However, unlike the relationship observed between EPA 1600 levels and illness for older ages (main text, Figure 2), there was not a gradual dose-response relationship between quartiles of the qPCR indicator and diarrhea incidence – instead, the risk increased most in the highest quartile of the distribution, with intermediate increases in quartiles 2 and 3 (Figure S5).

As noted in the main text, we found that there was some evidence effect modification of the association between the regulatory cutoff value for EPA 1600 and diarrhea incidence by whether or not the beach had a known point-source of treated human sewage nearby (main text, Figure 3). In contrast, we found no evidence for effect modification of the EPA 1611 qPCR level and diarrhea incidence by point-source versus non-point source beaches (liklihood ratio test $P = 0.63$) and the adjusted cumulative incidence ratio was similar across all conditions (Figure S6). Although there were relatively few individuals exposed above the regulatory cutoff for the EPA 1611 qPCR assay (N=2,338), these results suggest that values of the assay that exceed the regulatory guideline is a consistent indicator of health risk even at beaches with no known nearby point source of human fecal pollution.

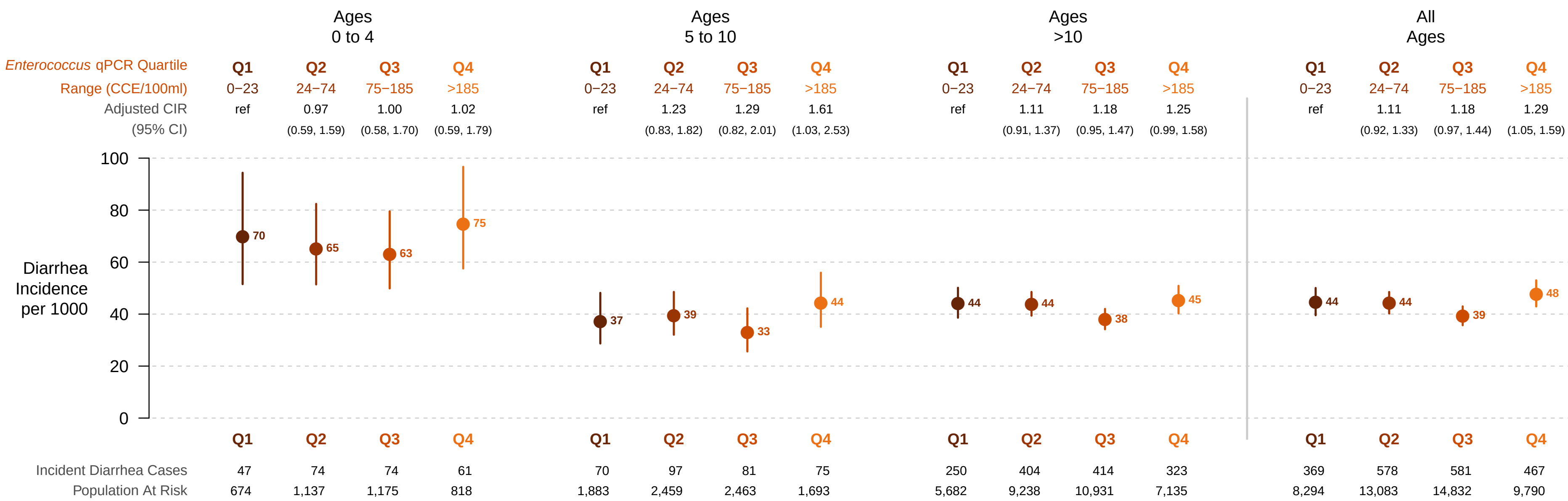

Figure S5: Incident Diarrhea Among Body Immersion Swimmers Associated with Quartiles of Enterococcus EPA qPCR 1611 Concentration, Stratified by Age.

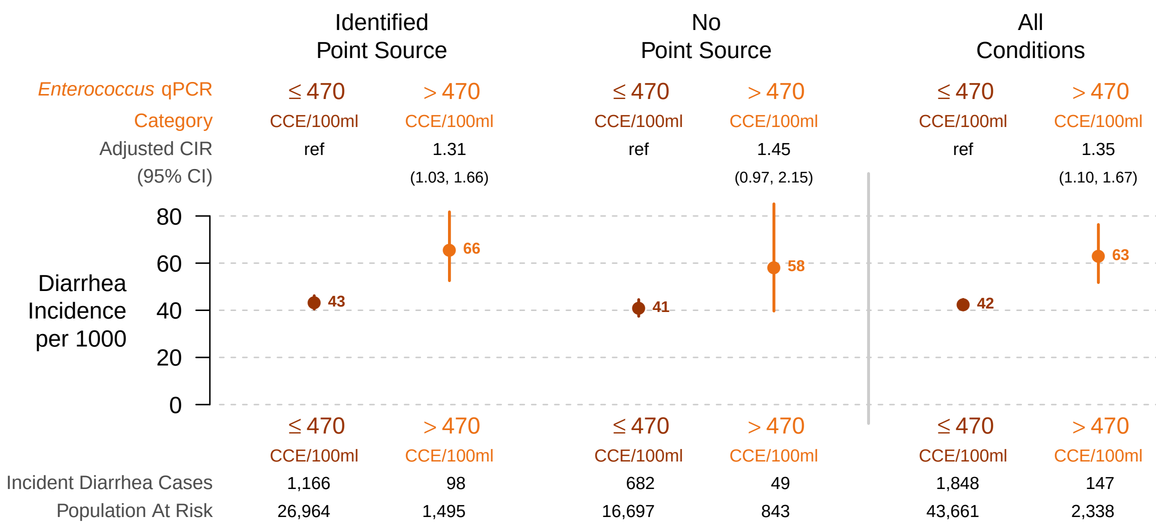

Figure S6: Incident Diarrhea Among Body Immersion Swimmers Associated with Enterococcus EPA qPCR 1611 Concentration Above and Below Regulatory Guidelines, Stratified by Beach Conditions (P-value for effect modification = 0.63).

4 Beach-Specific Results

This section reports beach-specific estimates of the association between: i) water exposure and diarrhea risk, and ii) *Enterococcus* exposure and diarrhea risk, corresponding to the first two study objectives.

