The following supplement accompanies the article

Integrated mixed-effect growth models for species with incomplete ageing histories: a case study for the loggerhead sea turtle *Caretta caretta*

Brandon E. Chasco*, James T. Thorson, Selina S. Heppell, Larisa Avens, Joanne Braun McNeill, Alan B. Bolten, Karen A. Bjorndal, Eric J. Ward

*Corresponding author: brandon.chasco@noaa.gov

Marine Ecology Progress Series 636: 221–234 (2020)

Text S1. File descriptions for recreating the results of this analysis.

We provide a list of R scripts to reproduce the figures and tables in the main body of the paper and the appendix material. All analysis files and simulation files are independent: that is they will read in, or simulate, the necessary data to recreate the results.

Table S1. Recreating the analysis of the observed data. The R script, TMB code, or data files used to recreate the analyses for the main body of the paper and the appendix material.

Location	File type	File	Paper reference
Main body of paper	Analysis of data	runIntegratedHumerusMR.r	Combines both the mark-recapture and stranding data
		runNonIntegratedHumerus.r	Runs just the stranding data
		runNonIntegratedMR.r	Runs just the mark-recapture tagging data.
	Data files	tagDat.dat	Tagging data
		strandingData.dat	Summary of non-LAG stranding data
		lagDiameter2.dat	LAG data associated with a stranded individual
	TMB cpp files	IntegratedHumerusMR.cpp	TMB code for integrated stranding and tagging data
		NonIntegratedHumerus.cpp	TMB code for stranding data alone
		NonIntegratedMR.cpp	TMB code for tagging data alone
Appendix material	Simulation of sample size effects	simulationModelSampleSIze.r	R script for the sample size analysis
	Simulation of model selection	simulationModelSelection.r	R script to estimate the effects of model selection criteria
	Simulation of variability effects	simulationPersistentAndTransientVar.r	R script to estimate the effects of different sources of variability
	The function for creating the simulated data	func_simulation.r	R script

Table S2. R scripts used to recreate the figures in this analysis: the name of the R script and the associated plot in the main body of the paper and the supplemental material.

Location	Output	R-script	Paper reference
Main body of paper	Figures	Fig_ggplotgrowthProcesses4.r	Figure 1
		Fig_comparisonMarkRecapture.r	Figure 2
		Fig_comparisonAvensHumerus.r	Figure 2
		Fig_ggplot_Age4.r	Figure 3
		func_quantAgeAtLen.r	Figure 3
		func_drawASM_percent.r	Figure 3
		Fig_ageErrorNyLength.r	Figure 4
		Fig_sizeAtAge_andLengthFreq.r	Figure 5
	Tables	table_modelSummary.r	Table 2.
Appendix	Figures	ggplot_modelSelectionExperiment.r	Figure-A.1
		ggplot_sampleSizeExperiment_Par.r	Figure-A.2
		ggplot_sampleSizeExperiment_sigma.r	Figure-A.3
		ggplot_persistentAndTransientVar_Par.r	Figure-A.4
		ggplot_persistentAndTransientVar_Sigma.r	Figure-A.5
		fig_simulationDataIntegration.r	Figure-A.6

Table S3. The instantaneous and incremental growth processes for the von Bertalanffy, Gompertz and logistic models. The parameters of the model are L_∞, L_0, and k , and a and δ are the age and time-at-liberty between observations, respectively.

Process	Model	Equation
Size-at-age	von Bertalanffy	$L(a) = L_\infty \times \left(1 - \frac{(L_\infty - L_0)}{L_\infty} \times \exp(-k \times a)\right)$
	Gompertz	$L(a) = \exp\left(\log(L_\infty) \times \left(1 - \frac{(\log(L_\infty) - \log(L_0))}{\log(L_\infty)} \times \exp(-k \times a)\right)\right)$
	logistic	$L(a) = L_\infty \times \frac{(1-b\times\exp(k\times a))}{b+(1-b)\times\exp(k\times a)}$ $b = 1 - \frac{L_0}{L_\infty}$
Growth increment	von Bertalanffy	$L(a+\delta) = L(a) + (L_\infty - L(a)) \times (1 - \exp(-k \times \delta))$
	Gompertz	$L(a+\delta) = \exp\left(\log\big(L(a)\big) + \big(\log(L_\infty) - \log\big(L(a)\big)\big) \times (1 - \exp(-k \times \delta))\right)$
	logistic	$L(a+\delta) = \left(\frac{1}{L(a)} + \left(\frac{1}{L_\infty} - \frac{1}{L(a)}\right) \times (1 - \exp(-k \times \delta))\right)^{-1}$

