Slope Estimates for Freshwater and Marine Epidemiology Studies

Timothy J. Wade

May 17, 2012

All information provided is preliminary and subject to change
Do not cite quote or distribute

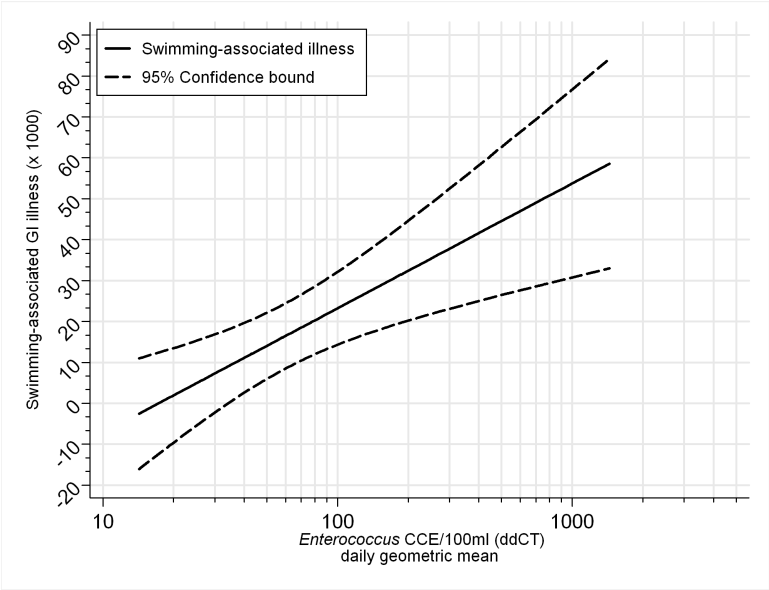
1 Introduction

This article provides preliminary revised regression relationships for the Attributable Risk of GI illness for and fecal indicator bacteria from the NEEAR Water Study [1]. The calculation for qPCR Calibarator Cell Equivalents (CCE) is described in [2].

2 *Enterococcus* $CCE_{\Delta\Delta}$

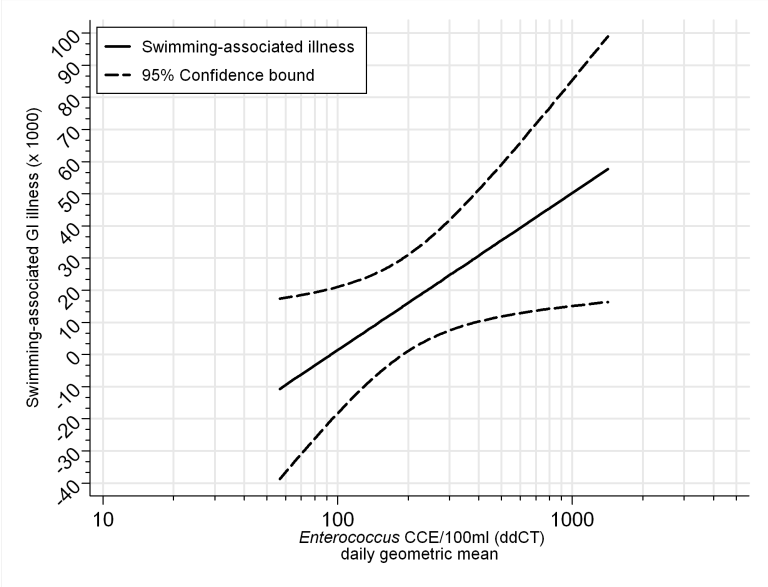
2.1 Freshwater Beaches

Figure 1: Swimming-Associated GI illness and Daily Average *Enterococcus* qPCR $CCE_{\Delta\Delta}$. All Subjects, freshwater beaches