Figure S7 summarizes the adjusted cumulative incidence ratio (CIR) of diarrhea associated with different levels of water exposure at each beach. The results are sorted by fresh versus marine water and then by the CIR for body immersion swimmers. There was evidence for effect modification by fresh versus marine water conditions, with stronger associations between water exposure and incident diarrhea at freshwater beaches: P-value for effect modification = 0.05 (body immersion), 0.08 (head immersion), 0.05 (swallowed water). Summary statistics of heterogeneity across beaches cite[Higgins2002, Higgins2003] were consistent with low to moderate levels of effect hetergeneity in the estimates: body immersion (Cohen's Q = 18.52, Pr(Q,df=12)=0.10, $I^2 = 35\%$), head immersion (Cohen's Q= 15.72, Pr(Q,df=12)=0.20, $I^2 = 24\%$), and swallowed water (Cohen's Q= 21.60, Pr(Q,df=12)=0.04, $I^2 = 44\%$).

Figure S8 summarizes the adjusted CIR of diarrhea associated with exposure to water above USEPA regulatory guidlines for body imerssion swimmers at each beach. We observed more heterogeneity in the effects compared to the swim exposure analysis, particularly for the EPA 1611 qPCR indicator analyses. Part of this is due to relatively small numbers of individuals exposed above the qPCR regulatory guideline of 470 CCE/100ml (Figure S4). Indeed, we have not presented the beach-specific CIRs associated with quartiles of *Enterococcus* concentrations due to small sample sizes at each beach for that type of granular analysis. There was some evidence for effect modification by point versus non-point source beaches for the EPA 1600 indicator (P-value for effect modification = 0.32). The indicator was associated with increased diarrhea incidence only at beaches with known point sources of treated human sewage (Figure S8). In contrast, we observed no effect modification by beach type of the relationship between the EPA 1611 qPCR indicator and diarrhea incidence (P-value for effect modification = 0.62).

XXX placeholder for heterogeneity statistics

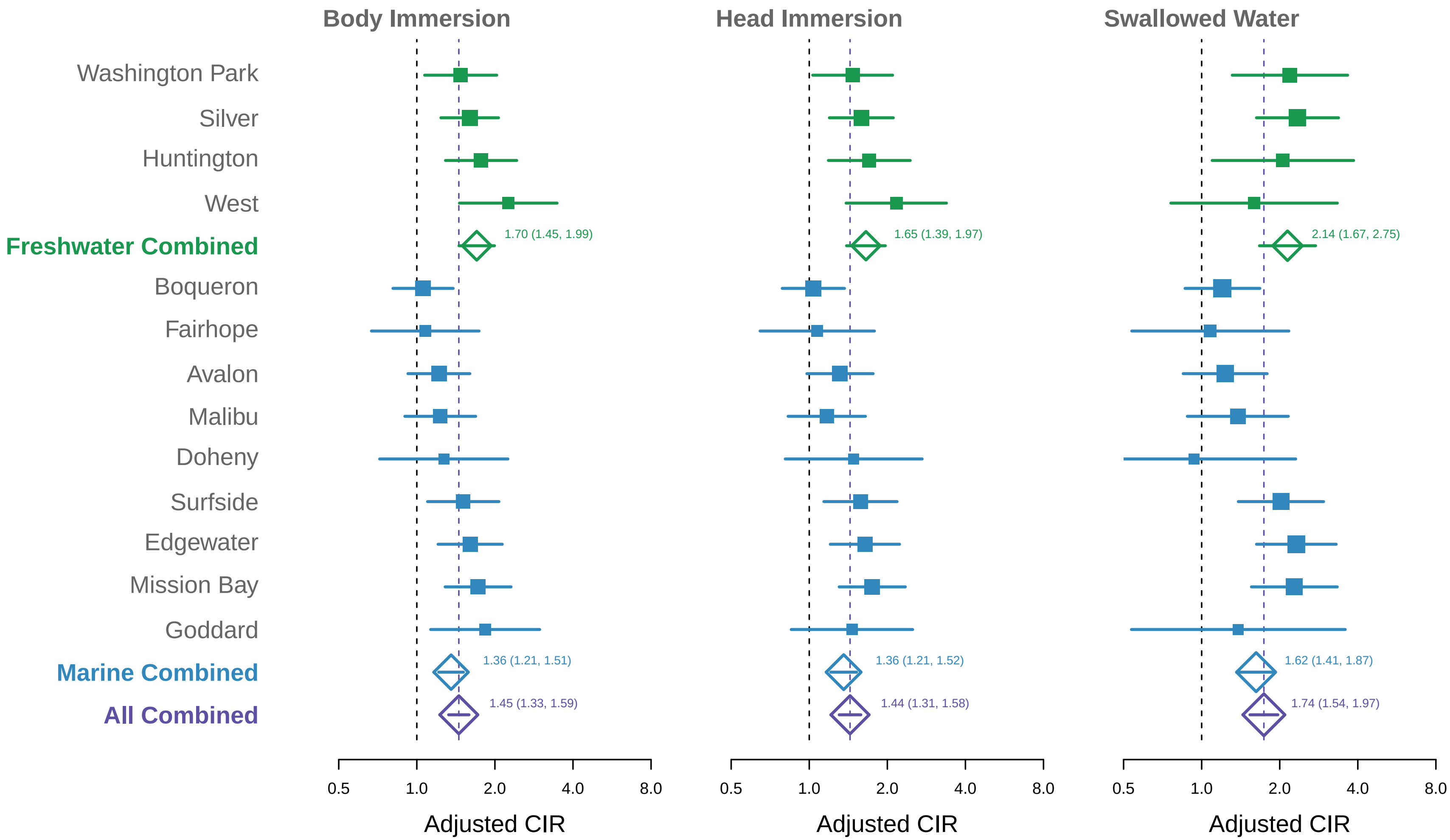

Figure S7: Forest Plot of Adjusted Cumulative Incidence Ratios (CIR) of Diarrhea Comparing Swimmers with Nonswimmers by Beach, Water Type, and Water Exposure Level. Beach-specific estimates (boxes) are scaled by the inverse of the standard error of the CIR, pooled estimates are represented by diamonds, and horizontal lines mark 95% confidence intervals.