Text S2: Methods for simulation testing

To determine the bias and precision of the parameters, and satisfy that the fixed and random effects of the model are identifiable, we compared parameter fits from the IME model to simulated data derived from known parameters. We conducted three experiments to evaluate statistical consistency of the estimators, and model misclassification using AIC based different random effects and growth processes. The first experiment examined the consistency of the estimators and misclassification of the true growth process based on AIC. The misclassification of the growth process was examined based on a 3 by 3 by 3 factorial design. We simulated the data for each of the three growth processes (i.e., von Bertalanffy, gompertz, and logistic) and then fitted all three growth processes to the simulated data. We saved the maximum likelihood parameter estimates for each estimation model and calculated the marginal AIC. We repeated this for sample sizes of 50, 150, and 300 stranded and tagged turtles. Thus, the 3 X 3 X 3 included three estimation models fitted to three simulated data sets for three different sample sizes. The second experiment was a 3 by 4 factorial design that evaluated how the magnitude of the transient and persistent variability effects both the consistency of the estimators and model misclassification using AIC. We simulated data from the von Bertalanffy growth process and with three levels of persistent and transient deviations $\left(\text{i.e., } \phi_{k_i}^d,\ \phi_{k_{i,j}}^d, \text{respectively}\right)$ of (0.1, 0.7), (0.4, 0.4) and (0.7, 0.1). For each of these three scenario, we then fitted four estimation models that included persistent random effects for k and L_∞, persistent effects for just k or L_∞, and then no persistent effects for either k or L_∞.

The third experiment examined the effects of independent and integrated data sets for estimating the ages of the sea turtles. In this last experiment, we simulated a random data set for both stranded and tagged turtles then we fitted the model to the integrated data, and to each independent data set. For the scenarios of each experiment, we conducted 100 simulations. Unless stated for a particular experimental design above, we have described base case parameters below.

The asymptotic size (μ_∞) was equal to 95 cm, and the size at age 0 (L_0) was equal to 5 cm. The growth coefficient k was equal to 0.07 year^{-1} for the von Bertalanffy, 0.09 year^{-1} for the gompertz, and 0.12 year^{-1} for the logistic. The average age of the first observation was eight for stranded turtles and 15 for tagged turtles. The standard deviations for the log-normally distributed age effects (ϕ_a^d) was equal to 0.8 stranded turtles and 0.3 for the tagged turtles. Based on work by Avens et al. (2013), we assumed no effect of sex on either the asymptotic size or growth rate parameter. The body proportional function (BPF) that describes the ratio between the LAG diameters and straight carapace length (β) was equal to 0.38mm cm^{-1} (Snover et al. 2007) and the observation error was normally distributed with a standard deviation equal to 2 mm. The observation errors of the LAG diameters for stranded turtles and carapace length from mark recapture turtles were log-normally distributed with standard deviations of 0.001 and 0.01, respectively. The amount of time (year) that represents the humerus growth beyond the last LAG was uniformly distributed ($\text{Uniform}(0,1)$).

The number of LAG per stranded turtle and the number of recaptures per tagged turtle were sampled with replacement from the observed numbers of LAG and recaptures for the Western Atlantic loggerheads (Braun-McNeill et al. 2008). While the time-at-liberty between the LAG observations was fixed to one year for the stranded turtles (Goshe et al. 2016), the time-at-liberty (years) for mark-recapture observations varies according to a gamma random variable (shape = 0.45, scale = 2.2) based on the observed times-at-liberty for loggerheads collected in the Western Atlantic. Additionally, there was a small number of stranded turtles with complete ageing histories (i.e., no missing LAG), which we set to 5% of the sample size.