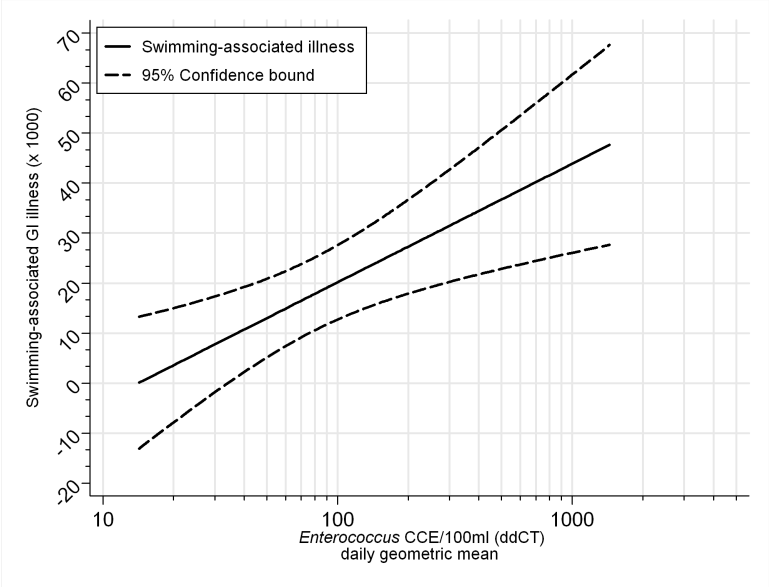
2.2 Marine Beaches

Figure 2: Swimming-Associated GI illness and Daily Average *Enterococcus* qPCR $CCE_{\Delta\Delta}$. All Subjects, marine beaches



2.3 Marine and Freshwater Beaches Combined

Figure 3: Swimming-Associated GI illness and Daily Average *Enterococcus* qPCR $CCE_{\Delta\Delta}$. All Subjects, marine and freshwater beaches



2.4 Comparison of Risks for Marine and Freshwater Beaches

Figure 4: Swimming-Associated GI illness and Daily Average *Enterococcus* qPCR $CCE_{\Delta\Delta}$. All Subjects, marine and freshwater beaches

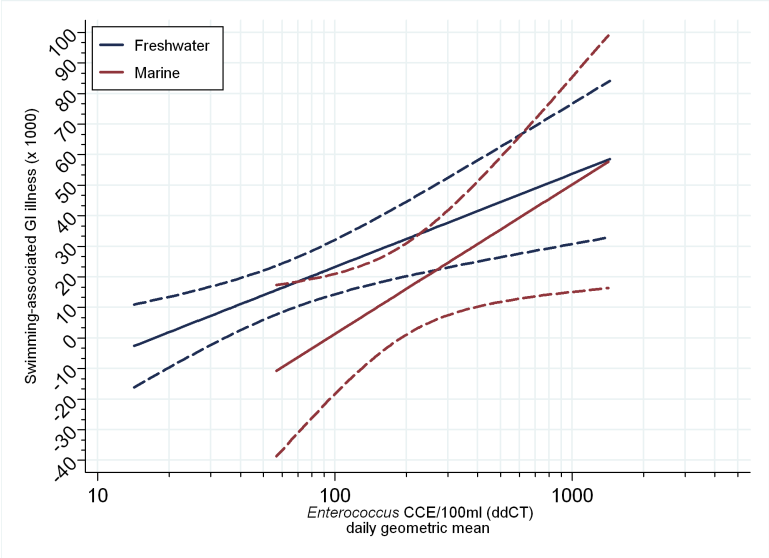


Table 1 shows that above approximately 126 *Enterococcus* $\frac{CCE}{100ml}$, risks for marine and freshwater beaches cannot be distinguished. Moreover, there is no difference in the slopes, or the rate of increase in risk per unit increase in *Enterococcus* CCE, between marine and freshwater beaches.

Table 1: Comparison of risks for marine, fresh and combined relationships *Enterococcus* qPCR CCE (per 100 ml) and GI illness