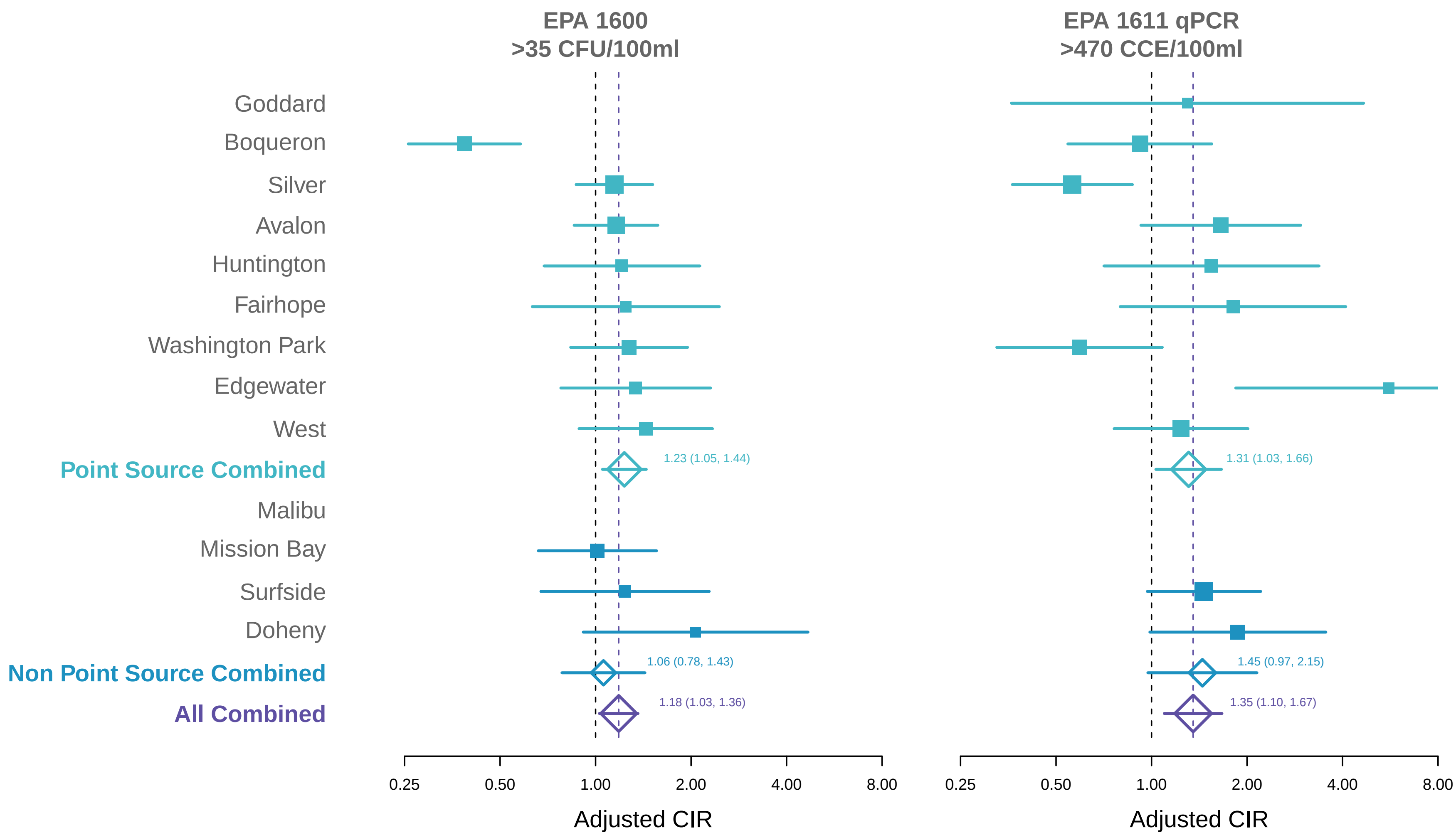

Figure S8: Forest Plot of Adjusted Cumulative Incidence Ratios (CIR) of Diarrhea Among Body Immersion Swimmers Associated with Levels of *Enterococcus* Above USEPA Regulatory Guidelines for EPA 1600 and EPA 1611 qCPR Indicators. There are no estimates for Malibu (1600 and 1611), Goddard (1600) or Mission Bay (1611) because too few swimmers were exposed above the regulatory level. Beach-specific estimates (boxes) are scaled by the inverse of the standard error of the CIR, pooled estimates are represented by diamonds, and horizontal lines mark 95% confidence intervals.

5 Negative Control and Sensitivity Analyses

This section includes a summary of pre-specified negative control analyses to test the robustness of the findings to bias from unmeasured confounding or measurement error, as well as pre-specified sensitivity analyses to ensure that our results were not an artifact of decisions made about swim exposure level, choice of modeling strategy (quartiles versus continuous exposure) or length of follow-up period (ranging from 1 to 10 days).

Negative Control Analyses: We matched *Enterococcus* levels measured in the water to people who were at the beach that day but did not have any water contact as a negative control exposure analysis xxcite[Lipsitch2010, Arnold2015, Colford2012] – without water exposure, the association between *Enterococcus* levels and diarrhea incidence should be null, and indeed that is what we observed (Figure S9). This lends additional credibility to the results because it suggests that the effects observed are not an artifact of unobserved confounding or measurement error xxcite[Lipsitch2010, Arnold2015].

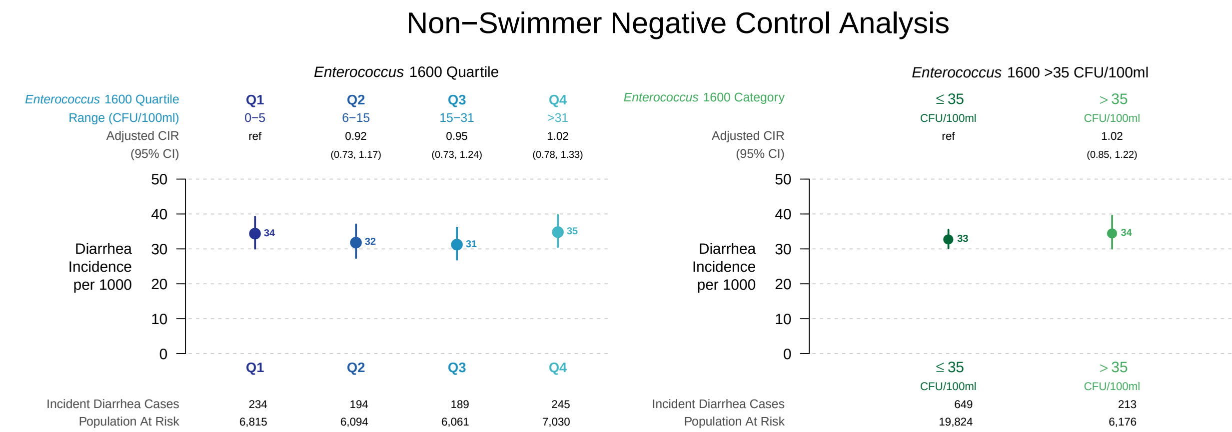

Figure S9: Negative Control Analyses. The relationship between *Enterococcus* EPA 1600 quartiles and diarrhea incidence among non-swimmers (left panel); and, the relationship between *Enterococcus* EPA 1600 above and below USEPA regulatory guidelines and diarrhea incidence among non-swimmers (right panel).

