

Supporting Information

Evaluating Elicited Judgements of Turtle Captures for Data-poor Fisheries Management

William N. S. Arlidge [0000-0002-1807-4150]^{1,2*}, Joanna Alfaro-Shigueto^{3,4,5}, Bruno Ibañez-Erquiaga⁶, Jeffrey C. Mangel^{3,4}, Dale Squires⁷, E.J. Milner-Gulland¹

¹ Interdisciplinary Centre for Conservation Science, Department of Zoology, University of Oxford OX1 3PS, UK

² Pembroke College, University of Oxford, Oxford OX1 1DW, UK

³ Pro Delphinus, Calle José Galvez 780-E, Miraflores 15074, Perú

⁴ School of Biosciences, University of Exeter, Cornwall Campus, Penryn, Cornwall, TR10 9FE, UK

⁵ Facultad de Biología Marina, Universidad Científica del Sur, Campus Villa, Lima 42, Perú

⁶ Departamento de Química y Biología, Universidad San Ignacio de Loyola, Lima, Perú

⁷ NOAA Fisheries Southwest Fisheries Science Center, La Jolla, CA, USA

* Corresponding author

Appendix S1. Demographic data

Table S1. Expert elicitation respondent characteristics.

Variable	Category	Number
Stakeholder group	Gillnet skipper	3
	Not-for-profit scientist	2
Gender	Male	4
	Female	1
Age	26–40	3
	41–55	2
Years lived in San Jose	0–10	2
	11–20	0
	21–30	1
	31–40	2
Years fishing	0–10	2
	11–20	3

Appendix S2. Tables of summary statistics - green turtle observer analysis

Table S2.1. Spearman's rho (rs) rank correlation test results for variable inclusion in binomial Generalized Linear Mixed Model (GLMM). GRT = gross registered tonnage, p = p-value. Correlation coefficients are highlighted in bold, p-values in plain text. Soak time is the length of time the net is in the water during each set (setting and hauling of the net). Fishers will often lay multiple sets during a fishing trip. To support interpretation, the correlation coefficient of GRT and season was 0.058, with a p-value of 0.005. The low level of the p-value indicated that 99.995% of the time the correlation is weak at an r of 0.058.

	GRT	Season	Year	Soak	Net length (km)	Crew number
GRT	1 (p<0.01)	0.058 (p=0.005)	-0.180 (p<0.001)	-0.022 (p=-0.296)	-0.132 (p<0.01)	-0.221 (p=0.286)
Season		1 (p<0.01)	0.01 (p=0.596)	-0.061 (p=0.002)	-0.081 (p<0.001)	-0.027 (p=0.173)
Year			1 (p<0.01)	-0.038 (p=0.054)	0.228 (p<0.01)	-0.341 (p<0.01)
Soak time				1 (p<0.01)	0.002 (p=0.914)	0.074 (p<0.001)
Net length (km)					1 (p<0.01)	0.062 (p=0.001)
Crew number						1 (p<0.01)

Table S2.2. Hausman test data where the null hypothesis is that the preferred model is random effects versus the alternative hypothesis that the preferred model is fixed effects. df = degrees of freedom.

Data: Green turtle bycatch (binomial) ~ GRT + Season + Year + Soak time + Net (km) + Crew. Panel data index = vessel identification.		
Chi squared = 6.6852	df = 6	p-value = 0.3509
alternative hypothesis: one model is inconsistent.		

Failing to reject the null hypothesis of random effects (against fixed effects) we proceeded with a Generalized Linear Mixed Model (GLMM) to integrate both fixed and random effect variables into our model.

Table S2.3. Results from the binomial generalized linear mixed model for predicting probability of green turtle catch, where vessel ID is a random effect (re). Models are ranked by Delta AIC scores, with Delta BIC scores also presented. df = degrees of freedom, GRT = gross registered tonnage. The chosen model is in bold text.

<i>Green turtle catch</i>				
Rank	Model	df	Δ AIC	Δ BIC
1	GRT + Year + Season + Vessel (re)	2501	0	0
2	GRT + Year + Season + Crew + Vessel (re)	2500	1.4	7.2
3	GRT + Year + Season + Soak time + Vessel (re)	2500	1.7	7.6
4	GRT + Year + Season + Soak time + Crew + Vessel (re)	2499	3.1	14.8
5	GRT + Year + Season + Net length (km) + Crew + Vessel (re)	2499	3.4	15.0
6	GRT + Year + Season + Soak time + Net length (km) + Vessel (re)	2499	3.7	15.4
7	GRT + Year + Season + Soak time + Net length (km) + Crew + Vessel (re)	2498	5.1	22.6
8	GRT + Year + Vessel (re)	2502	78.9	67.2
9	GRT + Year + Crew + Vessel (re)	2501	80.6	74.7
10	GRT + Year + Soak time + Net length (km) + Crew + Vessel (re)	2499	84.4	90.3
11	GRT + Season + Net length (km) + Crew + Vessel (re)	2501	145.0	139.2
12	GRT + Season + Soak time + Net length (km) + Crew + Vessel (re)	2500	146.8	146.8
13	GRT + Vessel (re)	2504	150.8	127.5
14	GRT + Crew + Vessel (re)	2503	152.7	135.2
15	GRT + Net length (km) + Crew + Vessel (re)	2502	153.6	141.9
16	GRT + Soak time + Net length (km) + Crew + Vessel (re)	2501	155.5	149.6

Table S2.3. Summary of observer coverage across the inshore/midwater fleet by vessel size class. GRT = gross registered tonnage.

