Supplements for "Using sonobuoys and visual surveys to characterize North Atlantic right whale (*Eubalaena glacialis*) calling behavior in the Gulf of St. Lawrence"

Kimberly J. Franklin[1,*], Timothy V. N. Cole[2], Danielle M. Cholewiak[2], Peter A. Duley[2], Leah M. Crowe[3], Philip K. Hamilton[4], Amy R. Knowlton[4], Christopher T. Taggart[1], Hansen D. Johnson[1]

[1] Oceanography Department, Dalhousie University, Halifax, Nova Scotia B3H 4R2 Canada
[2] NOAA Northeast Fisheries Science Center, Woods Hole, MA 02543 USA
[3] Integrated Statistics, under contract to the Northeast Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Woods Hole, MA 02543 USA
[4] Anderson Cabot Center for Ocean Life, New England Aquarium, Boston, MA 02110 USA

**Corresponding author: kimberly.j.franklin@dal.ca*

Supplement 1: Behavioral summary

This supplement includes additional details on how observed behaviors were categorized (Table S1).

Table S1. A comprehensive list of behaviors associated with the three behavioral categories. Mother and calf behaviors were not included in any category. 'Other' is not included in the manuscript but was categorized and investigated in statistical analyses before being removed as it lacked evidence of being correlated with other variables. Behaviors were defined according to Zani and Hamilton (2017).

Foraging	Socializing	'Other'
• Sub-surface feeding • Mouth closing • Skim feeding • Co-feeding • Coordinated feeding	• SAG (surface active group) • Body contact • Rolling • Belly up • Approacher to SAG • Belly to belly contact • Lobtailing • Underwater exhalation • Breaching • Bubbles • Tail slashing • Racing • Focal animal • Flippering/ Flipper Slapping	• Mud • Mouth open • Defecation • Head lift • Avoidance to Approaching Platform • Fluking • Posturing • Calf playing with object • Unusual

Supplement 2: Characterizing potential biases from time, space, entanglement, and the aircraft

Prior to performing statistical model analyses, Kruskal-Wallis comparison tests and graphs (where appropriate) were completed to evaluate potential sources of bias. The variables did not differ significantly among years (except for duration) but many did among months (Table S2). Additional tests were not done due to a low sample size and lack of obvious differences among years. Thus, the years were pooled together in further analyses, however this is not true among the months. To provide perspective on how well the aircraft could survey a sonobuoy acoustic range (30 km) estimated proportions of the aircraft track line within the 30 km acoustic range were calculated. The maximum proportion of area surveyed was estimated to be 0.26 (i.e., 26%) and on average was 0.14 ± 0.06 (i.e., 14% ± 6%) within the acoustic space scale 30 km and assuming the observers saw 1.5 km away from the aircraft (Table S3). There was no evidence that deployments with entangled whales differed from those without entangled whales. The only exception was for 'other' behavior, which were not considered in the main analysis (Table S4). Variation in call rates within each deployment was evaluated to determine the potential effect the aircraft had on calling. The aircraft was around the area the sonobuoy was deployed and would sometimes continue to record after leaving the immediate area. There was no evidence that calling rate changed systematically thorough the deployments or varied significantly between the first and second half of the deployments (Figure S1). The Kruskal-Wallis comparison tests were completed using the function kruskal.test() from the *stats* package in R (R Core Team 2020).

Table S2. Kruskal-Wallis test results comparing variables by year and month. All tests had two degrees of freedom and the double asterisks (**) indicate significance at $\alpha = 0.05$.