Sensitivity Analysis - Water Exposure Level: As described in our statistical analysis plan (available in the supporting information), we chose body immersion swimming as our primary swim exposure level for the *Enterococcus* analyses and population attributable risk analyses to ensure that we included the largest number of participants in the analysis and because we had observed in the individual studies that contributed to this meta-analysis that the risk was similar for body immersion swimming and head immersion swimming. When we estimated the association between *Enterococcus* levels and illness among swimmers with higher levels of exposure, we found similar results to those observed for body immersion swimmers in terms of relative risk measured by the cumulative incidence ratio, albeit with lower precision due to smaller sample sizes (Figure S10). Consistent with the swim exposure analyses, the overall incidence of diarrhea was higher among individuals who swallowed water, but the relative increase in risk with increasing concentration of *Enterococcus* was very similar to body immersion swimmers.

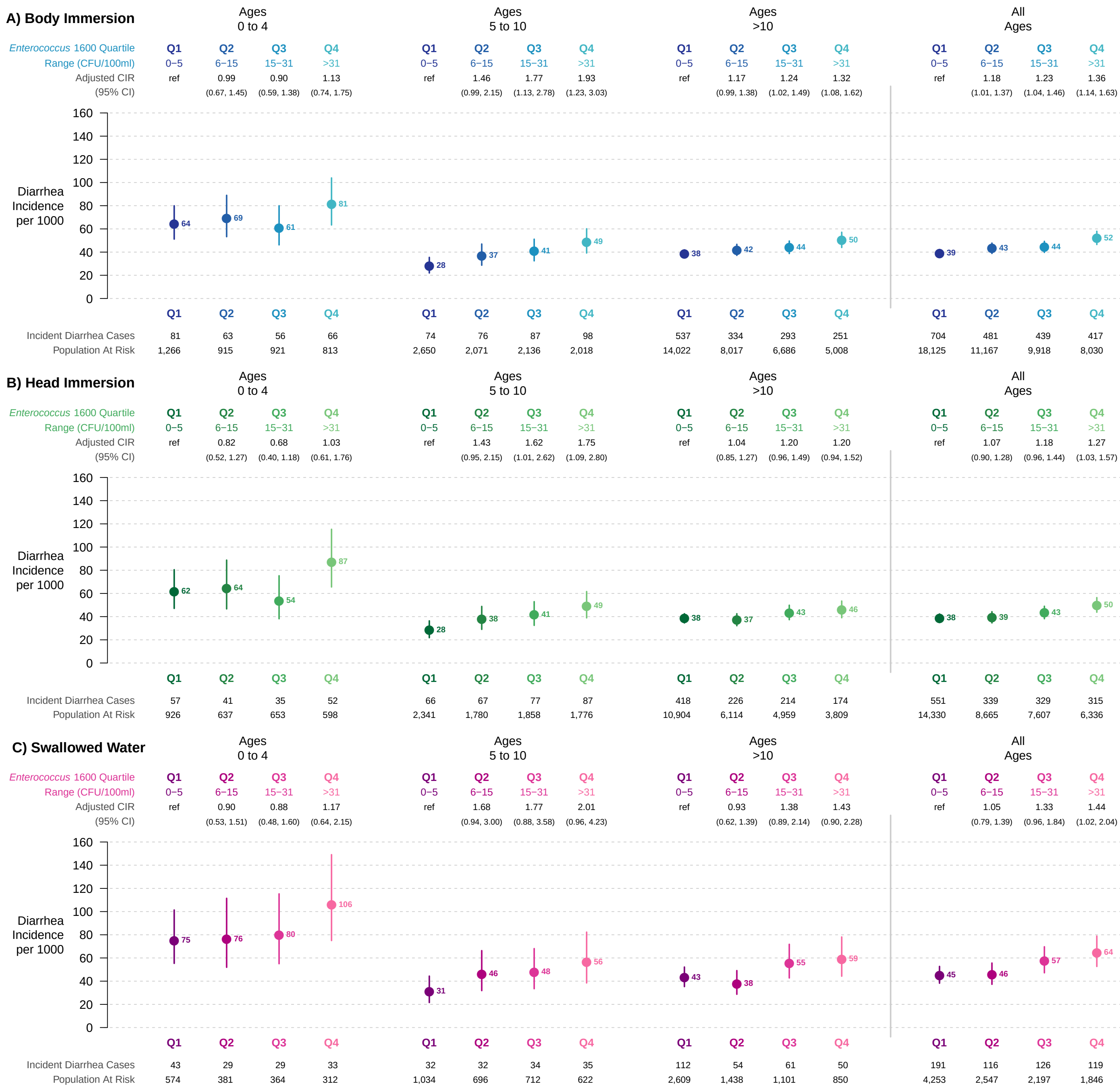

Figure S10: Association between *Enterococcus* EPA 1600 quartiles and diarrhea incidence by age. **A)** Among body immersion swimmers (identical to Figure 2 in the main text). **B)** Among head immersion swimmers. **C)** Among beachgoers who swallowed water.

Sensitivity Analysis - Continuous Exposure: To provide summary results that were consistent with the modeling approach used in the individual beach studies, we estimated the relationship between $\log_{10}$ *Enterococcus* concentrations on a continuous scale and the risk of diarrhea during follow-up. This analysis imposes stronger modeling assumptions on the relationship between *Enterococcus* concentrations and diarrhea risk, but is a complementary approach to the quartile analysis used in the main text. Consistent with our primary analysis, we modeled the probability of diarrhea during the 10 days of follow-up among body immersion swimmers using a log-linear, modified Poisson model with robust standard errors clustered at the household level. This model estimates the cumulative incidence ratio (CIR) associated with a $\log_{10}$ increase in *Enterococcus* concentration. We estimated adjusted CIRs using the same set of potential confounding variables used in the primary analysis (see the Statistical Analysis Plan in the Supporting Materials for details).

We calculated marginally adjusted exposure-response curves based on the model fit using marginal standardization over the empirical distribution of covariates in the study population xxcite[Ahern2009, Muller2014]. We calculated point-wise standard errors and 95% confidence intervals for the curves using a bootstrap that re-sampled households with replacement, stratified by beach and 1,000 iterations xxcite[Ahern2009].

We estimated that a $\log_{10}$ increase *Enterococcus* EPA 1600 diarrhea was associated with a 19% increase in the probability of diarrhea in the 10 days following the beach visit (CIR=1.19, 95%CI: 1.07, 1.34; Figure S11). Consistent with the primary analysis, we observed effect modification by age (likelihood ratio P=0.01), with no clear relationship among children ages 0-4 years, and the strongest relationship among children ages 5-10 years (Figure S12).