GRT class	Number of vessels	Number of trips	Number of sets
1<4	20	53	291
4<8	4	181	1099
8<12	7	208	1278
>12	1	3	17
Total	32	445	2685

Table S2.4. Extrapolated seasonal and annual green turtle catch estimates calculated from observer data without Generalized Linear Mixed Model (GLMM) class weightings. Temp. Grp. = temporal grouping; winter represents the cold weather months of June to November, summer represents the warm weather months of December to May.

Temp. Grp.	Mean	Min 90 CI	Max 90 CI
Winter	340.37	255.27	428.99
Summer	797.47	598.08	1005.10
Total net encounters p.a.	1137.85	853.35	1434.09

Table S2.5. Green turtle catch-per-unit-effort (CPUE) per trip weighted by vessel size class (gross registered tonnage). Gross registered tonnage for weightings were obtained from the Generalized Linear Mixed Model (GLMM). SE = standard error of the mean, CI = confidence interval.

GRT class	Coef.	SE	Green turtle weighted mean	Green turtle (-90 CI)	Green turtle (+90 CI)
1<4	0	0.000	0.71	0.56	0.86
4<8	0.036	0.030	0.75	0.60	0.75
8<12	0.050	0.026	0.76	0.61	0.76
>12	0.104	0.083	0.82	0.67	0.82

Table S2.6. Approximated inshore/midwater gillnet fleet size in San Jose. Actively fishing gillnet vessels in 2008 are based on expert opinion from researcher's surveying in San Jose that year (JAS & JCM). The fleet size was most recently recorded in a census survey in the winter of 2017 and key informant interviews provided estimates of the 2017 summer fleet size (Arlidge et al. in review). Seasonal differences reflect the proportional difference identified from data the 2017 census data, which is supported by key informant interviews and focus discussion groups held in San Jose. An incremental decay of three vessels per year were applied from 2007 to 2019. Gillnet fleet size is declining as skippers and crew change from gillnets to squid jigging. The winter fishing season is June-November and the summer fishing season is December-May.

Year	Vessel Number (Summer)	Vessel Number (Winter)
2007	63	48
2008	60	45
2009	57	42
2010	54	39
2011	51	36
2012	48	33
2013	45	30
2014	42	27
2015	39	24
2016	36	21
2017	33	18
2018	30	15

Table S2.7. Extrapolated green turtle capture estimates per season based on catch-per-unit-effort (CPUE) per trip. CPUE was weighted using Generalized Linear Mixed Model (GLMM) size class coefficients. Annual values were summed across gross registered tonnage weight classes. CI = confidence interval.

Year	Green turtle CPUE / Summer	Green summer (-90 CI)	Green summer (+90 CI)	Green turtle CPUE / Winter	Green winter (-90 CI)	Green winter (+90 CI)
2007	1199.54	953.15	1353.53	597.57	474.83	674.29
2008	1142.42	907.76	1289.08	560.22	445.15	632.14
2009	1085.29	862.37	1224.62	522.87	415.47	590.00
2010	NA	NA	NA	NA	NA	NA
2011	NA	NA	NA	NA	NA	NA
2012	NA	NA	NA	NA	NA	NA
2013	856.81	680.82	966.81	373.48	296.77	421.43
2014	799.69	635.43	902.35	336.13	267.09	379.29
2015	742.57	590.04	837.90	298.79	237.41	337.14
2016	685.45	544.65	773.45	261.44	207.74	295.00
2017	628.33	499.27	708.99	224.09	178.06	252.86
2018	571.21	453.88	644.54	186.74	148.38	210.71
2019	514.09	408.49	580.08	149.39	118.71	168.57
Mean	822.54	653.59	928.13	351.07	278.96	396.14

Bootstrap comparison of means methodology

To compare the means of our two datasets (expert elicited bycatch estimates and bycatch rates calculated from at-sea observer data) we ran bootstrap hypothesis tests (Efron & Tibshirani 1993). We tested the null hypothesis that, within each fishing season, the mean monthly number of green turtle captures in the San Jose inshore/midwater gillnet fleet calculated from the elicitation exercise is the same as the capture rate calculated from the observer data. This required comparing monthly elicited estimates of green turtle bycatch rates within summer (n=5) and winter (n=5) fishing seasons, with monthly capture rates within summer and winter across each fishing year we had observer data for (n=10 summer, n=10 winter).

Our monthly green turtle capture rates per season from the elicitation data represents the expected value based on multiple years of data, which may or may not be the same length (i.e., one expert may be drawing on 20 years of fishing experience when considering their monthly green turtle bycatch rates, whereas another may be drawing on 5 years fishing experience). Green turtle captures were estimated from the observer data for each year data was available (n=10 over a 13-year period) and then averaged across this period.

Consideration of the potential source of temporal bias between the two datasets is highlighted for results interpretation.

Appendix S3. Performance metrics calculations

We followed the performance metric calculations for participants of the IDEA protocol as outlined by Hemming et al. (2018b) and McBride et al. (2012) .

Accuracy

Accuracy of point estimates (‘Accuracy’) was measured by calculating the average log-ratio error (ALRE) for participants' judgements. In order to calculate ALRE, we first standardized each response by the range of responses for that question, known as range-coding (McBride et al. 2012; Hemming et al. 2018). Range-coding is used to minimize the effect that one or a few very divergent responses has on the accuracy measure (Burgman et al. 2011). The calculation involves standardising the best estimates $b_e^{n,r}$ from each participant e , for each question n , in each round r (including the realized outcome) by the range of responses for each question:

$$bc_e^{n,r} = \frac{(b_e^{n,r} - b_{min}^n)}{b_{max}^n - b_{min}^n} \quad (S1)$$

where, $bc_e^{n,r}$ is the range-coded response for participant e , in round r , b_{max}^n is the maximum best estimate response taken from the pool of responses (best estimates) from all participants for question n , across both Round 1 and Round 2, and b_{min}^n is the minimum best estimate response. The realized truth (x^n) for each question is also range-coded using equation 3.