	Year		Month	
Variable	Test statistic	p-value	Test statistic	p-value
Upcall count	3.11	0.21	8.99	0.011 **
Upcall rate	5.45	0.07	8.64	0.013 **
Upcall production rate	5.19	0.07	4.88	0.087
Gunshot count	0.96	0.62	21.0	< 0.001 **
Gunshot rate	2.26	0.32	21.6	< 0.001 **
Gunshot production rate	2.16	0.34	20.33	< 0.001 **
Tonal count	2.22	0.33	21.65	< 0.001 **
Tonal rate	3.68	0.16	21.21	< 0.001 **
Tonal production rate	3.27	0.19	19.34	< 0.001 **
Whale count	1.03	0.6	9.51	0.009 **
Male/female ratio	0.35	0.84	3.96	0.14
Adult male	1.60	0.45	11.80	0.003 **
Adult female	1.43	0.49	4.44	0.11
Juvenile male	1.81	0.40	7.60	0.022 **
Juvenile female	3.12	0.21	5.70	0.06
Calf male	2.7	0.26	6.4	0.04 **
Calf female	5.73	0.06	1.43	0.49
Unknown demographic	0.88	0.64	2.84	0.24
Foraging rate	3.98	0.14	5.55	0.06
Socializing rate	4.89	0.09	13.98	0.001 **
‘Other behavior’ rate	3.01	0.22	5.07	0.08
Duration	19.85	< 0.001 **	2.41	0.3

Table S3. Estimated proportion of area surveyed for each deployment. The proportion is the estimated visual area surveyed (1.5 km radius from the aircraft) within 30 km radius of a sonobuoy deployment divided by the acoustic area surveyed.

Deployment date	Deployment time (UTC)	Estimated proportion of area surveyed
2017-06-27	14:27:27	0.12
2017-06-29	18:51:22	0.18
2017-07-05	14:35:00	0.13
2017-07-08	14:54:19	0.14
2017-07-20	15:23:48	0.22
2017-07-21	15:22:44	0.15
2017-07-24	15:18:34	0.16
2017-07-26	14:07:03	0.21
2018-06-06	16:34:45	0.20
2018-06-07	14:06:10	0.21
2018-06-11	19:08:54	0.15
2018-06-15	14:07:01	0.23
2018-06-17	13:32:50	0.13
2018-06-26	17:02:38	0.09
2018-06-29	18:25:40	0.18
2018-07-19	16:30:26	0.17
2018-07-21	14:26:13	0.20
2018-07-27	17:08:00	0.18
2018-08-03	13:46:56	0.18
2018-08-06	12:59:36	0.21
2018-08-12	12:48:16	0.26
2019-06-04	17:05:35	0.08
2019-06-05	13:07:35	0.14
2019-06-07	15:41:42	0.08
2019-06-09	16:08:53	0.13
2019-06-10	15:38:52	0.12
2019-06-13	14:45:41	0.16
2019-07-11	19:14:05	0.06
2019-07-16	17:31:08	0.06
2019-07-19	13:41:55	0.04
2019-07-19	17:59:47	0.14
2019-08-14	16:52:47	0.09
2019-08-15	14:31:22	0.08
2019-08-16	14:10:29	0.03
2019-08-21	11:22:38	0.16
2019-08-25	13:32:57	0.10
2019-08-26	16:04:19	0.05

Table S4. Kruskal-Wallis test results comparing variables for deployments that contained entangled whales (n = 5) and those that did not (n = 32). All tests had one degree of freedom and the double asterisks (**) indicate significance at $\alpha = 0.05$.

Variables	Test Statistic	p-value
Upcall count	2.865	0.091
Upcall rate	2.498	0.114
Upcall production rate	1.61	0.205
Gunshot count	0	1
Gunshot rate	0.005	0.946
Gunshot production rate	0.113	0.736
Tonal count	0.128	0.72
Tonal rate	0.338	0.561
Tonal production rate	0.018	0.893
Whale count	0.714	0.398
Adult male	0.793	0.373
Male female ratio	0.024	0.876
Adult female	1.197	0.274
Juvenile male	0.116	0.733
Juvenile female	0.197	0.657
Calf male	0.321	0.571
Calf female	0.493	0.483
Unknown demographic	0.427	0.514
Foraging rate	0.931	0.335
Socializing rate	0.33	0.565
‘Other behavior’ rate	3.925	0.048 **
Duration	3.318	0.069