Contrast	**p-value**[1]	**AR-fresh**[2]	**AR-marine**	**AR-combined**
Test of Slopes	.4512			
Test at 32 CCE	.1201	7.98	-23.1	8.28
Test at 40 CCE	.1023	11.03	-18.22	10.66
Test at 50 CCE	.0847	14.08	-13.33	13.03
Test at 63 CCE	.0685	17.13	-8.44	15.4
Test at 79 CCE	.0551	20.18	-3.55	17.78
Test at 100 CCE	.0467	23.23	1.34	20.15
Test at 126 CCE	.0462	26.28	6.23	22.52
Test at 158 CCE	.059	29.33	11.12	24.89
Test at 200 CCE	.0963	32.38	16	27.27
Test at 251 CCE	.1709	35.43	20.89	29.64
Test at 316 CCE	.284	38.47	25.78	32.01
Test at 398 CCE	.4185	41.52	30.67	34.38
Test at 501 CCE	.5536	44.57	35.56	36.76
Test at 631 CCE	.6762	47.62	40.45	39.13
Test at 794 CCE	.7817	50.67	45.34	41.5
Test at 1000 CCE	.8703	53.72	50.23	43.88
Test at 1259 CCE	.9441	56.77	55.11	46.25

1: p-value for test of freshwater and marine risks. 2: Attributable risk per 1000

2.5 Comparison with Children

Previous studies [1] suggested children 10 years of age and under may be at increased susceptibility to illness following exposure. Although this effect could not be confirmed at the marine beach sites, the smaller sample size at the marine sites reduced the power for subgroup analysis. Having established in the above section that slope and risk estimates of marine and freshwater beaches do not differ appreciably, children were examined separately as a subset of all beach sites combined. Results are compared for 10 and under are compared to all respondents as well as to those above 10. The approach used for testing cross-model hypotheses of non-nested models, similar to that described in Section 2.4 for testing differences for marine and freshwater beaches has been described previously [3, 4]. As shown in Figures 5 and 6, although the slope for children is steeper, there is considerable overlap in risk estimates across the ranges of exposure.

Figure 5: Comparisons of risks for children 10 and under with participants over 10 years of age. Swimming-Associated GI illness and Daily Average *Enterococcus* qPCR $CCE_{\Delta\Delta}$.

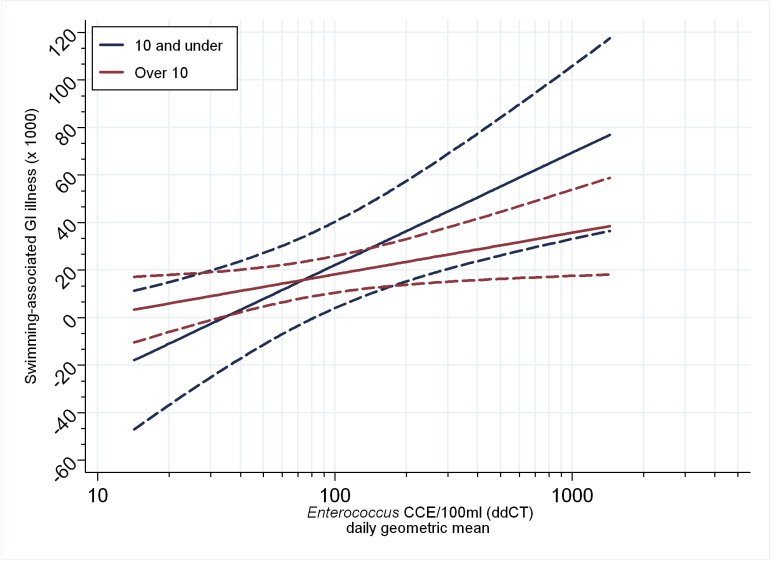


Figure 6: Comparisons of risks for children 10 and under to all subjects combined. Swimming-Associated GI illness and Daily Average *Enterococcus* qPCR $CCE_{\Delta\Delta}$.