ALRE is then calculated using the range-coded values generated:

$$ALRE_i = \frac{1}{N^r} \sum_{n=1}^N \left(\log_{10} \left(\frac{x^{n+1}}{bc_e^{n,r} + 1} \right) \right) \quad (S2)$$

where, N_r is the number of quantities assessed in any round r , $bc_e^{n,r}$ is the range-coded prediction, and x^n is the range-coded observed value (‘realized truth’) for question n . Ones were added to both the range-coded observed value and the range-coded prediction to avoid taking the log of zero (which occurs when the realisation is standardized). The \log_{10} ratio provides a measure that emphasizes order of magnitude errors rather than linear errors. In other words, a judgement that is five times the observed value x would be weighted the same as a value that is one-fifth the value of the observed value x . Smaller ALRE scores indicate more accurate responses (Hemming et al. 2018). For any given question, the log ratio scores have a maximum possible range of $\log_{10}(2)(0.31)$, which occurs when the true answer

coincides with either the group minimum or group maximum, and a best possible score of zero (McBride et al. 2012).

Calibration

Calibration of interval judgement ('Calibration') measures the proportion of questions answered by a respondent for which their intervals capture the realized truth. Following the protocol outlined in Hemming et al. (2018b), we used the standardized upper and lower values of participants' intervals and the standardised level of confidence associated with those intervals. Using participants' standardised intervals to score calibration is possible as the participants receive feedback on their standardisations between Round 1 and Round 2 of the modified Delphi approach and are informed they can (and should) adjust their estimates if they are not in accordance with their true beliefs. We standardized intervals to 90%, as such a participant with perfect calibration is assumed to capture the observed value 90% of the time. The actual number of realisations captured was calculated using:

$$C_e^r = \frac{t^r}{N^r} \times 100 \quad (\text{S3})$$

where, C_e^r is the score for calibration for participant e in Round r , t is the number of standardized intervals provided by the participant which contained the realized truth, and N_r is the total number of questions answered by the participant in round r . Because it is possible for participants to obtain a high calibration by providing very wide (uninformative) intervals, this calibration measure is considered alongside a measure of informativeness (described below).

Informativeness

Informativeness of interval judgement ('Informativeness') measures the width (i.e., precision) of the of the participant's intervals relative to the total range provided by participants for a question. First, we calculated the width of standardized intervals (e.g. 90%) supplied by participants for each question in each round:

$$w_e^{n,r} = u_e^{n,r} - l_e^{n,r} \quad (\text{S4})$$

where, $w_e^{n,r}$ is the width of the standardized interval of participant e for question n , in round r , while $u_e^{n,r}$ is the upper standardized estimate provided by participant e for question n , in

Round r , and $l_e^{n,r}$ is the lower standardized estimate provided by participant e for question n .

Then for each question, a background range was calculated:

$$w_{max}^n = u_{max}^n - l_{min}^n \quad (S5)$$

where w_{max}^n is the background range created for question n , u_{max}^n is the highest standardised upper bound estimate provided for question n across Round 1 and Round 2 by any participant, and l_{min}^n is the lowest standardized lower bound estimate provided for question n across Round 1 and Round 2 by any participant.

Finally, the average informativeness score of each participant per round was calculated by:

$$I_e^r = \frac{1}{N^r} \sum_{n=1}^N \left(\frac{w_e^{n,r}}{w_{max}^n} \right) \quad (S6)$$

where, I_e^r is the average informativeness of participant e in Round r over all questions in Round r , $w_e^{n,r}$ is the width of the interval provided by participant e in Round r for question n , w_{max}^n is the background range for question n , and N^r is the total number of questions answered in Round r . Scores range between 0 and 1 with higher numbers relating to less informative individuals. To ensure that participants are not rewarded for reporting no uncertainty when they are not certain of the true value of observed value x , this measure is considered in conjunction with the calibration measure.

Appendix S4. Additional leatherback turtle analysis

Elicited judgements for leatherback turtle capture rate with gillnets

In addition to eliciting participants' judgements for capture rates of green turtles, we also asked participants to quantify capture rates for leatherback turtles (*Dermochelys coriacea*) in gillnets set by inshore/midwater vessels operating from San Jose, Lambayeque, Peru (6°46' S, 79°58' W).

Participants' judgements for leatherback turtle captures were 4.8 – 15.2 individuals per month (Fig. S4.3). Skippers' judgements (best estimates) were higher for leatherback turtles than the not-for-profit employees. We then used participants' monthly turtle capture rates to infer seasonal capture rates, as well as calculating an annual capture rate for leatherback turtles by adding the summer and winter encounters together.

Comparison of participant judgements with onboard observer data

Following the analysis undertaken for green turtles presented in the main text, GLMMs were used to estimate the predictive power of vessel weight class for leatherback turtle catch while controlling for seasonal and annual temporal variations, fishing effort (gillnet soak time), and inter-vessel variation within the fleet as a random effect. GRT and a random effect for vessel resulted in the best model (Tables S4.1 and S4.2).

In contrast to green turtles, vessel size was found to be weakly negatively correlated to leatherback turtle catch, however, following correcting for serial correlation, no significance was identified (Table S4.2). The low and sporadic catch rate of leatherback turtles across the observer dataset (n=7) resulted in the GLMM model having little predictive power in terms of leatherback catchability in relation to our vessel size classes. We caution readers when interpreting the presented outputs in this analysis presented in supporting information. Based on this observer data and GLMM output, we extrapolated gillnet leatherback turtle capture estimates for the inshore/midwater gillnet fleet in San Jose as an estimated 19.18 (5 – 32) individuals per year (Table S4.3).