The increase in dirrhea incidence associated with a $\log_{10}$ increase *Enterococcus* EPA 1611 qPCR was similar to the EPA 1600 assay (CIR=1.15, 95%CI: 1.03, 1.28; Figure S13). We observed some effect modification by age (likelihood ratio P=0.003), though differences between age groups were less pronounced than for the EPA 1600 assay (Figure S14).

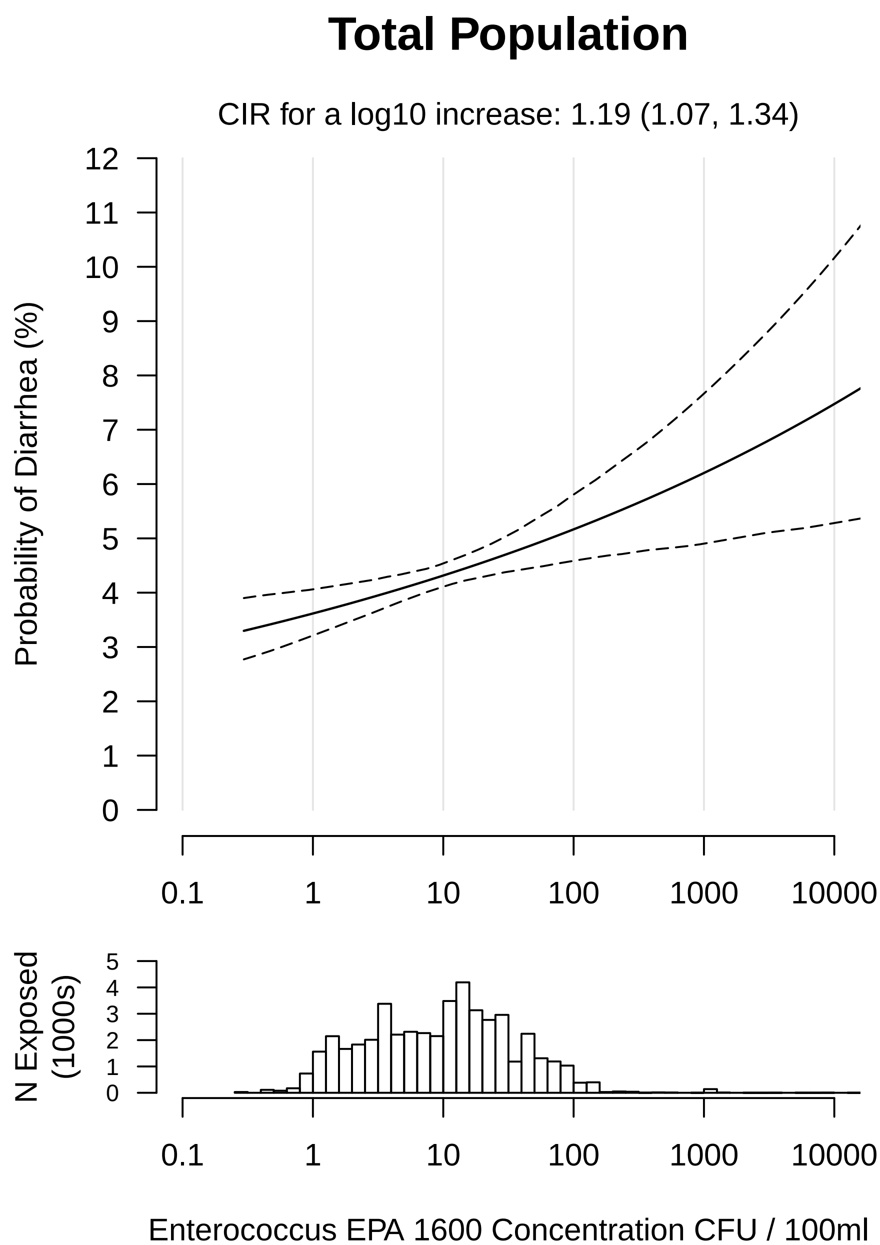

Figure S11: Association between *Enterococcus* EPA 1600 concentration and incident diarrhea among body immersion swimmers. The cumulative incidence ratio (CIR) and exposure-response curve were estimated using an adjusted log-linear regression model. Dashed lines are bootstrapped 95% confidence intervals.