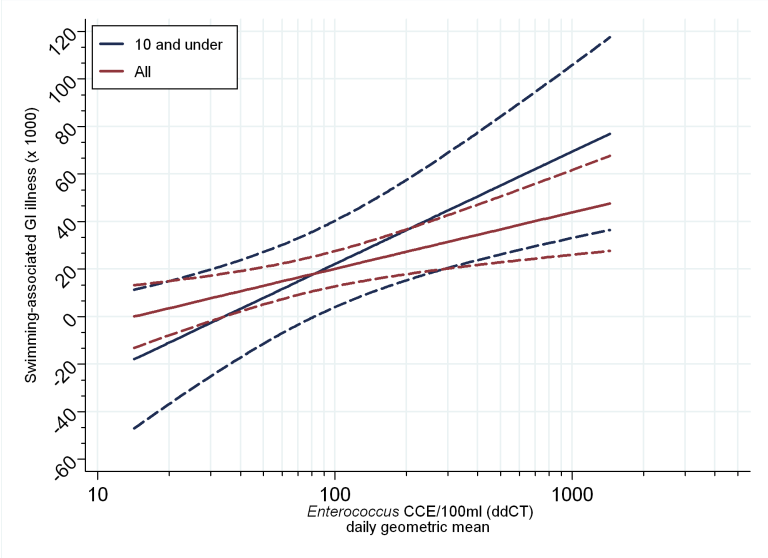


Table 2: Comparison of risks for children 10 and under, all subjects and those over 10 *Enterococcus* qPCR $CCE_{\Delta\Delta}$ (per 100 ml) and GI illness

Contrast	**p-value**[1]	**p-value**[2]	**AR-10 and under**[3]	**AR-over 10**	**AR-all**
Test of Slopes	.0537				
Test of Slopes		.0661			
Test at 32 CCE	.3178	.3421	-1.49	9.8	8.28
Test at 40 CCE	.4174	.4503	3.23	11.57	10.66
Test at 50 CCE	.5578	.6042	7.96	13.35	13.03
Test at 63 CCE	.7436	.8069	12.69	15.12	15.4
Test at 79 CCE	.9652	.9573	17.42	16.9	17.78
Test at 100 CCE	.8044	.7237	22.15	18.68	20.15
Test at 126 CCE	.5979	.5268	26.88	20.45	22.52
Test at 158 CCE	.4355	.3814	31.61	22.23	24.89
Test at 200 CCE	.3194	.2826	36.34	24	27.27
Test at 251 CCE	.2406	.2177	41.07	25.78	29.64
Test at 316 CCE	.188	.1752	45.8	27.55	32.01
Test at 398 CCE	.1528	.1469	50.53	29.33	34.38
Test at 501 CCE	.1287	.1275	55.26	31.11	36.76
Test at 631 CCE	.1118	.1138	59.99	32.88	39.13
Test at 794 CCE	.0997	.1039	64.71	34.66	41.5
Test at 1000 CCE	.0907	.0966	69.44	36.43	43.88
Test at 1259 CCE	.084	.0911	74.17	38.21	46.25

1: p-value for under 10 vs. all subjects

2: p-value for under 10 vs. over 10

3: Attributable risk

2.6 Summary Estimates-***Enterococcus*** $CCE_{\Delta\Delta}$

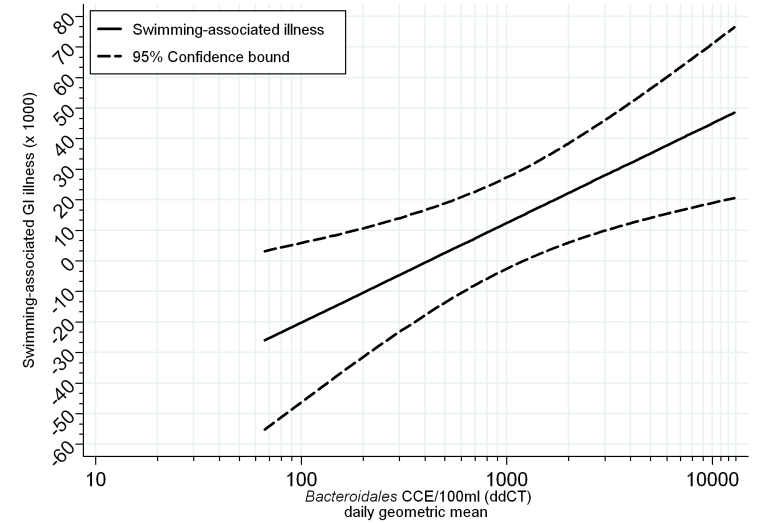
Table 3: Summary of Slope, Intercept and Variance components for *Enterococcus CCE*

Indicator	**Slope**	**Intercept**	**Vswim**	**Vexp**	**Cov**
Enterococcus CCE-Freshwater	0.03048980	-0.03775042	0.000270767	0.000080585	-0.000143137
Enterococcus CCE-Marine	0.04888604	-0.09643296	0.002755392	0.000515744	-0.001179437
Enterococcus CCE-Combined	0.02372795	-0.02730777	0.000217944	0.000056644	-0.000107531
Enterococcus CCE-Children	0.04729246	-0.07243349	0.000928537	0.000226596	-0.000437267

3 Bacteroidales $CCE_{\Delta\Delta}$ (Marine sites only)

textitBacteroidales $CCE_{\Delta\Delta}$ were not measured at freshwater beaches, so results are presented only for marine beaches.