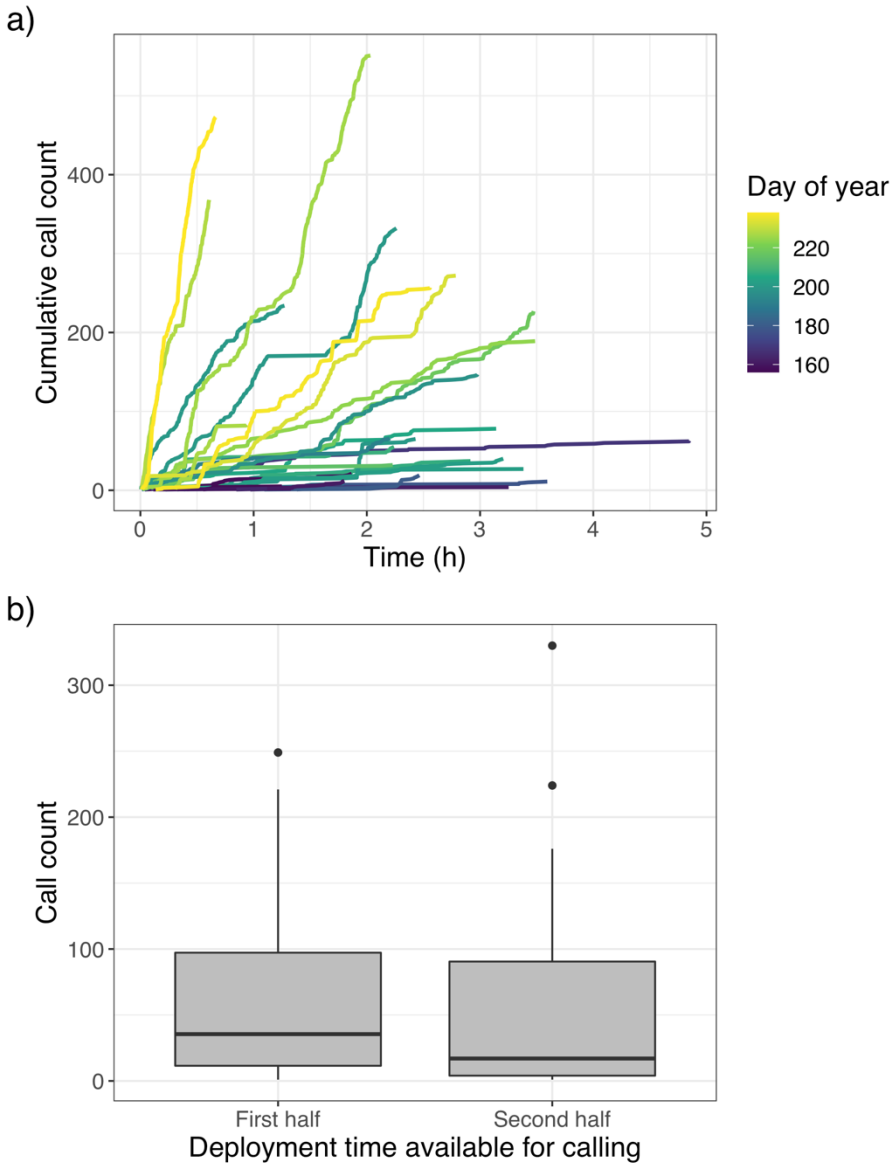

Figure S1. Cumulative call counts over time or all deployments (a), and boxplots of calls recorded during the first half and second half of each deployment (b). There was not a significant difference ($\alpha = 0.05$) between the number of calls produced during the first half of a deployment to the second half (Kruskal-Wallis test statistic = 0.407, p-value = 0.524).

Supplement 3: Evaluating model assumptions

This section describes testing the linear and negative binomial regression model assumptions. To determine if collinearity would be an issue in all the models, a Spearman correlation matrix (Figure S2) and variance inflation factors (VIF; Table S5) were calculated. The correlation matrix only compares two variables at a time while the VIF consider all variables in the model. All the call rates and day of year variables were significantly correlated (Figure S2) and the VIF were relatively high among the call types, particularly gunshot and tonal rates (Table S5). The VIF of the call rate negative binomial regression models were much lower and suggested collinearity was less likely to be an issue.