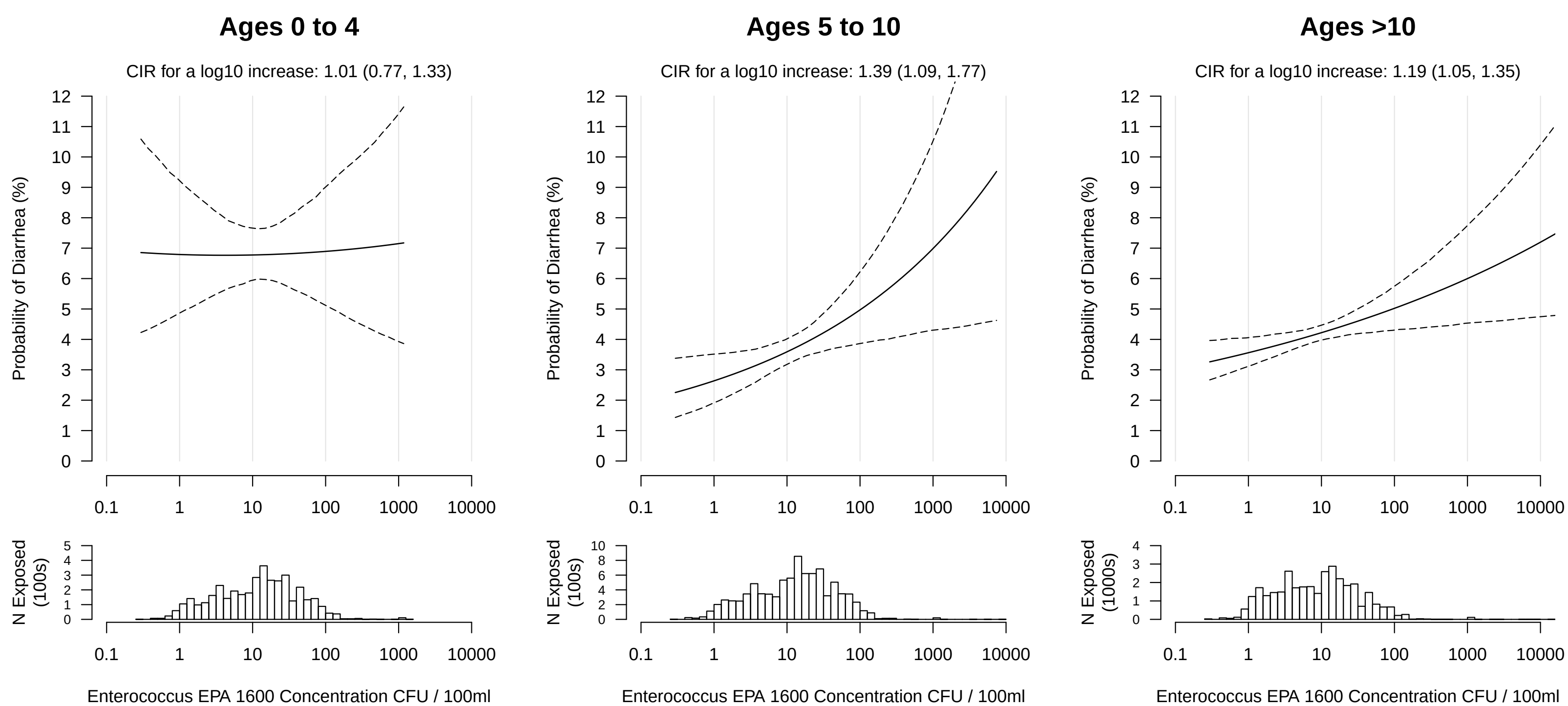

Figure S12: Association between *Enterococcus* EPA 1600 concentration and incident diarrhea among body immersion swimmers, stratified by age group. Cumulative incidence ratios (CIRs) and curves were estimated using a log-linear regression model. Note that the Y-scale differs on the number exposed across plots to display the distributions more clearly given differences in sample size across age groups. Dashed lines are bootstrapped 95% confidence intervals.

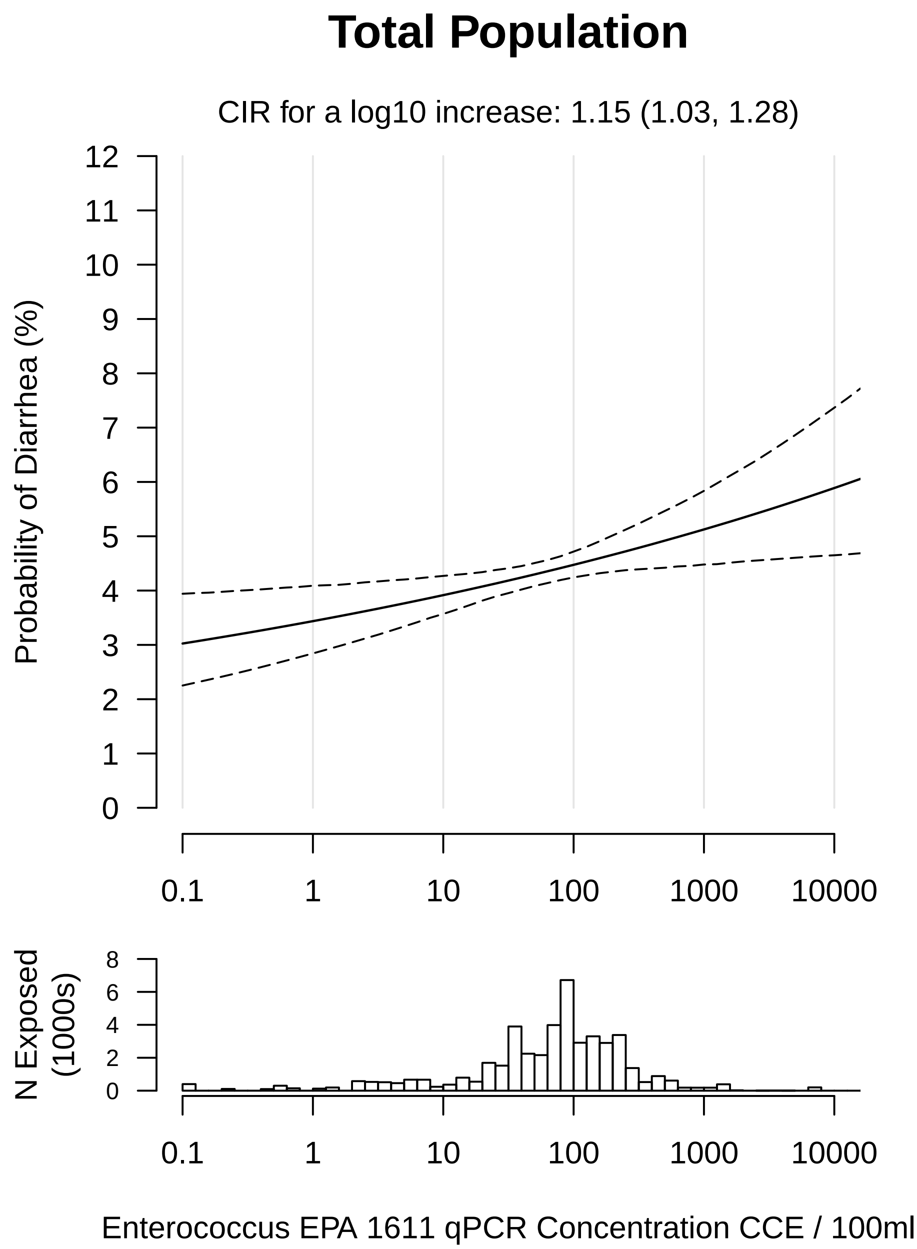

Figure S13: Association between *Enterococcus* EPA 1611 qPCR concentration and incident diarrhea among body immersion swimmers. The cumulative incidence ratio (CIR) and exposure-response curves were estimated using a log-linear regression model. Dashed lines are bootstrapped 95% confidence intervals.