Figure 7: Swimming-Associated GI illness and Daily Average *Bacteroidailes* qPCR $CCE_{\Delta\Delta}$. All Subjects, marine beaches



3.1 Comparison with Children-*Bacteroidales* CCE

As shown in Figures 8 and 9, although the slope for children is steeper, there is considerable overlap in risk estimates across the ranges of exposure.

Figure 8: Comparisons of risks for children 10 and under with participants over 10 years of age. Swimming-Associated GI illness and Daily Average *Bacteroidales* qPCR $CCE_{\Delta\Delta}$.

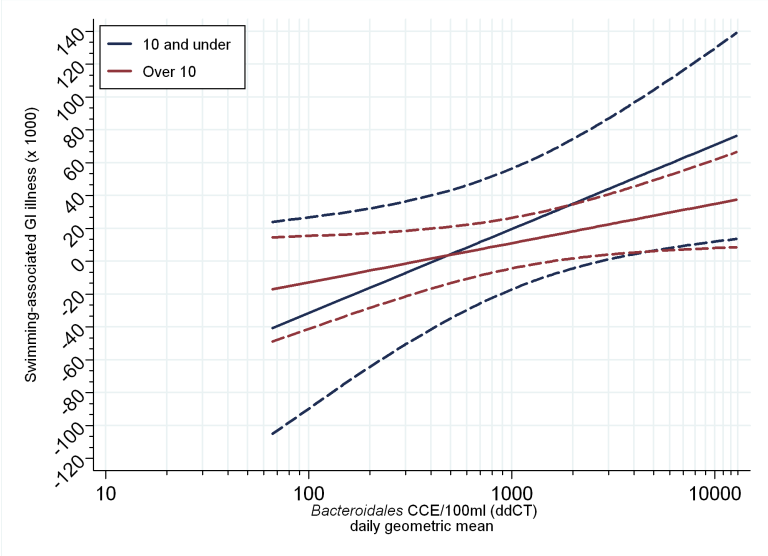


Figure 9: Comparisons of risks for children 10 and under to all subjects combined. Swimming-Associated GI illness and Daily Average *Bacteroidales* qPCR $CCE_{\Delta\Delta}$.

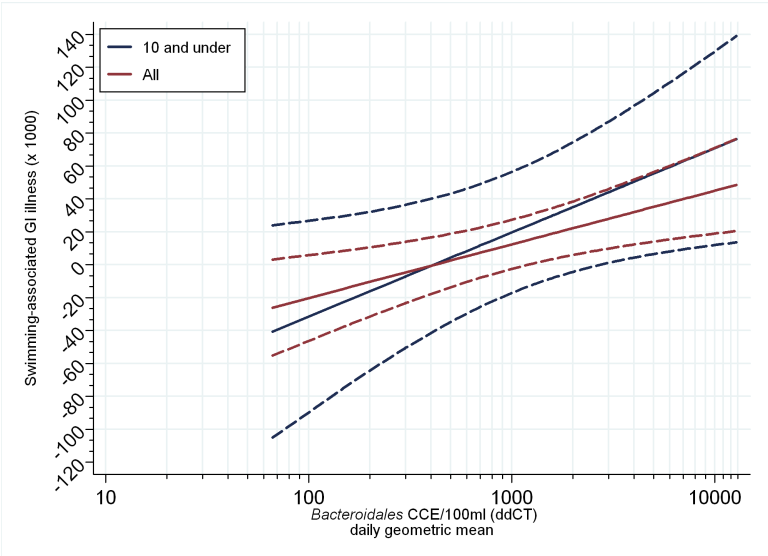


Table 4: Comparison of risks for children 10 and under, all subjects and those over 10 *Bacteroidales* qPCR $CCE_{\Delta\Delta}$ and GI illness associations (Marine sites only)