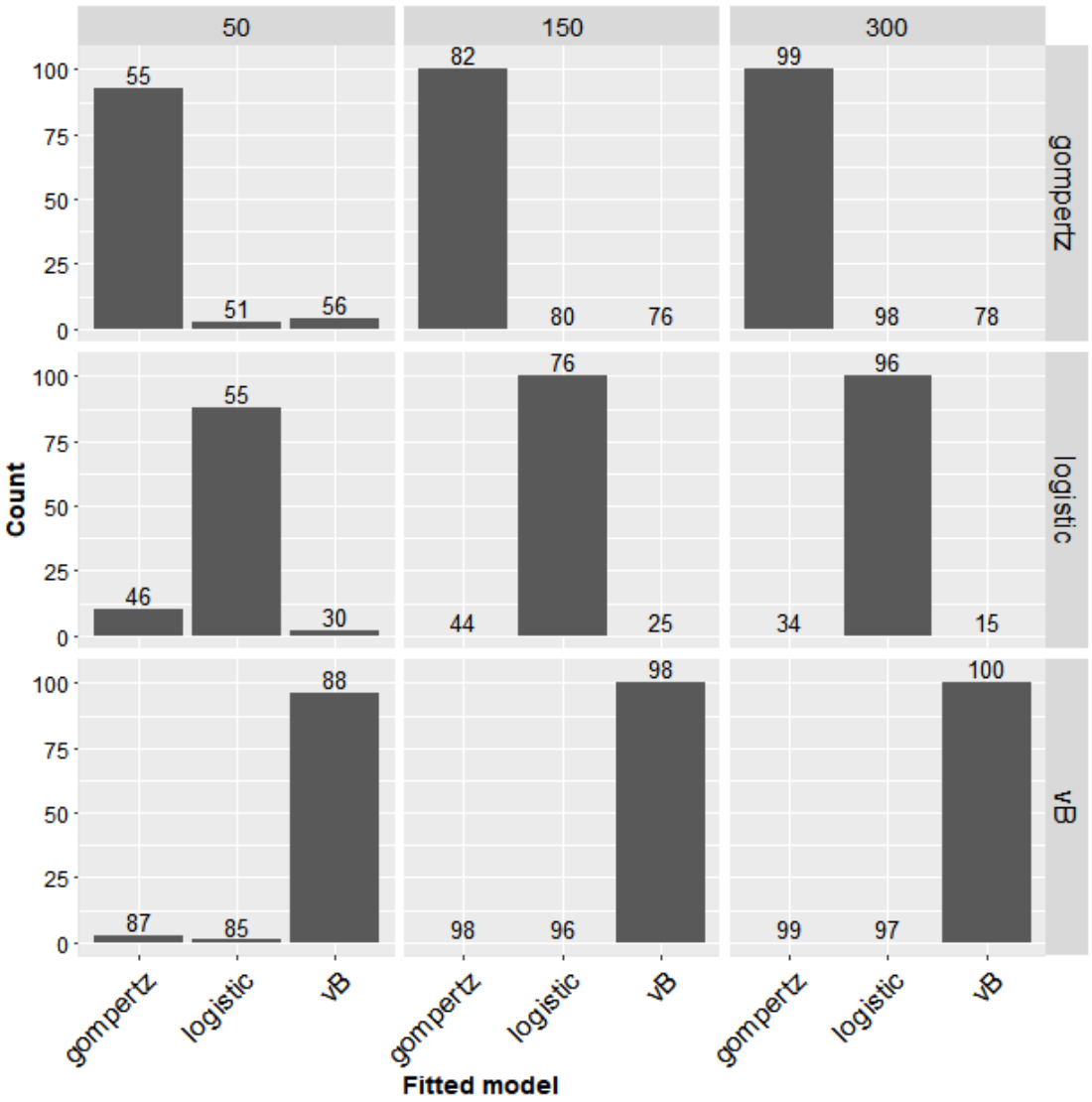

Figure S1. Model selection experiment with 3 X 3 X 3 factorial design for sample size, simulated model, and fitted model. The height of the bar represents the percent of the simulations for which the fitted growth process was selected for a given simulated growth process, and the numbers above the bars equal the number of models that converged with gradient less than 1e-4 and a positive definite hessian.