Assessing participant performance

Participants' judgements were more precise at estimating catch rates for leatherback turtles than green turtles. Leatherback turtles are infrequently captured in this fishery (Table 1 – Main text); it is possible that these fishers were able to recall these rare capture events with

more precision than for green turtles, which are more frequently captured, due to the lasting impression that encountering this species leaves. Leatherback turtles are more easily differentiated in size and by their distinct soft leather-like shell from the other hard-shelled sea turtles that are captured in the San Jose fishing system (green, hawksbill, and olive ridley turtles; Alfaro-Shigueto et al. 2011; Alfaro-Shigueto et al. 2018). Indeed, good recollection of rare capture events is reflected in the findings of other studies eliciting local knowledge for species counts (van der Hoeven et al. 2004; Brittain et al. 2018). Participant L05 (not-for-profit) submitted accurate leatherback turtle judgements despite very tight confidence bounds. Participant L01 (gillnet skipper) accurately estimated leatherback turtle captures in winter.

Table S4.1. Results from the binomial generalized linear mixed model for predicting probability of leatherback turtle catch, where vessel ID is a random effect (re). Models are ranked by Delta AIC scores, with Delta BIC scores also presented. df = degrees of freedom, GRT = gross registered tonnage. The chosen model in bold text.

<i>Leatherback turtle catch</i>				
Rank	Model	df	Delta AIC	Delta BIC
1	GRT + Vessel (re)	2504	0	0
2	GRT + Crew + Vessel (re)	2503	1.528	7.367
3	GRT + Year + Vessel (re)	2502	2.777	14.454
4	GRT + Net length (km) + Crew + Vessel (re)	2502	3.247	14.924
5	GRT + Year + Season + Vessel (re)	2501	3.628	21.143
6	GRT + Season + Net length (km) + Crew + Vessel (re)	2501	3.882	21.397
7	GRT + Year + Crew + Vessel (re)	2501	4.408	21.924
8	GRT + Year + Season + Crew + Vessel (re)	2500	5.179	28.533
9	GRT + Soak time + Net length (km) + Crew + Vessel (re)	2501	5.24	22.755
10	GRT + Year + Season + Soak time + Vessel (re)	2500	5.614	28.968
11	GRT + Season + Soak time + Net length (km) + Crew + Vessel (re)	2500	5.867	29.221
12	GRT + Year + Season + Net length (km) + Crew + Vessel (re)	2499	6.894	36.086
13	GRT + Year + Season + Soak time + Crew + Vessel (re)	2499	7.162	36.354
14	GRT + Year + Season + Soak time + Net length (km) + Vessel (re)	2499	7.309	36.501
15	GRT + Year + Soak time + Net length (km) + Crew + Vessel (re)	2499	8.209	37.401
16	GRT + Year + Season + Soak time + Net length (km) + Crew + Vessel (re)	2498	8.877	43.908

Table S4.2. Best fit model for predicting probability of leatherback turtle catch chosen following AIC and BIC ranking criteria.

<i>Leatherback turtle bycatch</i>				
Reference	Random effects	Intercept	Residual	n
Vessel	Std. dev	1.05E-02	4.85E-02	32
	Fixed effects	Coefficient	SE ¹	p-value
	Intercept	0.006	0.007	0.41
GRT	4<8 GRT	-0.009	0.008	0.2732
<i>reference = 0<4 GRT</i>	8<12 GRT	-0.008	0.006	0.2281
	>12 GRT	-0.010	0.016	0.5314
	Net (km)	0.001	0.002	0.5962
	Crew number	0.001	0.001	0.5032

¹ Serial correlation-consistent standard errors

Table S4.3. Extrapolated seasonal and annual reductions in leatherback turtle captures with small-scale fishery gillnets set from vessels launching from San, Jose, Peru, based on elicited monthly estimates of the efficacy of the bycatch reduction strategy scenario of gear switching from gillnets to potting or trolling, and onboard observer data obtained from the period August 2007–May 2019. Temp. Grp. = temporal grouping; winter represents the cold weather months of June to November, summer represents the warm weather months of December to May. CI = credible interval. Note that no weighting by GLMM coefficients were applied to the extrapolated bycatch rates calculated from the observer data.

Temp. Grp.	Expert elicitation data			Observer data		
	Mean best (B)	Std. lower 90 CI (lsi)	Std. upper 90 CI (usi)	Mean	Min 90 CI	Max 90 CI
Monthly/winter	7.17	4.76	9.48	2.24	0.59	4.00
Monthly/summer	12.03	9.65	15.16	3.20	0.84	5.70
Total winter	43	28.57	56.89	5.74	1.52	10.23
Total summer	72.2	57.92	90.95	13.44	3.55	23.98
Annual	115.2	86.49	147.84	19.18	5.07	34.21

Table S4.4. Leatherback turtle catch-per-unit-effort (CPUE) per trip weighted by vessel size class (gross registered tonnage). Gross registered tonnage for weightings were obtained from the Generalized Linear Mixed Model (GLMM). SE = standard error of the mean, CI = confidence interval.

GRT class	Coef.	SE	Leatherback turtle weighted mean	Leatherback turtle (-90 CI)	Leatherback turtle (+90 CI)
1<4	0.000	0.000	0.015	0.01	0.02
4<8	-0.009	0.008	0.006	0.00	0.02
8<12	-0.008	0.006	0.007	0.00	0.02
>12	-0.010	0.016	0.005	0.00	0.01

Table S4.5. Extrapolated seasonal and annual leatherback turtle catch estimates calculated from observer data without Generalized Linear Mixed Model (GLMM) class weightings. Temp. Grp. = temporal grouping; winter represents the cold weather months of June to November, summer represents the warm weather months of December to May.

