

Fish intraspecies body size distributions.

Freddie J. Heather Shane A. Richards Micheal Spence?
Simon Brandl Asta Audzijonyte

Abstract goes here

Background

Body size is a key trait in determining how organisms interact with their environment, this is particularly true for marine organisms, where body size is a strong predictor of trophic position¹, growth rate, and predation mortality², among many other life-history parameters. With warming waters, there is no consensus on whether wild populations of fish are getting larger or smaller³, **TSR, James' Rule**). The problem with this question however is that a species cannot be defined by a single body size - what does it really mean to say that a species is getting smaller? It is theoretically possible for a species mean size to decrease, and yet the 'maximum' body size to increase. In order to understand how environmental drivers such as warming waters and increased fishing pressures are influencing the body size of fish species it is important to considerer the entire species body size distribution. Empirical intra-species body size distributions are generally limited to a small number of fish species (**refs**), the majority of which are commercially important species (**refs**).

Estimations of unexploited biomass are necessary for the stock-assessments, yet these are often based on expert advice, which may lead to subjective or biased estimates of exploitation levels^{4,5}. Among fisheries scientists, there appears to be no consensus on the body size distribution of an unexploited fish population, partly due to the majority of body size data coming from fishery-dependant sources with size selective biases (e.g., fishing net mesh size). Given a snapshot in time, one might expect the greatest number of individuals in the smallest size classes, as small individuals give rise to larger individuals, and a proportion of which succumb to predation or natural mortality. Despite many modelling approaches assuming this monotonically declining abundance with increasing body size^{6,7}, observational data do not support this (e.g. in mammals⁸ **other refs**). Fisheries-independent fish body size distributions often display a uni-modal shape, with peak abundance occurring at some intermediate body size (**refs**). In this study we examine the body size distributions of 3K (**get exact number**) populations, from 1064 reef fish species, predominantly throughout Australian continent but also globally. We compare two statistical distributions, the lognormal and the normal distribution,

in their ability to describe observed fisheries-independent body size distributions. Further, we test a conjecture by Giometto et al.⁹ who used unicellular protists under experimental laboratory conditions to claim that the mean size of a protist is sufficient to describe the entire body size distribution, to see if this hold to wild populations of fish species.

We fit and compare the likelihood of each of these distributions using a combination of two data collection methods, 1) a continental scale underwater visual census method of 12 million+ individuals which involves body size estimation into pre-defined body size classes, and 2) a global-scale (8 thousand+ individuals) fine-mesh netting approach targetting small cryptobenthic species, providing continuous body size distributions.

Results & discussion

Our observations from 607 unique fish species support the conjecture⁹ that a single body size metric (e.g. mean body size) is sufficient to approximate the body size distribution of fish species. We show that 95% of estimates of the coefficient of variation (CV) of both the normal and lognormal distributions fell within the range of 0.14 to 1.64, with a median coefficient of variation of 0.33 (Figure 1). Only one of the 139 cryptobenthic populations (i.e., species at a location) body size distributions (continuous data) was better described by a normal distribution than a lognormal distribution (triangles in Figure 1 A), this population had also the smallest mean size of any population. On the other hand, more of the 3089 binned body size distributions (i.e., species at a $1^\circ \times 1^\circ$ latitude-longitude grid cell) from Reef Life Survey populations were better described by a normal distribution (64%, n = 2989 populations) than a lognormal distribution (36%, n = 1100 populations). The coefficient of variation of body size distributions that were better described by lognormal distributions were not significantly different (median CV = 0.33 ± 0.007) than the those better described by a normal distribution (median CV = 0.32 ± 0.008 , Figure 1 B). These results suggests that irrespective of the ‘preferred’ distribution of the population, a consistent coefficient of variation 0.3 enable the estimation of the entire body size distribution (either lognormal or normal) based on the mean or maximum (e.g. 95th quantile) size of the species.

Body size distributions, once scaled by the total abundance of the population (i.e. abundance density, or probability-at-size), and scaled to the mean size, show remarkable inter-population consistency (Figure 2) and to body size distributions of single-cellular organisms under experimental conditions⁹. It is expected that a single statistical distribution will never fully describe all the observed body size distributions, under a range of environmental conditions and size-selective pressures (such as fishing). Here we show that the normal distribution (truncated at the smallest size class) and the lognormal distribution provide good approximations of the body size distribution of wild fish populations. Given the mean (or even maximum) body size of a population, we can reconstruct the body size distribution based on the assumption of a constant coefficient of variation with mean body size. Here we show that approximately 72% of the variation in body size (71.1% for lognormal, 72.7% for normal distribution, linear mixed