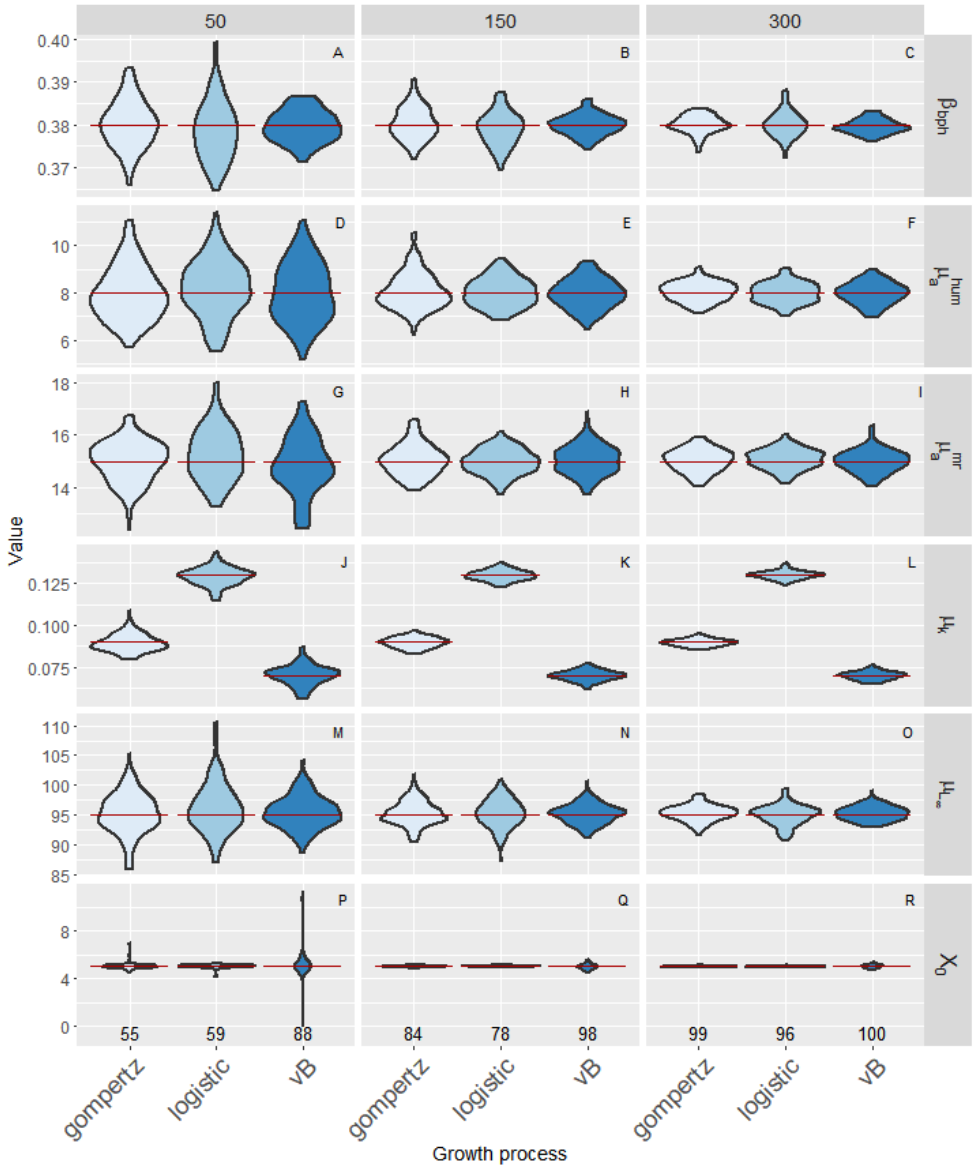

Figure S2. Power analysis for the biological parameters related to the three growth processes. Red horizontal lines represent the simulated values and plots represent the density of the fitted values. Numbers above each growth process on the x-axis represent the percentage of models that converged based on maximum gradient less than 1e-4 and a positive definite hessian.