Contrast	**p-value**[1]	**p-value**[2]	**AR-10 and under**[3]	**AR-over 10**	**AR-all**
Test of Slopes	.3324				
Test of Slopes		.273			
Test at 63 CCE	.5989	.4995	-41.69	-17.58	-26.72
Test at 79 CCE	.626	.5252	-36.58	-15.19	-23.46
Test at 100 CCE	.6576	.5559	-31.46	-12.8	-20.2
Test at 126 CCE	.6947	.5926	-26.35	-10.42	-16.93
Test at 158 CCE	.7382	.6367	-21.23	-8.03	-13.67
Test at 200 CCE	.7891	.6897	-16.12	-5.64	-10.41
Test at 251 CCE	.8483	.7534	-11	-3.25	-7.15
Test at 316 CCE	.9165	.8289	-5.89	-.86	-3.89
Test at 398 CCE	.9934	.9167	-.77	1.53	-.63
Test at 501 CCE	.922	.9842	4.34	3.92	2.64
Test at 631 CCE	.8325	.8772	9.45	6.31	5.9
Test at 794 CCE	.7422	.7679	14.57	8.7	9.16
Test at 1000 CCE	.6556	.6633	19.68	11.09	12.42
Test at 1259 CCE	.5771	.5697	24.8	13.48	15.68
Test at 1585 CCE	.5096	.4907	29.91	15.87	18.94
Test at 1995 CCE	.4541	.4275	35.03	18.26	22.21
Test at 2512 CCE	.4101	.3786	40.14	20.65	25.47
Test at 3162 CCE	.3761	.3418	45.26	23.04	28.73
Test at 3981 CCE	.3503	.3144	50.37	25.43	31.99
Test at 5012 CCE	.331	.2941	55.48	27.82	35.25
Test at 6310 CCE	.3167	.2791	60.6	30.21	38.51
Test at 7943 CCE	.3061	.268	65.71	32.6	41.78
Test at 10000 CCE	.2983	.2598	70.83	34.99	45.04
Test at 12589 CCE	.2926	.2537	75.94	37.38	48.3

1: p-value for under 10 vs. all subjects

2: p-value for under 10 vs. over 10

3: Attributable risk

3.2 Summary Estimates, *Bacteroidales* CCE

Table 5: Summary of Slope, Intercept and Variance components for *Bacteroidales* $CCE_{\Delta\Delta}$

Indicator	Slope	Intercept	Vswim	Vexp	Cov
Bacteroidales CCE-10 and under	0.05114420	-0.13374966	0.003629911	0.000369534	-0.001113911
Bacteroidales CCE	0.03261677	-0.08542899	0.001120855	0.000118242	-0.000354508

References

[1] Wade TJ, Calderon RL, Brenner KP, Sams E, Beach M, Haugland R, et al. High sensitivity of children to swimming-associated gastrointestinal illness: results using a rapid assay of recreational water quality. Epidemiology. 2008;19(3):375–383.

[2] Wade T, Sams E, Brenner K, Haugland R, Chern E, Beach M, et al. Rapidly measured indicators of recreational water quality and swimming-associated illness at marine beaches: a prospective cohort study. Environmental Health. 2010;9(1):66. Available from: `http://www.ehjournal.net/content/9/1/66`.

[3] Weesie J. Seemingly unrelated estimation and the cluster-adjusted sandwich estimator. Stata Technical Bulletin. 1999;52:34–47.

[4] Clogg CC, Petkova E, Haritou A. Statistical methods for comparing regression coefficients between models. American Journal of Sociology. 1995;100(5):1261–1293.

[5] Cumming G. Inference by eye: reading the overlap of independent confidence intervals. Stat Med. 2009 Jan;28(2):205–220. Available from: `http://dx.doi.org/10.1002/sim.3471`.

[6] Wolfe R, Hanley J. If we're so different, why do we keep overlapping? When 1 plus 1 doesn't make 2. CMAJ. 2002 Jan;166(1):65–66.

[7] Altman DG, Bland JM. Interaction revisited: the difference between two estimates. BMJ. 2003 Jan;326(7382):219.