Temp. Grp.	Mean	Min 90 CI	Max 90 CI
Winter	7.28	2.78	12.07
Summer	17.06	6.52	28.29
Total net encounters p.a.	24.33	9.30	40.36

Table S4.5. Extrapolated leatherback turtle capture estimates per season based on catch-per-unit-effort (CPUE) per trip. CPUE was weighted using Generalized Linear Mixed Model (GLMM) size class coefficients. Annual values were summed across gross registered tonnage weight classes. CI = confidence interval.

Year	Leatherback turtle CPUE / Summer	Leatherback summer (-90 CI)	Leatherback summer (+90 CI)	Leatherback turtle CPUE / Winter	Leatherback winter (-90 CI)	Leatherback winter (+90 CI)
2007	19.60	5.18	34.97	9.77	2.58	17.42
2008	18.67	4.93	33.30	9.16	2.42	16.33
2009	17.74	4.68	31.64	8.55	2.26	15.24
2010	NA	NA	NA	NA	NA	NA
2011	NA	NA	NA	NA	NA	NA
2012	NA	NA	NA	NA	NA	NA
2013	14.00	3.70	24.98	6.10	1.61	10.89
2014	13.07	3.45	23.31	5.49	1.45	9.80
2015	12.14	3.21	21.65	4.88	1.29	8.71
2016	11.20	2.96	19.98	4.27	1.13	7.62
2017	10.27	2.71	18.32	3.66	0.97	6.53
2018	9.34	2.47	16.65	3.05	0.81	5.44
2019	8.40	2.22	14.99	2.44	0.64	4.35
Mean	13.44	3.55	23.98	5.74	1.52	10.23

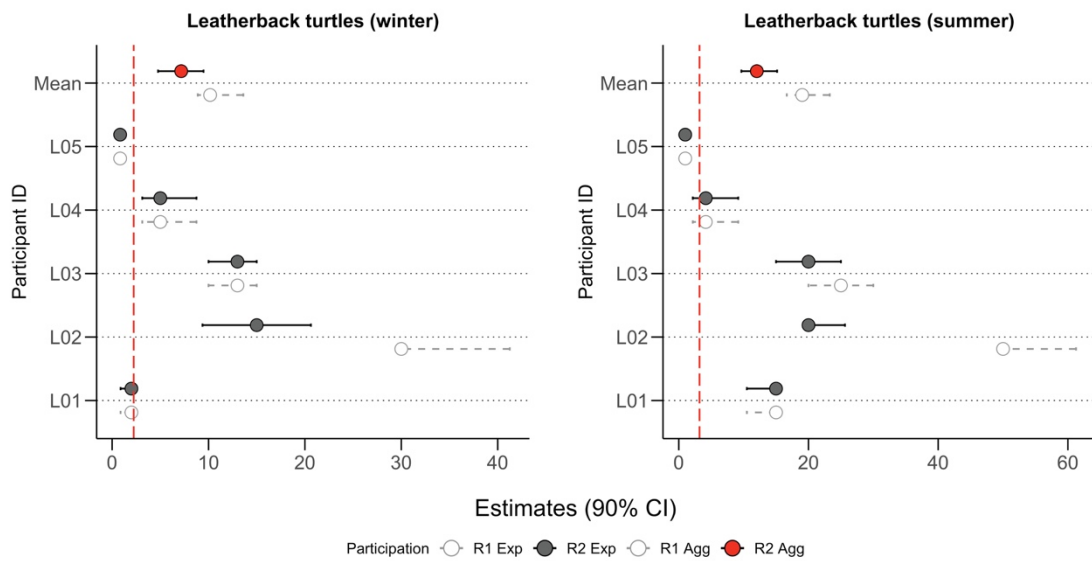


Figure S4.1. Comparing elicited seasonal estimates of the number of leatherback turtles saved from encountering gillnets set by the inshore/midwater fleet as a result of a total gear switch from gillnets to a fishing gear that results in no turtle bycatch (such as lobster potting or trolling) to current bycatch rates calculated from voluntary, at-sea human observer data in San Jose, Peru. Elicited estimates are divided into cold weather months and warm weather months fishing seasons. Experts assumed 100% compliance with the total gear switch scenario. Uncertainty bars have been adjusted to reflect 90% credible intervals for each expert's response. Red dotted line shows the extrapolated estimates of turtle catch from the observer data.

Appendix S5. Elicitation Data

Question 1: How many green turtles in winter per month would be saved using a total gillnet ban, with gear switching to lobster potting or hand line fishing required?