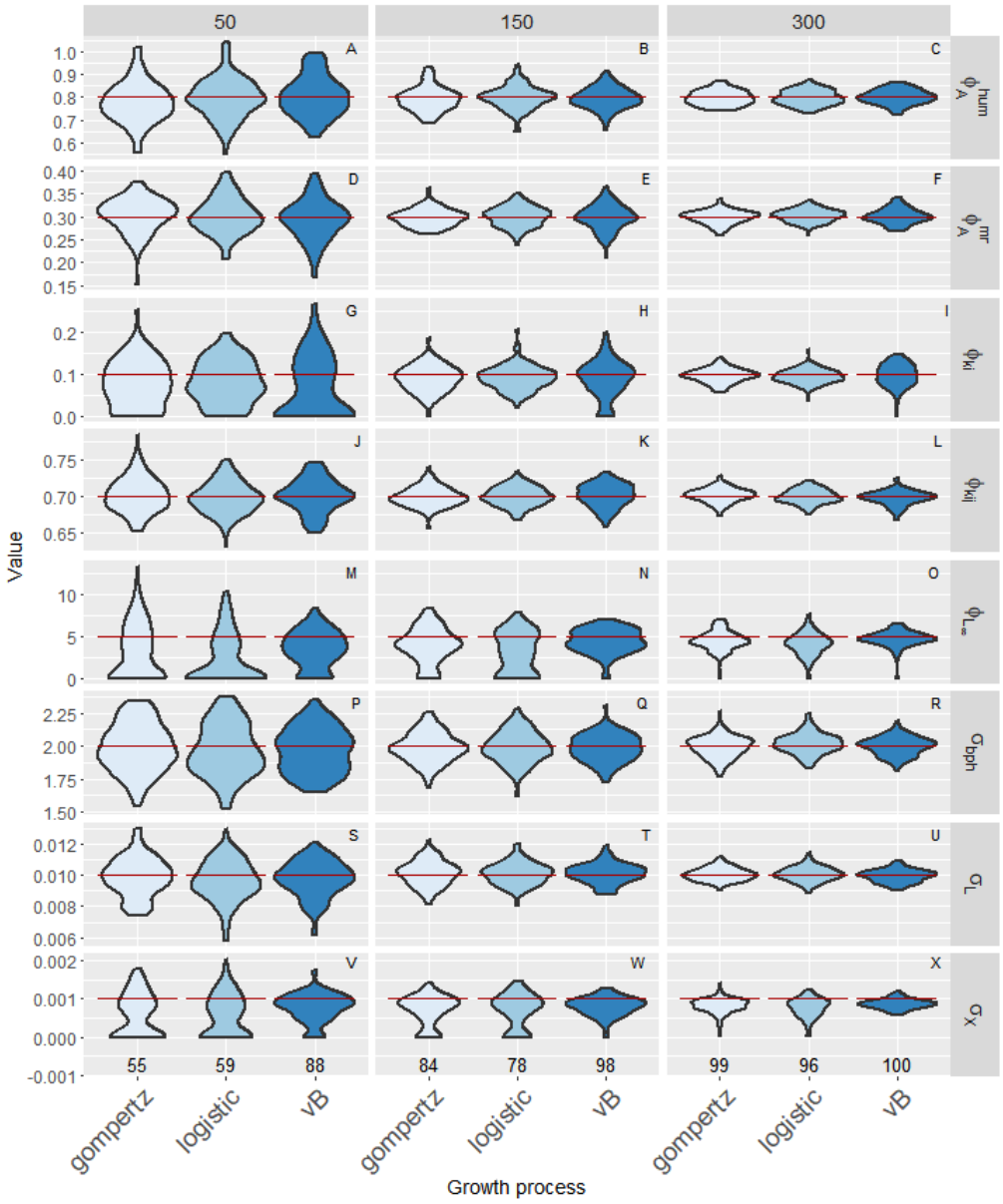

Figure S3. Power analysis for the variance parameters related to the observation and process for the three growth processes. Red horizontal lines represent the simulated values and plots represent the density of the fitted values. Numbers above each growth process on the x-axis represent the percentage of models that converged based on maximum gradient less than 1e-4 and a positive definite hessian.