Appendices

A Methods

The AR is a difference estimated from a linear combination of two parameters from a regression model with an identity (linear) link function and binomial error structure (linear probability model). In brief, the regression model is as described:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_i X_i \ldots + \beta_j X_j \quad (1)$$

Where:

- Y is the risk (probability) of illness
- α is the overall intercept
- X_1 is a 1 or 0 indicator of swimming exposure (body immersion)
- X_2 is the log_{10} transformed fecal indicator daily-averaged geometric mean for those who reported body contact with water, zero otherwise
- $X_i \ldots X_j$ represent other covariates such as beach, age, etc.

Other covariates included in the model are described in [1]. For the delta-delta CT methods reported, the upper CT value was restricted to 38.45 for the freshwater and 2007 marine samples. For the 2005 marine analysis the upper CT was restricted the upper 95% limit of the Y-intercept of the calibrator standard curves. Samples failing the salmon assay criterion were replaced with the average of samples at the same depth and time which met the criterion. One-half the calculated cell equivalent value was used for non-detects. The adjusted attributable risk, or swimming associated illness can be described from the above model as:

$$AR = \beta_1 + ENT \times \beta_2 \quad (2)$$

Where ENT=log_{10} qPCR Cell Equivalents (or other fecal indicator measure) The standard error of the AR associated with a particular level of indicator is calculated as follows:

$$SE_{AR} = \sqrt{Vswim + ENT^2 \times Vexp + 2 \times ENT \times Cov} \quad (3)$$

Where

- ENT is the log_{10} qPCR cell equivalents
- Vswim is the variance of the swimming coefficient (X_1)
- Vexp is the variance of the indicator coefficient (X_2)
- Cov is the covariance of the indicator and swimming coefficient

The 95% confidence limits for any given attributable risk estimate are then calculated as follows:

$$AR_{95} = AR \pm 1.96 \times SE_{AR} \tag{4}$$

This can be generalized to any other confidence level by the following:

$$AR_{CL} = AR \pm \Phi^{-1}_{CL+\frac{1-CL}{2}} \times SE_{AR} \tag{5}$$

Where $\Phi^{-1}_{CL+\frac{1-CL}{2}}$ is the inverse cumulative normal density associated with the desired confidence limit (CL) at $CL + \frac{1-CL}{2}$. For example, the inverse normal density multiplier for a 90% confidence limit is: $\Phi^{-1}_{0.90+0.05} = 1.645$

A.1 Model selection

Covariates were in a process similar to that described previously [1, 2], with a few modifications. The process is summarized below.

- Fit logistic model with variables for swimming exposure and water quality measure and full set of covariates. Covariates included the following individual characteristics age, sex, other swimming exposure, contact with unkown animals, contact with other ill people, race, consumption of raw or undercooked meat, frequency of visits to this beach, miles travelled to the beach, and presence of chronic illnesses.

- Environmental measures to be included in the models were determined through principal components analysis. Rather than including the components themselves, which were specific to a particular site, major components were selected. These included: average air temperature, bather density and rainfall. At marine sites wave height was substituted for rainfall as it was a major contributing component.

- Select covariates whose inclusion in the model changes the coefficient of the water quality parameter more than 5% or the swimming parameter more than 10%.

The major differences from previous analyses are described below.

- For freshwater beaches, a study day at Silver Beach in 2004 when a large festival, the Venetian Festival, which drew 200,000 people to the beach was excluded from the analysis. This was done beacause of the large number (hundreds) of food and drink vendors on site which was likely to inflate the occurrence of gastrointestinal illness unrealted to water quality.

- For environmental measures, principal components themselves were not used directly in regression models as has been done previously.

- Swimmers were defined as those with body immersion exposure

- Non-swimmers were those who did not enter the water as well as those who entered the water without immersing their body.