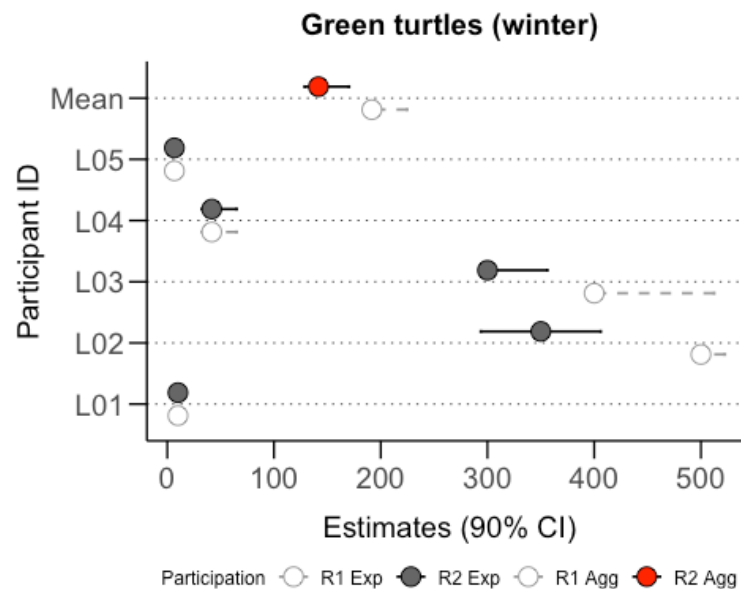


Figure S5.1. Round 1 and 2 estimates for Question 1: How many green turtles in winter per month would be saved using a total gillnet ban, with gear switching to lobster potting or hand line fishing required? Intervals have been standardised to 90%. The graph shows estimates for each expert in Round 1 (R1 Exp), and Round 2 (R2 Exp) and the aggregations in Round 1 (R1 Agg), Round 2 (R2 Agg).

Table S5.1. Round 1 and 2 estimates for Question 1: How many green turtles in winter per month would be saved using a total gillnet ban, with gear switching to lobster potting or hand line fishing required? Intervals have been standardised to 90%.

ID	Stakeholder group	Round	Lower	Upper	Best	Conf	Respondent comments
Mean	Mean	NA	189.79	224.85	191.67	90%	NA
Mean	Mean	NA	128.54	170.35	141.67	90%	NA
L01	Gillnet skipper	1	10.00	16.43	10.00	90%	I think all turtles that are usually captured in San Jose nets would be saved with this strategy. We don't fish as much in winter as we do in summer, so I am considering the differences in how much we fish between seasonal periods.
L01	Gillnet skipper	2	10.00	16.43	10.00	90%	I don't have any changes to make to my original estimates.
L02	Gillnet skipper	1	500.00	522.50	500.00	90%	Green turtles are captured in the highest numbers in our nets. I think this trend is consistent across the fleet.
L02	Gillnet skipper	2	293.75	406.25	350.00	90%	I am readjusting my estimate down as I had a really high monthly turtles saved, I was thinking too much about fishing further north and not considering more southern inshore/midwater boats launching from San Jose.
L03	Gillnet skipper	1	400.00	512.50	400.00	90%	I think that all the turtles that are usually captured in nets would be saved if there was a total ban and we switched to these fishing methods.
L03	Gillnet skipper	2	300.00	356.25	300.00	90%	When I consider that more green turtles are often caught when we head north rather than south, I'm going to readjust my estimate down as captures may not be evenly spaced
L04	Not-for-profit	1	32.29	65.10	41.67	90%	No comment.
L04	Not-for-profit	2	32.29	65.10	41.67	90%	I don't have any changes to make to my original estimates.
L05	Not-for-profit	1	6.67	7.74	6.67	90%	Here I am thinking about the total number of turtles likely captured by the San Jose fleet.
L05	Not-for-profit	2	6.67	7.74	6.67	90%	I don't have any changes to make to my original estimates.

Question 2: How many green turtles in summer per month would be saved using a total gillnet ban, with gear switching to lobster potting or hand line fishing required?

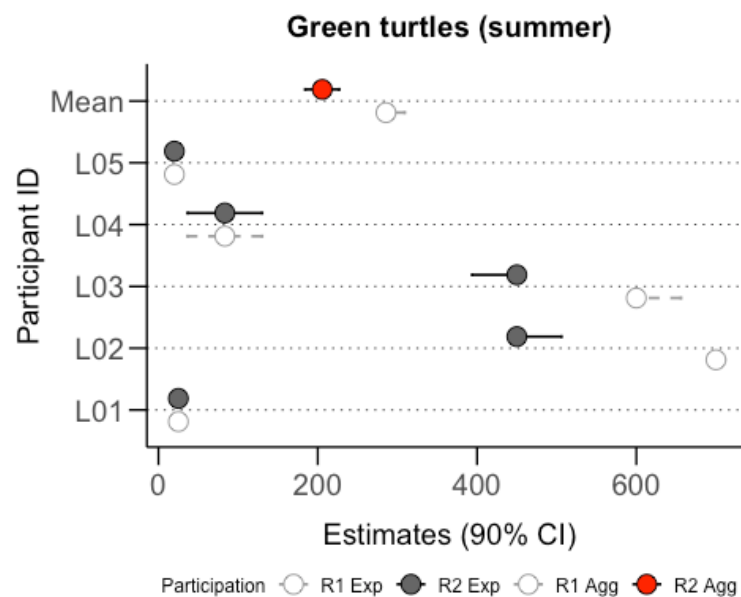


Figure S5.2. Round 1 and 2 estimates for Question 2: How many green turtles in summer per month would be saved using a total gillnet ban, with gear switching to lobster potting or hand line fishing required? Intervals have been standardised to 90%. The graph shows estimates for each expert in Round 1 (R1 Exp), and Round 2 (R2 Exp) and the aggregations in Round 1 (R1 Agg), Round 2 (R2 Agg).

Table S5.2. Round 1 and 2 estimates for Question 2: How many green turtles in summer per month would be saved using a total gillnet ban, with gear switching to lobster potting or hand line fishing required? Intervals have been standardised to 90%.