The Shapiro Wilk normality test was used for each variable and both linear and negative binomial models (Table S6 and Table S7, respectively). The studentized Breusch-Pagan homoscedasticity test was performed on the linear regression models (Table S8). Overdispersion and zero-inflation were performed on the negative binomial models (Table S9 and Table S10, respectively).

Table S6 shows that whale count was normally distributed while call rates were not. Furthermore, every linear regression where a call rate was the dependent variable also failed the normality test (Table S6). None of the whale count linear regressions showed evidence of non-normality (Table S6). A studentized Breusch-Pagan homoscedasticity test was done on linear regression models where whale count was the dependent variable to test for homoscedasticity. The regressions did not violate this assumption (Table S8). Thus, using a linear regression for all models with whale count as the dependent variable appeared to be an appropriate fit for the data.

The models with call rates as the dependent variable violated the normality assumption for linear regressions, therefore another regression model type was needed to fit the data accordingly (Table S7). The data were initially modelled using a generalized linear model (GLM) with a Poisson distribution, but a dispersion test revealed that all the call rates were over dispersed (Table S9). In addition, the call rate data were zero inflated, so a comparison was done between Poisson, quasi-Poisson, and negative binomial models to see which model predicted zeros and how many (Table S10). The negative binomial regressions were most robust to zero inflation and over dispersion so this type of GLM was chosen for the main analysis.

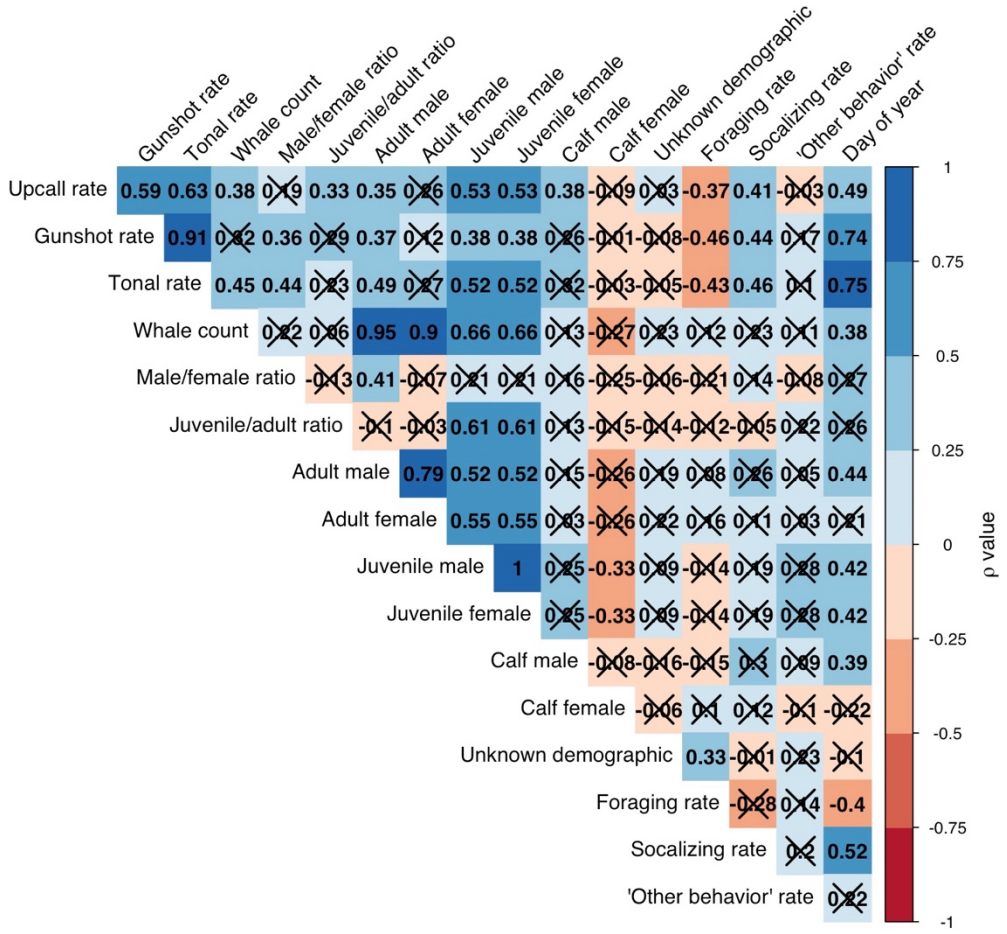

Figure S2. Spearman correlation matrix of all call rates, whale count, male/female ratio, all demographic groups, all behavior rates, and day of year where the ρ value is written in each tile and shown by the colour scale on the right side of the matrix. Correlations that were not significant (at $\alpha = 0.05$) are illustrated with an X in the center of the tile.