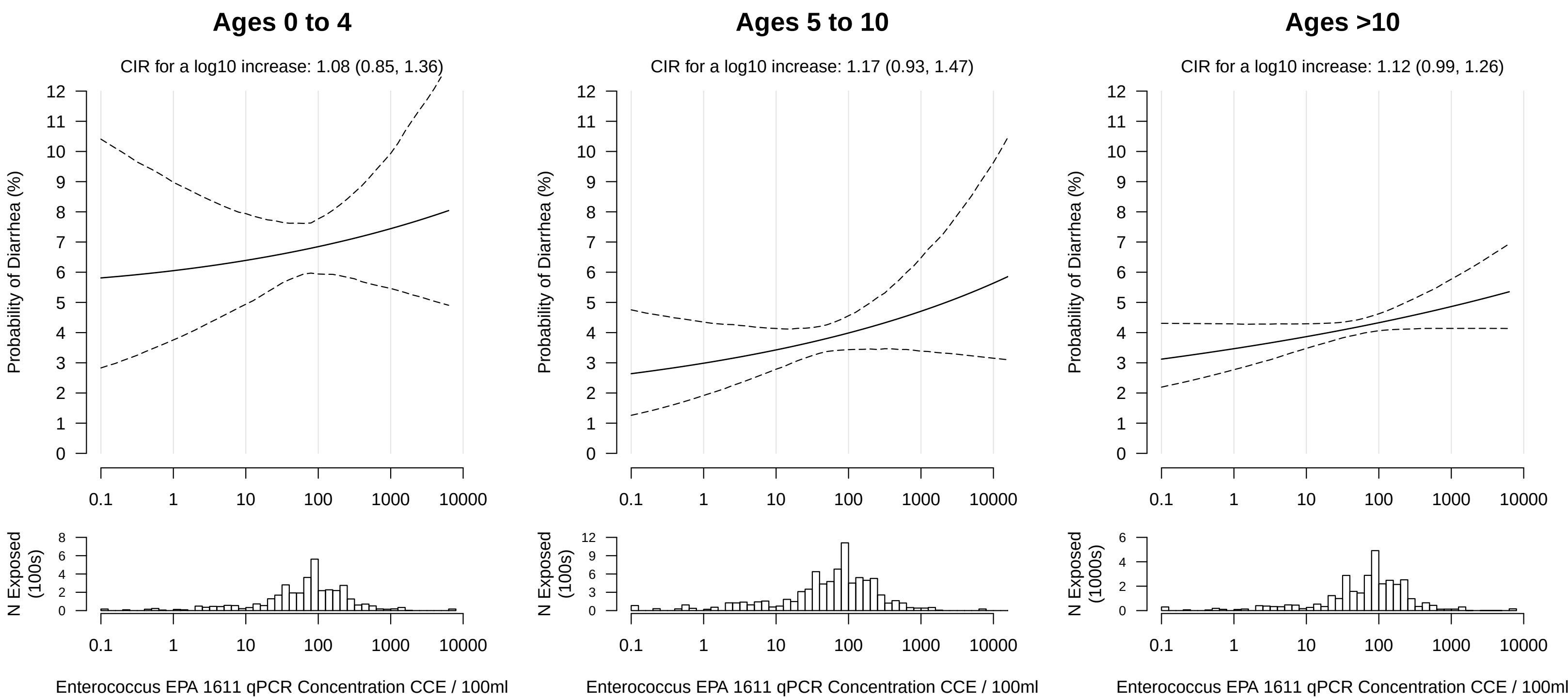

Figure S14: Association between *Enterococcus* EPA 1600 concentration and incident diarrhea among body immersion swimmers, stratified by age group. Cumulative incidence ratios (CIRs) and curves were estimated using an adjusted log-linear regression model. Note that the Y-scale differs on the number exposed across plots to display the distributions more clearly given differences in sample size across age groups. Dashed lines are bootstrapped 95% confidence intervals.

Sensitivity Analysis - Length of Follow-up: An earlier analysis from the Malibu cohort demonstrated that the greatest increase daily diarrhea incidence among swimmers compared to non-swimmers was in the first 3 days following the beach visit xxcite[Arnold2013]. When we examined daily incidence patterns across all cohorts, we observed a similar pattern (Figure S15). We additionally re-estimated the association between body immersion swim exposure and incident diarrhea, as well as *Enterococcus* exposure and incident diarrhea among swimmers (Figure S16). Consistent with the analysis of the Malibu cohort along xxcite[Arnold2013], there was an attenuation of the cumulative incidence ratio (CIR) associated with body immersion swim exposure with follow-up longer than 3 days (Figure S16, A). However, length of follow-up period did not greatly influence the association between *Enterococcus* exposure and diarrhea incidence – longer follow-up periods had similar CIR estimates to shorter periods, but with narrower confidence intervals due to substantially larger numbers of cumulative incident cases (Figure S16, B-C).

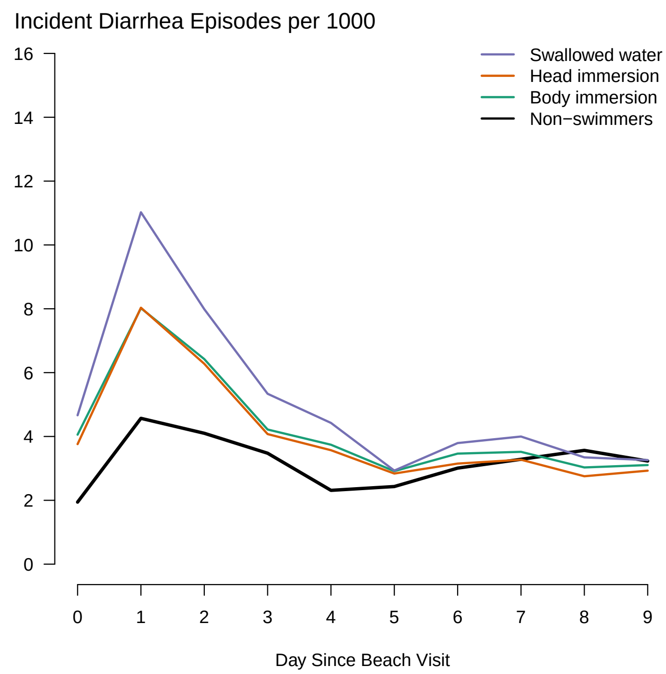

Figure S15: Daily Incidence of Diarrhea by Level of Water Exposure.