ID	Stakeholder group	Round	Lower	Upper	Best	Conf	Respondent comments
Mean	Mean	NA	275.39	309.34	285.67	90%	NA
Mean	Mean	NA	184.14	227.09	205.67	90%	NA
L01	Gillnet skipper	1	20.50	25.00	25.00	90%	My summer estimates are higher than my winter estimates as we are generally fishing more days when the sea is not so rough and they do not close the beach due to danger from waves.
L01	Gillnet skipper	2	20.50	25.00	25.00	90%	I don't have any changes to make to my original estimates. I don't catch that many green turtles in my nets. I think some of these other estimates are too high.
L02	Gillnet skipper	1	700.00	711.25	700.00	90%	I am considering how many green turtles encounter my nets and multiplying out. Sometimes we get a single haul with between 30-40 green turtles in it. These are all released, but in summer numbers can be high. In winter captures are lower as we fish less.
L02	Gillnet skipper	2	450.00	506.25	450.00	90%	The same as my green turtle winter estimate - I am going to readjust my estimate down as I think my first estimates were too high due to not considering how turtle captures are often lower when we fish further south.
L03	Gillnet skipper	1	600.00	656.25	600.00	90%	This strategy would be highly effective for reducing turtle encounters with nets; I don't think any turtles would be captured using handlines or potting and we often encounter these in nets, so I imaging across the fleet this would be reasonably high numbers.
L03	Gillnet skipper	2	393.75	450.00	450.00	90%	It seems like some other skippers may capture lower numbers of green turtles than I do, so perhaps I was overestimating. I would like to readjust.
L04	Not-for-profit	1	36.46	130.21	83.33	90%	No comment.
L04	Not-for-profit	2	36.46	130.21	83.33	90%	I don't have any changes to make to my original estimates.
L05	Not-for-profit	1	20.00	24.00	20.00	90%	Thinking about the total number of turtles likely captured by the San Jose fleet.
L05	Not-for-profit	2	20.00	24.00	20.00	90%	I don't have any changes to make to my original estimates.

Question 3: How many leatherback turtles in winter per month would be saved using a total gillnet ban, with gear switching to lobster potting or hand line fishing required?

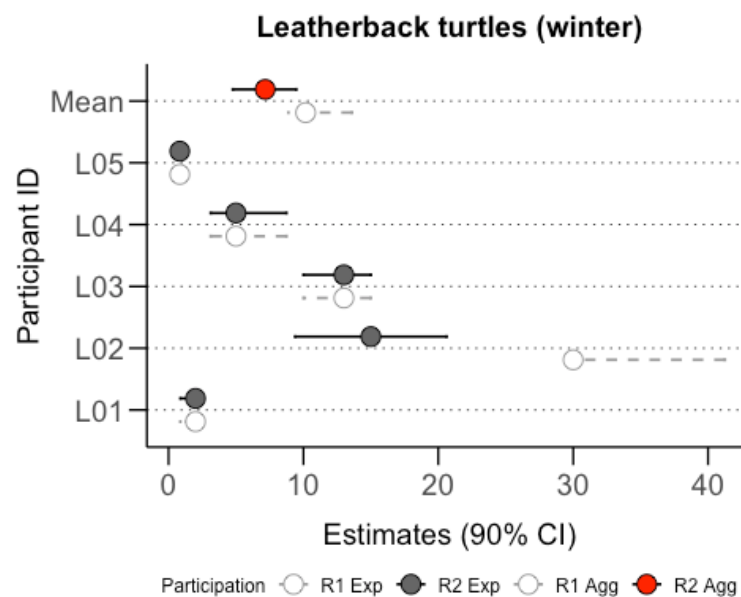


Figure S5.3. Round 1 and 2 estimates for Question 7: How many leatherback turtles in winter per month would be saved using a total gillnet ban, with gear switching to lobster potting or hand line fishing required? Intervals have been standardised to 90%. The graph shows estimates for each expert in Round 1 (R1 Exp), and Round 2 (R2 Exp) and the aggregations in Round 1 (R1 Agg), Round 2 (R2 Agg).

Table S5.3. Round 1 and 2 estimates for Question 3: How many leatherback turtles in winter per month would be saved using a total gillnet ban, with gear switching to lobster potting or hand line fishing required? Intervals have been standardised to 90%.

ID	Stakeholder group	Round	Lower	Upper	Best	Conf	Respondent comments
Mean	Mean	NA	8.89	13.61	10.17	90%	NA
Mean	Mean	NA	4.76	9.48	7.17	90%	NA
L01	Gillnet skipper	1	0.88	2.00	2.00	90%	All turtles that are usually captured in nets would be saved with this strategy
L01	Gillnet skipper	2	0.88	2.00	2.00	90%	I don't have any changes to make to my original estimates.
L02	Gillnet skipper	1	30.00	41.25	30.00	90%	No comment.
L02	Gillnet skipper	2	9.38	20.62	15.00	90%	I am considering the captures of laud (leatherback turtles) that I hear about and then extrapolating out. Captures definitely occur.
L03	Gillnet skipper	1	10.00	15.00	13.00	90%	All turtles that are usually captured in nets would be saved if there was a total ban, I would think a maximum of 15 turtles per month in winter would be a good estimate.
L03	Gillnet skipper	2	10.00	15.00	13.00	90%	I don't have any changes to make to my original estimates.
L04	Not-for-profit	1	3.12	8.75	5.00	90%	No comment
L04	Not-for-profit	2	3.12	8.75	5.00	90%	I don't have any changes to make to my original estimates.
L05	Not-for-profit	1	0.43	1.03	0.83	90%	Thinking about the total number of turtles likely captured by the San Jose fleet
L05	Not-for-profit	2	0.43	1.03	0.83	90%	I don't have any changes to make to my original estimates.

Question 4: How many leatherback turtles in summer per month would be saved using a total gillnet ban, with gear switching to lobster potting or hand line fishing required?