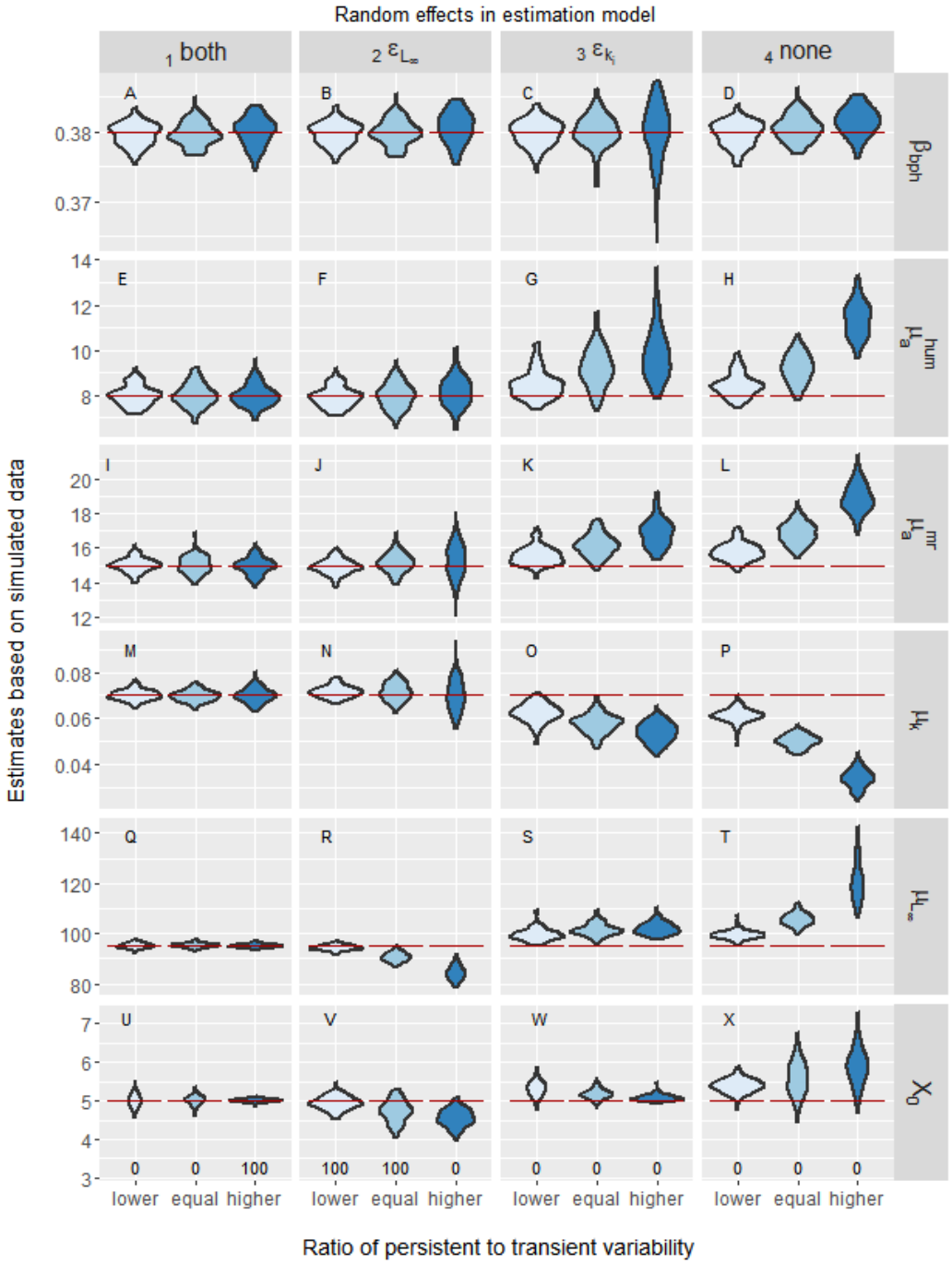

Figure S4. The bias and precision of the von Bertalanffy growth process based on 3 X 4 factorial design examining the ratio between persistent and transient variability in the growth coefficient, and the levels of persistent variability in the estimation model. The ratio between the persistent and transient variability include: lower; ϕ_{k_i} equal 0.1 and $\phi_{k_{ij}}$ equals 0.7, equal; ϕ_{k_i} equal 0.4 and $\phi_{k_{ij}}$ equals 0.4, and higher; ϕ_{k_i} equal 0.7 and $\phi_{k_{ij}}$ equals 0.1. Numbers along the x-axis indicate the percentage of estimation models selected using AIC. If AIC correctly match the estimation to the simulated data, all ratio experiments in the first column should have 100 over the top of them. Red lines are the true values in the simulation models and density plots represent the variability in the parameters from the estimation model.