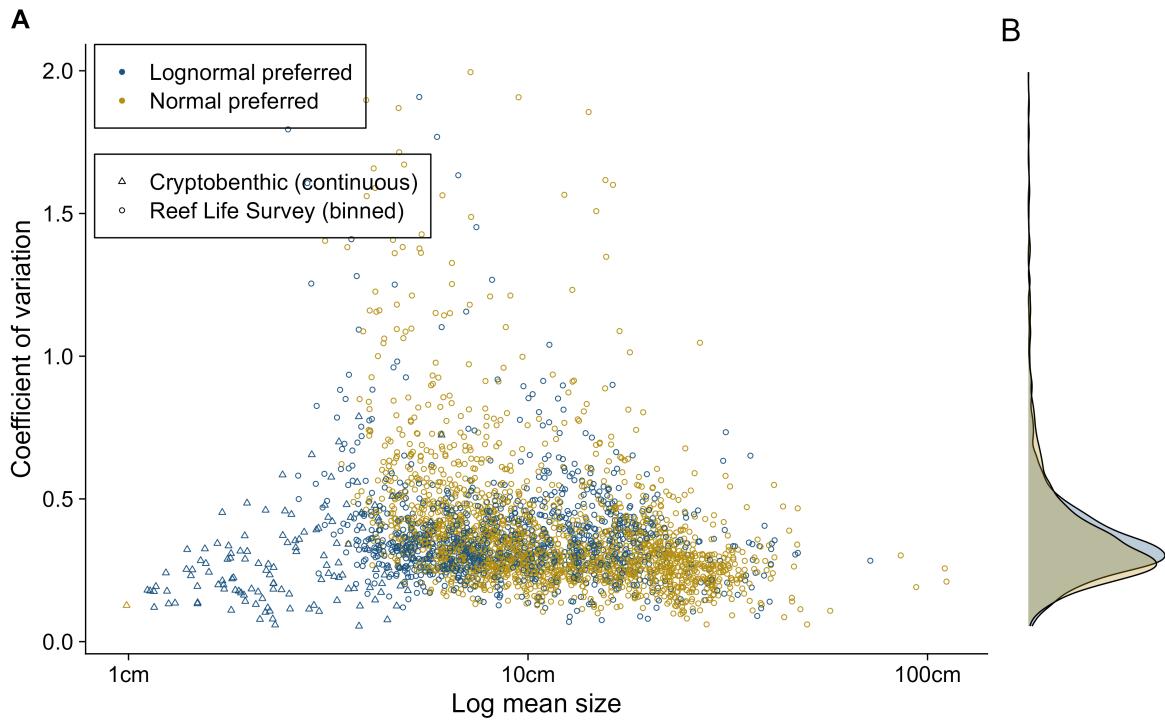


Figure 1: Relatively constant coefficient of variation with increasing species mean body length, across three orders of magnitude in body length. Each datum corresponds to a single fish population (species at a location) from two data-sources, cryptobenthic fishes with continuous body length measures (triangles), and Reef Life Survey with binned estimates of body size (circles). The colour of the point represents the best fitting distribution for that population's body size distribution. The coefficient of variation is calculated from the parameters of the fit of the preferred distribution (lognormal or normal). The y-axis has been limited to an upper value of 2 for visualisation purposes, 97.8% of the data are shown here, see supplementary figure XX for all populations.

effects regression model) can be explained purely based on the mean body size of the species and a constant coefficient of variation of 0.33 (Figure 3).

In this empirical study we do not aim to identify the drivers of the best fitting distribution, rather the observation that even intra-species populations may exhibit a preference for a normal or lognormal distribution. It is likely that size-selective pressures, such as the removal of larger bodied individuals via targeted fishing, will result in shorter right tails to the body size distribution - this may well result in a shift from a preference of a lognormal distribution to a normal distribution (E.g. Figure Figure 2 E).

Further, we show that whilst some populations are better described by a lognormal or normal distribution, their selection makes very little difference in the estimation of the observed body size distribution based on the mean body size of the species.

Reef ecosystems are not necessarily reflective of pelagic fish populations, due to the size-based refugia of tropical and temperate reefs¹⁰. Thus this work should be extended with further data-sources to test this constant CV in other marine, or even terrestrial populations.

Methods

Body size data sources

Fish body size data were obtained through two sources: 1) visual census data from Reef Life Survey (RLS; filtered to Australian populations only, over 12 million individuals, 497 species, binned body size estimates)¹¹, and 2) lethal sampling of Cryptobenthic reef fishes (CBF; global spatial coverage, over 8000 individuals, 127 species, continuous body size measurements to the nearest 0.1mm)¹².

RLS surveys involve an underwater visual census method along 50m transect line, with a diver searching 5m either side of the transect line, the body size of the fish is estimated to the nearest body size bin (2.5, 5, 7.5, 10, 15, 20, 25, 30, ..., 400cm).

Another data source was from field collections using enclosed clove-oil stations at six reefs; Moorea, GoO, AG, Lizard Island, Panama, and Belize. Reef outcrops are selected, measured, and covered with a bell-shaped fine mesh and tarpaulin, before being sprayed with a clove-oil:ethanol solution (1:5). Fish are collected with tweezers and placed in ziplock bags. See Brandl et al.¹² for full methods.