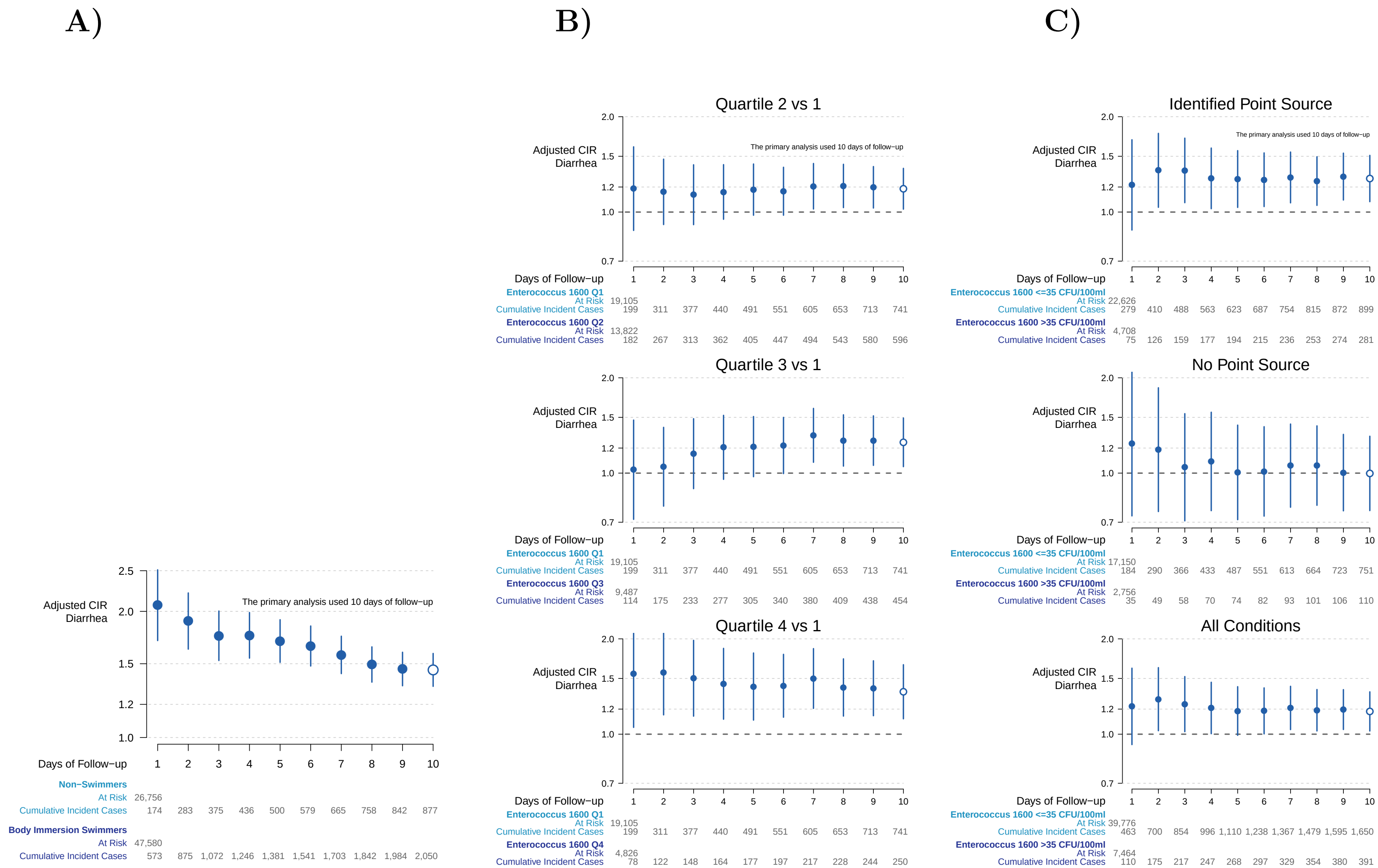

Figure S16: Sensitivity analysis of the length of follow-up period on cumulative incidence ratio (CIR) estimates. **A)** Body immersion swim exposure analysis. **B)** *Enterococcus* EPA 1600 quartile analysis. **C)** *Enterococcus* EPA 1600 >35 CFU/100ml analysis, stratified by beach type.

Population Attributable Risk, Stratified by Beach Conditions: As noted in the main text, we found that the relationship between *Enterococcus* EPA 1600 and swimmer illness was modified by the type of fecal inputs to the beach – the association between *Enterococcus* levels above USEPA regulatory guidelines and diarrhea incidence was clear for beaches with known "point sources" of treated human feces sewage nearby the beach, wheras there was no consistent association for "non-point source" beaches (main text, Figure 3). We therefore calculated estimates for the population attributable risk (PAR) and population attributable fraction (PAF) associated with swim exposure above USEPA regulatory guidelines for *Enterococcus* EPA 1600 (>35 colony forming units per 100ml), stratified by beach conditions (Table S2).

Consistent with the associations presented in the main text, the risk attributable to swimming in water above regulatory guidelines for *Enterococcus* EPA 1600 was primarily concentrated among swimmers at point source beaches. However, due to the overall small attributable risks, there was limited power to assess effect modification in this outcome.

Table S2: Population Attributable Risk Among Body Immersion Swimmers Due to Swimming in Water That Exceeds the USEPA Guideline of *Enterococcus* >35 CFU/100ml. Missed paid work and medical visits due to gastrointestinal illness were too rare to calculate stratified estimates (results in Table 2 of the main text).

	N Events	N At Risk	Predicted Incidence[1] per 1000: Observed Exposure	Predicted Incidence[1] per 1000: All $\leq$35 CFU/100ml	Population Attributable Risk[2] (95% CI)		Population Attributable Fraction[3] (95% CI)	
Diarrhea, episodes								
All Conditions	2,041	47,240	43	42	1.3	(0.2, 2.4)	3%	(1%, 6%)
Beach Conditions								
Point Source	1,264	28,459	44	42	2.1	(0.6, 3.9)	5%	(1%, 9%)
Non-point Source	777	18,781	41	41	0.2	(-1.0, 1.4)	1%	(-2%, 3%)
Gastrointestinal illness[4], episodes								
All Conditions	2,942	47,240	62	61	1.0	(-0.3, 2.3)	2%	(-0%, 4%)
Beach Conditions								
Point Source	1,804	28,459	63	62	1.5	(-0.3, 3.3)	2%	(-0%, 5%)
Non-point Source	1,138	18,781	61	60	0.4	(-1.1, 2.0)	1%	(-2%, 3%)
Missed Daily Activities[5], days								
All Conditions	2,677	47,240	57	56	0.6	(-1.5, 2.6)	1%	(-3%, 5%)
Beach Conditions								
Point Source	1,716	28,459	60	60	0.9	(-2.0, 3.9)	1%	(-3%, 6%)
Non-point Source	961	18,781	51	51	0.3	(-2.6, 3.4)	1%	(-5%, 7%)

1. Predicted incidence per 1000 among body immersion swimmers under the empirical distribution of *Enterococcus* exposure (observed) and under a counterfactual scenario where nobody entered the water in conditions >35 CFU/100ml. Estimates are from a multivariable regression model adjusted for a range of potential confounders and beach level fixed-effects (see text for details).

2. Population Attributable Risk is the number of events per 1000 swimmers that would be prevented if the exposure of swimming in water with *Enterococcus* $\geq$35 CFU/100ml were removed from the population. The proportion of swimmers exposed to water with *Enterococcus* EPA 1600 >35 CFU/100ml was: all conditions (16%), point source conditions (19%), non-point source conditions (12%).

3. Population Attributable Fraction is the percentage of events among swimmers attributable to swimming in water with *Enterococcus* $\geq$35 CFU/100ml.

4. Gastrointestinal illness was defined as (i) diarrhea or (ii) vomiting or (iii) stomach cramps and missed daily activities or (iv) nausea and missed daily activities.

5. Includes days of school, work, or vacation missed because of gastrointestinal illness.