As shown in Figure 4, the confidence bounds for risk estimates of swimming-associated illness (AR) for marine and freshwater slopes overlap across the entire range of observed exposures. However, since overlapping confidence intervals do not necessarily indicate no difference at $\alpha \leq 0.05$ [5, 6], a test statistic for the difference is as follows [7]:

$$z = AR_1 - AR_2 / SE\left(AR_1 - AR_2\right) \tag{6}$$

where:

- AR_1 is the estimated risk at a level of exposure in freshwater and AR_2 is the level in marine water as described in Equation 2.
- $SE\left(AR_1 - AR_2\right)$ is the standard error of the differences which has the standard error: $\sqrt{SE_{AR_1}^2 + SE_{AR_2}^2}$, where SE_{AR} is shown in Equation 3.

B Full models

Table 6: Full regression models and covariates used to develop slope estimates for *Enterococcus* qPCR $CCE_{\Delta\Delta}$

	Fresh	Marine	Combined	Ten and Under	Over Ten
Body Immersion	-0.038*	-0.096	-0.027	-0.072*	-0.017
Entero CCE	0.030***	0.049*	0.024**	0.047**	0.018*
Beach-FB	0.000	-0.015	-0.007	0.008	-0.010
Beach-GB	0.000	-0.040***	-0.034***	-0.032	-0.035**
Beach-HB	0.027***		0.016	0.034	0.009
Beach-SB	-0.011*		-0.021*	0.001	-0.023*
Beach-WB	-0.002		-0.009	0.013	-0.014
Beach-WP	0.000	0.000	-0.010	0.001	-0.010
Age (years)	-0.000**	-0.000	-0.000**		-0.000
Chronic Illness	0.032***				0.032***
Frequency visit beach		0.005			0.004
Miles travelled		-0.009**	-0.003	-0.005	-0.000
Wave Height		-0.034			
Bathers in water				-0.003	
Sex					0.015***
Animal contact					0.032***
Raw or rare meat					0.026***
Body buried in sand					0.013
Intercept	0.069***	0.106***	0.092***	0.084*	0.040**
N	19218	5953	24993	5237	19680

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 8: Full regression models and covariates used to develop slope estimates for *Bacteroidales* qPCR $CCE_{\Delta\Delta}$

	Marine	Ten and Under	Over Ten
Body Immersion	-0.085*	-0.134*	-0.053
Bacter. CCE	0.033**	0.051**	0.022
Beach-FB	-0.003	0.019	-0.010
Beach-GB	-0.032**	-0.012	-0.038***
Wave Height	-0.043		
Age (years)	-0.000		
Intercept	0.091***	0.065**	0.081***
N	6052	1286	4766

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

C Comparison with logistic models

C.1 *Enterococcus*

Figure 10: Comparisons of risks estimated by linear, and two logistic models. Swimming-Associated GI illness and Daily Average *Enterococcus* qPCR $CCE_{\Delta\Delta}$.

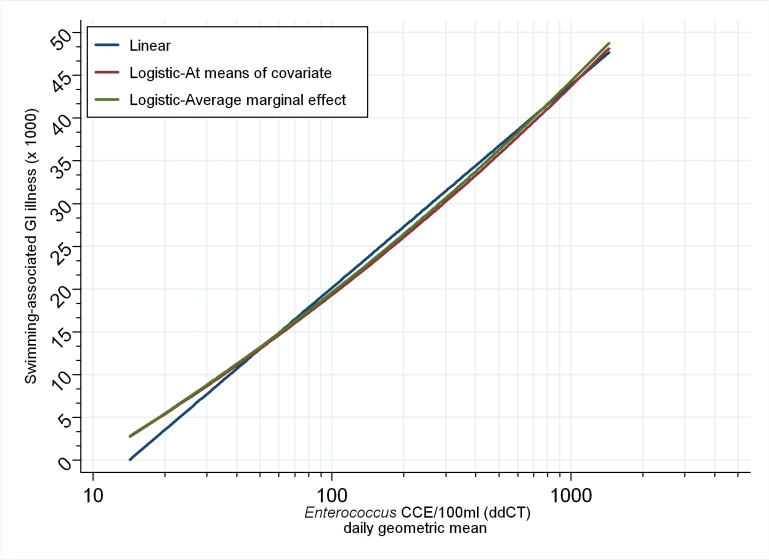


Figure 11: Comparisons of risks estimated by linear, and two logistic models. Children 10 and under. Swimming-Associated GI illness and Daily Average *Enterococcus* qPCR $CCE_{\Delta\Delta}$.