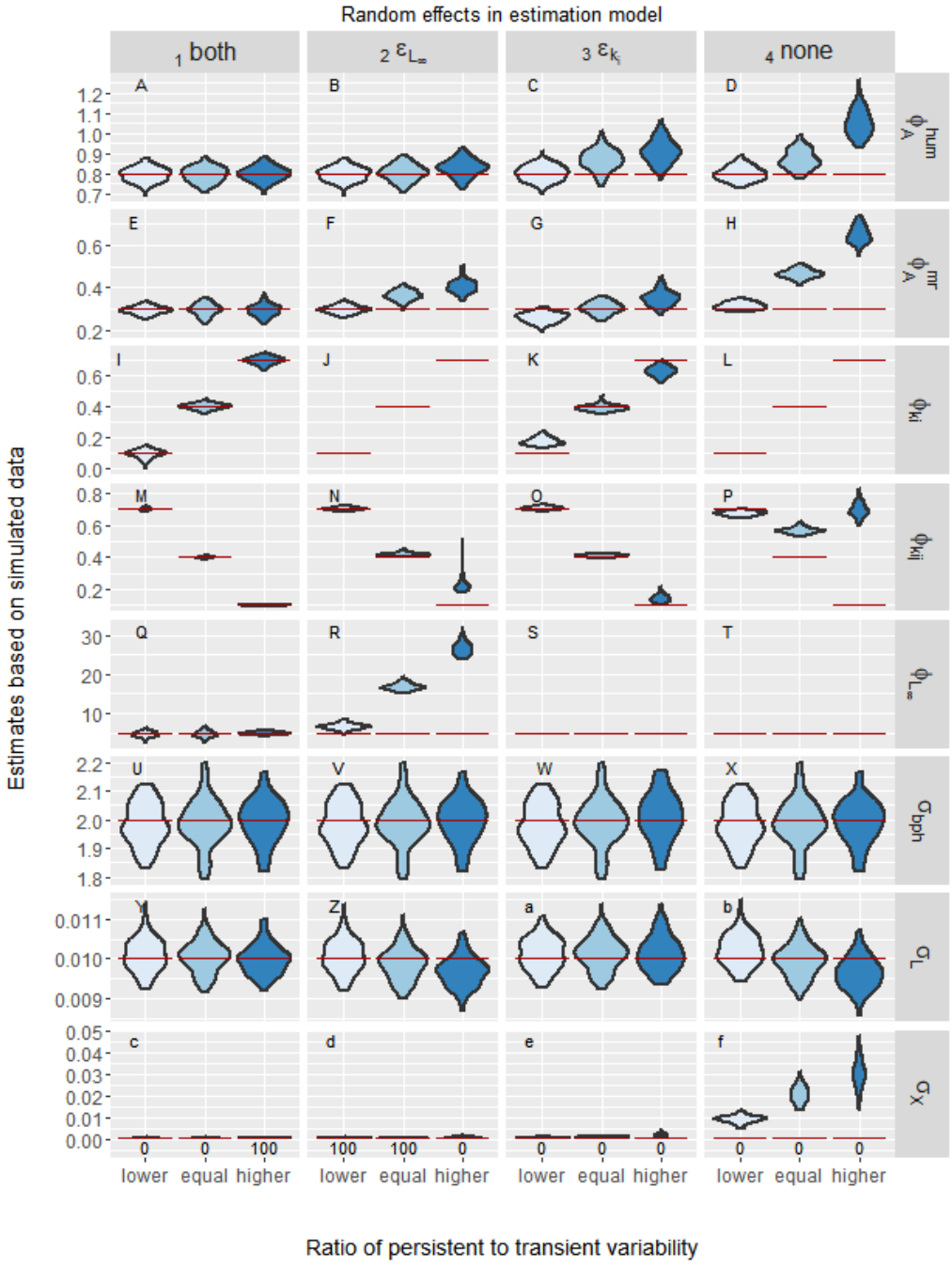

Figure S5. The bias and precision of the observation and process errors based on 3 X 4 factorial design examining the ratio between persistent and transient variability in the growth coefficient, and the levels of persistent variability in the estimation model. The ratio between the persistent and transient variability includes: low; ϕ_{k_i} equal 0.1 and $\phi_{k_{ij}}$ equals 0.7, equal; ϕ_{k_i} equal 0.4 and $\phi_{k_{ij}}$ equals 0.4, and higher; ϕ_{k_i} equal 0.7 and $\phi_{k_{ij}}$ equals 0.1. Numbers along the x-axis indicate the percentage of estimation models selected using AIC. If the AIC correctly match the estimation to the simulated data, all the experiments in the first column should have 100 over the top of them. Red lines are the true values in the simulation models and density plots represent the variability in the parameters for the estimation model.