Table S5. Variance inflation factors (VIF) for the whale count linear regression model as well as the upcall, gunshot and tonal call rate negative binomial regression models. VIF were calculated using the vif() function from *car* package in R (Fox & Weisberg 2019).

Model	Day of year	Upcall rate	Gunshot rate	Tonal rate	Male/ female ratio	Foraging rate	Socializing rate
Whale count ~ upcall rate + gunshot rate + tonal rate + day of year	1.56	1.92	3.32	3.53	NA	NA	NA
Upcall ~ day of year + male/female ratio + socializing rate + foraging rate + offset(log(duration))	1.97	NA	NA	NA	1.28	1.16	1.75
Gunshot ~ day of year + male/female ratio + socializing rate + foraging rate + offset(log(duration))	1.94	NA	NA	NA	1.29	1.15	1.78
Tonal ~ day of year + male/female ratio + socializing rate + foraging rate + offset(log(duration))	1.93	NA	NA	NA	1.29	1.14	1.8

Table S6. Shapiro Wilk normality test results for individual variables, where the double asterisks (**) indicate significance at $\alpha = 0.05$. These tests were completed using function shapiro.test() from the *stats* package in R (R Core Team 2020).

Variable	Test statistic (W)	p-value
Whale count	0.97	0.34
Upcall rate	0.4	< 0.001 **
Gunshot rate	0.66	< 0.001 **
Tonal rate	0.58	< 0.001 **
Day of year	0.93	0.03 **
Male/female ratio	0.93	0.025 **
Foraging rate	0.71	< 0.001 **
Social rate	0.76	< 0.001 **

Table S7. Shapiro Wilk normality test results for linear regressions for only single variable models where the double asterisks (**) indicate significance at $\alpha = 0.05$. These tests were completed using function shapiro.test() from the *stats* package in R (R Core Team 2020).

Model	Test statistic (W)	p-value
Whale count ~ upcall rate	0.97	0.3
Whale count ~ gunshot rate	0.96	0.24
Whale count ~ tonal rate	0.96	0.28
Whale count ~ day of year	0.95	0.12
Upcall rate ~ day of year	0.59	< 0.001 **
Upcall rate ~ male/female ratio	0.83	< 0.001 **
Upcall rate ~ whale count	0.39	< 0.001 **
Upcall rate ~ foraging rate	0.45	< 0.001 **
Upcall rate ~ socializing rate	0.82	< 0.001 **
Gunshot rate ~ day of year	0.86	< 0.001 **
Gunshot rate ~ male/female ratio	0.79	< 0.001 **
Gunshot rate ~ whale count	0.7	< 0.001 **
Gunshot rate ~ foraging rate	0.76	< 0.001 **
Gunshot rate ~ socializing rate	0.91	0.004 **
Tonal rate ~ day of year	0.83	< 0.001 **
Tonal rate ~ male/female ratio	0.79	< 0.001 **
Tonal rate ~ whale count	0.6	< 0.001 **
Tonal rate ~ foraging rate	0.68	< 0.001 **
Tonal rate ~ socializing rate	0.9	0.002 **

Table S8. Studentized Breusch-Pagan test results to test for homoscedasticity in models containing whale count as the dependent variable. All tests had one degree of freedom and $\alpha = 0.05$. The tests were completed using function bptest() from *lmtest* package in R (Zeileis & Hothorn 2002).