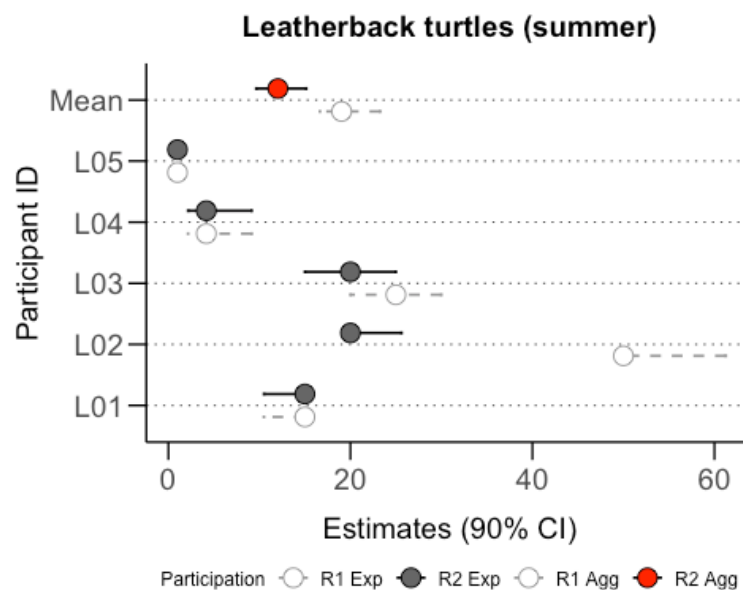


Figure S5.4. Round 1 and 2 estimates for Question 4: How many leatherback turtles in summer per month would be saved using a total gillnet ban, with gear switching to lobster potting or hand line fishing required? Intervals have been standardised to 90%. The graph shows estimates for each expert in Round 1 (R1 Exp), and Round 2 (R2 Exp) and the aggregations in Round 1 (R1 Agg), Round 2 (R2 Agg).

Table S5.4. Round 1 and 2 estimates for Question 3: How many leatherback turtles in summer per month would be saved using a total gillnet ban, with gear switching to lobster potting or hand line fishing required? Intervals have been standardised to 90%.

ID	Stakeholder group	Round	Lower	Upper	Best	Conf	Respondent comments
Mean	Mean	NA	16.65	23.28	19.03	90%	NA
Mean	Mean	NA	9.65	15.16	12.03	90%	NA
L01	Gillnet skipper	1	10.50	15.00	15.00	90%	I think laud (leatherback turtles) in particular wouldn't be captured by handlines, there would be little overlap as handline fishers don't venture too far from the coast.
L01	Gillnet skipper	2	10.50	15.00	15.00	90%	I don't have any changes to make to my original estimates.
L02	Gillnet skipper	1	50.00	61.25	50.00	90%	No comment.
L02	Gillnet skipper	2	20.00	25.62	20.00	90%	I am readjusting my estimate down. When I consider how many leatherback turtles I hear about, I think that these estimates were too high at first. I think these occur more in summer. We are meant to let the local IMARPE officer know if we catch them
L03	Gillnet skipper	1	20.00	30.00	25.00	90%	No comment.
L03	Gillnet skipper	2	15.00	25.00	20.00	90%	Considering that I rarely hear about laud (leatherback turtles) being captured and looking at the other estimates, I think I was too high with my first estimate
L04	Not-for-profit	1	2.17	9.17	4.17	90%	No comment.
L04	Not-for-profit	2	2.17	9.17	4.17	90%	I don't have any changes to make to my original estimates.
L05	Not-for-profit	1	0.60	1.00	1.00	90%	Thinking about the total number of turtles likely captured by the San Jose fleet
L05	Not-for-profit	2	0.60	1.00	1.00	90%	I don't have any changes to make to my original estimates.

Literature cited

- Alfaro-Shigueto J, Mangel JC, Darquea J, Donoso M, Baquero A, Doherty PD, Godley BJ. 2018. Untangling the impacts of nets in the southeastern Pacific: Rapid assessment of marine turtle bycatch to set conservation priorities in small-scale fisheries. *Fisheries Research* **206**:185-192.
- Alfaro-Shigueto J, Mangel JC, Bernedo F, Dutton PH, Seminoff JA, Godley BJ. 2011. Small-scale fisheries of Peru: a major sink for marine turtles in the Pacific. *Journal of Applied Ecology* **48**:1432-1440.
- Arlidge WNS, Firth JA, Alfaro-Shigueto J, Ibañez-Erquiaga B, Mangel JC, Squires D, Milner-Gulland EJ. in review. Understanding the potential for information spread about fisheries bycatch reduction initiatives using cross-contextual information-sharing networks. *Science Advances*.
- Brittain S, Bata MN, De Ornellas P, Milner-Gulland E, Rowcliffe M. 2018. Combining local knowledge and occupancy analysis for a rapid assessment of the forest elephant *Loxodonta cyclotis* in Cameroon's timber production forests. *Oryx*:1-11.
- Burgman MA, McBride M, Ashton R, Speirs-Bridge A, Flander L, Wintle B, Fidler F, Rumpff L, Twardy CJPO. 2011. Expert status and performance. **6**:e22998.
- Efron B, Tibshirani R. 1993. *An Introduction to the Bootstrap*. New York: Chapman & Hall/CRC.
- Hemming V, Walshe TV, Hanea AM, Fidler F, Burgman MA. 2018. Eliciting improved quantitative judgements using the IDEA protocol: A case study in natural resource management. *PloS one* **13**:e0198468.
- McBride MF, Fidler F, Burgman MAJD, Distributions. 2012. Evaluating the accuracy and calibration of expert predictions under uncertainty: predicting the outcomes of ecological research. **18**:782-794.
- van der Hoeven CA, de Boer WF, Prins HH. 2004. Pooling local expert opinions for estimating mammal densities in tropical rainforests. *Journal for nature conservation* **12**:193-204.