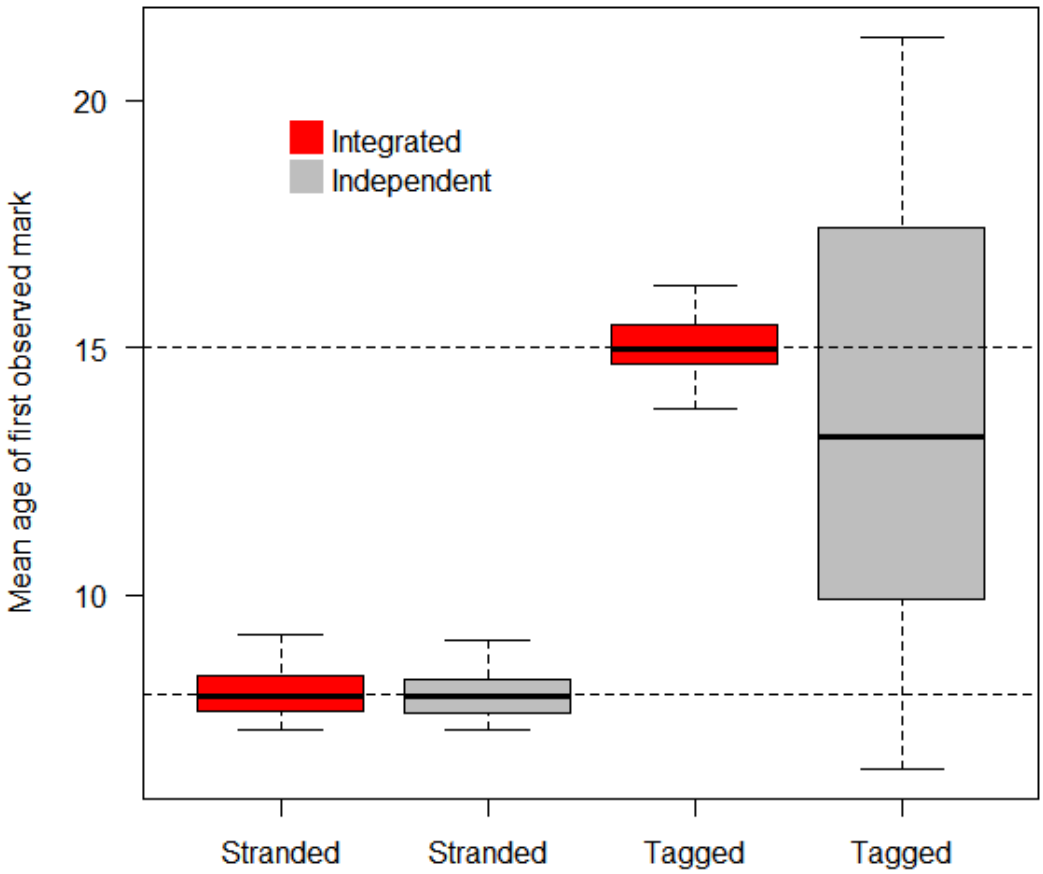

Figure S6. Comparison between the estimated mean ages of the first mark for stranded (red) and tagged (grey) turtles using independent and integrated data sets. Horizontal dashed lines represent the true simulated mean ages for the stranded (eight years) and tagged (15 years) turtles. Box and whiskers represent the 75% and 95% confidence intervals for the estimated mean ages, and the horizontal black lines represent the median estimates.

Text S3. Estimated size-at-sexual maturity.

Based on comparisons between the rapprochement size associated with sexual maturation and asymptotic size (Avens et al. 2015), we assumed the size-at-sexual maturity for each turtle was 95% of the asymptotic size ($SSM_{0.95}$). The age associated with $SSM_{0.95}$ was the estimated age-at-sexual maturity ($ASM_{0.95}$). To account for variability between size-at-sexual maturity we assumed a standard deviation of 0.02, ($SSM_i \sim Beta(\bar{x} = 0.95, \sigma = 0.02) \times L_{\infty,i}$). Using the best-fit model, a von Bertalanffy growth process with persistent variation for only the asymptotic size, we compared the ages and $ASM_{0.95}$ for both the integrated data and non-integrated data.

References

Avens, L., Goshe, L.R., Pajuelo, M., Bjorndal, K.A., MacDonald, B.D., Lemons, G.E., Bolten, A.B., and Seminoff, J.A. (2013) Complementary skeletochronology and stable isotope analyses offer new insight into juvenile loggerhead sea turtle oceanic stage duration and growth dynamics. Mar. Ecol. Prog. Ser. 491: 235–251

Braun-McNeill, J., Epperly, S.P., Avens, L., Snover, M.L., and Taylor, J.C. (2008) Growth rates of loggerhead sea turtles (Caretta caretta) from the western North Atlantic. Herpetol. Conserv. Biol. 3(2): 273–281

Goshe, L.R., Snover, M.L., Hohn, A.A., and Balazs, G.H. (2016) Validation of back-calculated body lengths and timing of growth mark deposition in Hawaiian green sea turtles. Ecol. Evol. 6(10): 3208–3215