Model	Test statistic (bp)	p-value
Whale count ~ upcall rate	0.05	0.83
Whale count ~ gunshot rate	1.44	0.23
Whale count ~ tonal rate	2.62	0.11
Whale count ~ day of year	0.17	0.68

Table S9. Overdispersion test results for full Poisson GLM call rate models. All tests had one degree of freedom and the double asterisks (**) indicate significance at $\alpha = 0.05$. These tests were completed using function dispersontest() from the *AER* package in R (Kleiber & Zeileis 2008).

Model	Test statistic (z)	p-value	Dispersion
Upcall ~ day of year + whale count + male/female ratio + social rate + foraging rate + offset(log(duration))	3.25	0.001 **	21.67
Gunshot ~ day of year + whale count + male/female ratio + social rate + foraging rate + offset(log(duration))	2.13	0.017 **	70.78
Tonal ~ day of year + whale count + male/female ratio + social rate + foraging rate + offset(log(duration))	2.52	0.006 **	123.41

Table S10. Zero inflation test results for full call rate models (for all call types) with Poisson, quasi-Poisson and negative binomial distributions. Tests were completed using check_zeroinflation() function from *performance* package in R (Lüdecke et al. 2020).

Model type	Model	No. of zeros in the model estimates	No. of zeros in the data	Ratio
Poisson	Upcall ~ day of year + whale count + male/female ratio + socializing rate + foraging rate + offset(log(duration))	0	6	0
Quasi-Poisson	Upcall ~ day of year + whale count + male/female ratio+ socializing rate + foraging rate + offset(log(duration))	0	6	0
Negative binomial	Upcall ~ day of year + whale count + male/female ratio+ socializing rate + foraging rate + offset(log(duration))	4	6	0.67
Poisson	Gunshot ~ day of year + whale count + male/female ratio+ socializing rate + foraging rate + offset(log(duration))	0	10	0
Quasi-Poisson	Gunshot ~ day of year + whale count + male/female ratio+ socializing rate + foraging rate + offset(log(duration))	0	10	0
Negative binomial	Gunshot ~ day of year + whale count + male/female ratio + socializing rate + foraging rate + offset(log(duration))	8	10	0.8
Poisson	Tonal ~ day of year + whale count + male/female ratio + socializing rate + foraging rate + offset(log(duration))	0	9	0
Quasi-Poisson	Tonal ~ day of year + whale count + male/female ratio + socializing rate + foraging rate + offset(log(duration))	0	9	0
Negative binomial	Tonal ~ day of year + whale count + male/female ratio + socializing rate + foraging rate + offset(log(duration))	6	9	0.67

Supplement 4: Model selection and comparison

This supplementary section contains additional details on the selection and comparison of models used to characterize call rates and whale count models. Negative binomial likelihood ratio comparison tests were used for comparing call rates for all call types (upcall, gunshot, and tonal; with variables day of year, male/female ratio, whale count, foraging rate, and socializing rate) to null models which only contained the intercept and offset (Table S11). All predictor variables (day of year, male/female ratio, whale count, foraging rate, and socializing rate) were combined into a single negative binomial model for each call type (upcall, gunshot, and tonal) and an AIC stepwise selection procedure was used to identify the most parsimonious model (Table S12). The models produced from the AIC stepwise procedure were then compared to null models (contained only the intercept and offset) using the negative binomial likelihood ratio test (Table S13). Lastly, ANOVA (analysis of variance) tests were performed for the linear regression models used to determine if whale count (with variables upcall counts, gunshot counts, tonal counts, and day of year) differed from null models (Table S14).

Table S11. Likelihood ratio test results for each call type for each variable of interest. All null models had residual degrees of freedom (df) of 36 and all single variable models had a residual df of 35. The degrees of freedom were one for each test. The models were compared to a null model which contained only the intercept and offset terms for each dependent variable. The double asterisks (**) indicate significance at $\alpha = 0.05$.

Model	Upcall		Gunshot		Tonal	
	likelihood ratio statistic	p-value	likelihood ratio statistic	p-value	likelihood ratio statistic	p-value
Day of year + offset(log(duration))	17.56	< 0.001 **	18.46	< 0.001 **	23.33	< 0.001 **
Male/female ratio + offset(log(duration))	9.81	0.002 **	2.14	0.144	5.46	0.019 **
Whale count + offset(log(duration))	0.32	0.57	0.88	0.348	1.15	0.28
Foraging rate + offset(log(duration))	6.22	0.013 **	8.95	0.003 **	8.64	0.003 **
Socializing rate + offset(log(duration))	22.23	< 0.001 **	6.32	0.012 **	7.99	0.005 **

Table S12. Backward AIC stepwise procedure of negative binomial GLM model for each call rate type (upcall, gunshot, tonal) where the first model is the full model, and the last model is the most parsimonious model for estimating the corresponding call rate. Null model degrees of freedom for all steps were 36.

Model	Log link	AIC	Deviance	Residual df
Upcall				
Upcall ~ day of year + male/female ratio + foraging rate + socializing rate + whale count + offset(log(duration))	-139.9	293.79	43.81	31
Upcall ~ male/female ratio + foraging rate + socializing rate + whale count + offset(log(duration))	-139.39	290.79	43.68	32
Upcall ~ foraging rate + socializing rate + whale count + offset(log(duration))	-139.4	288.8	43.69	33
Upcall ~ socializing rate + whale count + offset(log(duration))	-140.06	288.12	43.81	34
Gunshot				
Gunshot ~ day of year + male/female ratio + foraging rate + socializing rate + whale count + offset(log(duration))	-153.47	320.94	42.62	31
Gunshot ~ day of year + foraging rate + socializing rate + whale count + offset(log(duration))	-153.11	318.23	42.26	32
Gunshot ~ day of year + foraging rate + socializing rate + offset(log(duration))	-153.31	316.62	42.35	33
Gunshot ~ day of year + foraging rate + offset(log(duration))	-153.41	314.82	42.32	34
Gunshot ~ day of year + offset(log(duration))	-154.25	314.5	41.86	35
Tonal				
Tonal ~ day of year + male/female ratio + foraging rate + socializing rate + whale count + offset(log(duration))	-167.24	348.48	42.76	31
Tonal ~ day of year + male/female ratio + socializing rate + whale count + offset(log(duration))	-166.41	344.83	42.73	32
Tonal ~ day of year + male/female ratio + whale count + offset(log(duration))	-166.5	343	42.66	33
Tonal ~ day of year + male/female ratio + offset(log(duration))	-166.83	341.67	42.8	34

Table S13. Likelihood ratio test results for parsimonious models produced from stepwise procedure for each call type compared to a null model. The null model contains the intercept and offset term for each call type. The double asterisks (**) indicate significance $\alpha = 0.05$.

	Model	Residual df	df	likelihood ratio statistic	p-value
Upcall					
	Upcall ~ offset(log(duration))	36			
	Upcall ~ socializing rate + whale count + offset(log(duration))	34	2	25.94	< 0.001 **
Gunshot					
	Gunshot ~ offset(log(duration))	36			
	Gunshot ~ day of year + offset(log(duration))	35	1	18.46	< 0.001 **
Tonal					
	Tonal ~ offset(log(duration))	36			
	Tonal ~ day of year + male/female ratio + offset(log(duration))	34	2	26.67	< 0.001 **

Table S14. ANOVA tables for linear regression models where the response variable is whale count and the predictor variables upcall rate, gunshot rate, tonal rate, and day of year, respectively. The double asterisks (**) indicate significance at $\alpha = 0.05$.

	Degrees of freedom	Sums Squared	Mean Squared	F value	p-value
Upcall rate	1	13.09	13.09	0.09	0.76
Residuals	35	4989.24	142.55		
Gunshot rate	1	192.98	192.98	1.4	0.24
Residuals	35	4809.34	137.41		
Tonal rate	1	197.35	197.35	1.44	0.24
Residuals	35	4804.97	137.28		
Day of year	1	850.29	850.29	7.17	0.011 **
Residuals	35	4152.04	118.63		

Supplement 5: Model results without entangled whales

This section contains a table similar to Table 3 in the main paper, where all statistical models were repeated with the deployments that did not contain entangled whales (n = 32). The results in Table S15 are almost identical to the model results containing deployments with entangled whales.

Table S15. All models and associated p-values for all deployments without entangled whales for a) single variable negative binomial generalized linear models (GLMs) used to characterize call rates; b) stepwise-selected negative binomial GLMs used to characterize call rates; and c) linear models used to characterize whale count. The p-values for a) and b) were derived from likelihood ratio tests for negative binomial regressions, while those for c) were derived from ANOVA tables. Models a) and b) used an offset term (log of deployment duration) that is not shown here. The double asterisks (**) indicate significance at $\alpha = 0.05$.

Model	p-value
a) Characterizing call rates	
upcall = (0.033 ± 0.008) * **day of year** + (-12.247 ± 1.649)	< 0.001 **
upcall = (0.852 ± 0.401) * **male/female ratio** + (-7.243 ± 0.742)	0.003**
upcall = (0.011 ± 0.024) * **whale count** + (-5.71 ± 0.632)	0.709
upcall = (-3.861 ± 1.444) * **foraging rate** + (-5.157 ± 0.317)	0.02**
upcall = (5.363 ± 1.191) * **socializing rate** + (-6.762 ± 0.253)	< 0.001 **
gunshot = (0.061 ± 0.011) * **day of year** + (-17.445 ± 2.111)	< 0.001 **
gunshot = (0.537 ± 0.571) * **male/female ratio** + (-5.865 ± 1.054)	0.221
gunshot = (0.029 ± 0.03) * **day of year** + (-5.58 ± 0.808)	0.357
gunshot = (-6.782 ± 2.014) * **foraging rate** + (-4.408 ± 0.403)	0.007**
gunshot = (4.735 ± 1.954) * **socializing rate** + (-5.804 ± 0.409)	0.011**
tonal = (0.061 ± 0.009) * **day of year** + (-17.221 ± 1.798)	< 0.001 **
tonal = (1.146 ± 0.531) * **male/female ratio** + (-6.522 ± 0.981)	0.044**
tonal = (0.031 ± 0.029) * **whale count** + (-5.102 ± 0.775)	0.291
tonal = (-6.574 ± 1.824) * **foraging rate** + (-3.904 ± 0.383)	0.005**
tonal = (4.876 ± 1.836) * **socializing rate** + (-5.367 ± 0.384)	0.003**
b) Characterizing call rates - stepwise regressions	
upcall = (5.44 ± 1.12) * **socializing rate** + (0.03 ± 0.02) * **whale count** + (-7.65 ± 0.45)	< 0.001 **
gunshot = (0.06 ± 0.01) * **day of year** + (-17.45 ± 2.11)	< 0.001 **
tonal = (0.06 ± 0.01) * **day of year** + (-17.22 ± 1.8)	< 0.001 **
c) *Characterizing whale count*	
whale count = (0.05 ± 0.04) * **gunshot rate** + (22.27 ± 2.46)	0.25
whale count = (0.03 ± 0.02) * **tonal rate** + (22.47 ± 2.38)	0.26
whale count = (0.01 ± 0.06) * **upcall rate** + (23.56 ± 2.35)	0.84
whale count = (0.21 ± 0.07) * **day of year** + (-16.5 ± 13.58)	0.005**

Literature Cited

Fox J, Weisberg S (2019) An {R} companion to applied regression. Third Edition. Sage, Thousand Oaks CA. Available from https://socialsciences.mcmaster.ca/jfox/Books/Companion/

Kleiber C, Zeileis A (2008) Applied econometrics with R. Springer-Verlag, New York. Available from https://CRAN.R-project.org/package=AER

Lüdecke, Makowski, Waggoner, Patil (2020) Assessment of Regression Models Performance. CRAN. Available from https://easystats.github.io/performance/

R Core Team (2020) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Zani MA and Hamilton PK (2017) North Atlantic Right Whale Consortium Photographic Database/Catalog Submission. 74 pp. https://www.narwc.org/identification-database.html

Zeileis A, Hothorn T (2002) Diagnostic checking in regression relationships. R News 2:7-10. Available from https://CRAN.R-project.org/doc/Rnews/