Data filtering

For binned body size data (visual census, Reef Life Survey), we excluded body size distributions that spanned fewer than four body size bins (e.g. 2.5cm, 5cm, and 7.5cm size bins only), we did not apply an exclusion for the continuous data (lethal sampling, cryptobenthic fishes) based on the range of body sizes. For the continuous data we set a minimum count of 10 individuals

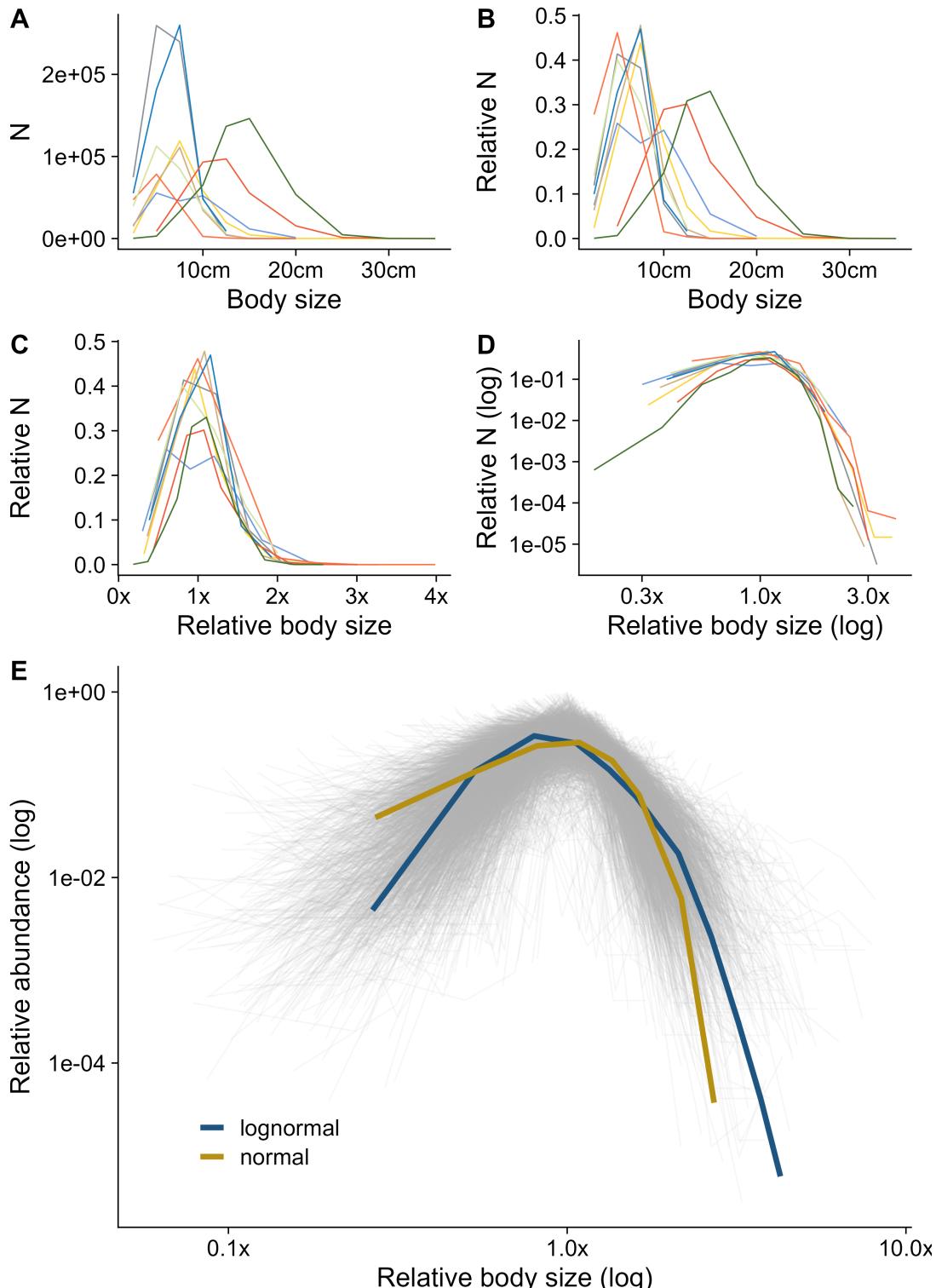


Figure 2: **Observed fish population body size distributions show similar overall shape.** Fish populations have variable median body sizes, total abundances, and range of body sizes (Panel A). If we scale the abundance by the total abundance of the population (Panel B), and scale the body size by the mean body size, we observed very similar shapes in body size distributions on the natural (Panel C) and the logarithmic scale (Panel D). Panel E shows all populations in the study (grey lines) and the median parameter estimates from the normal-preferred (gold coloured) and lognormal preferred (blue coloured) distributions. Panel's A-D show a random sample of 10 populations.

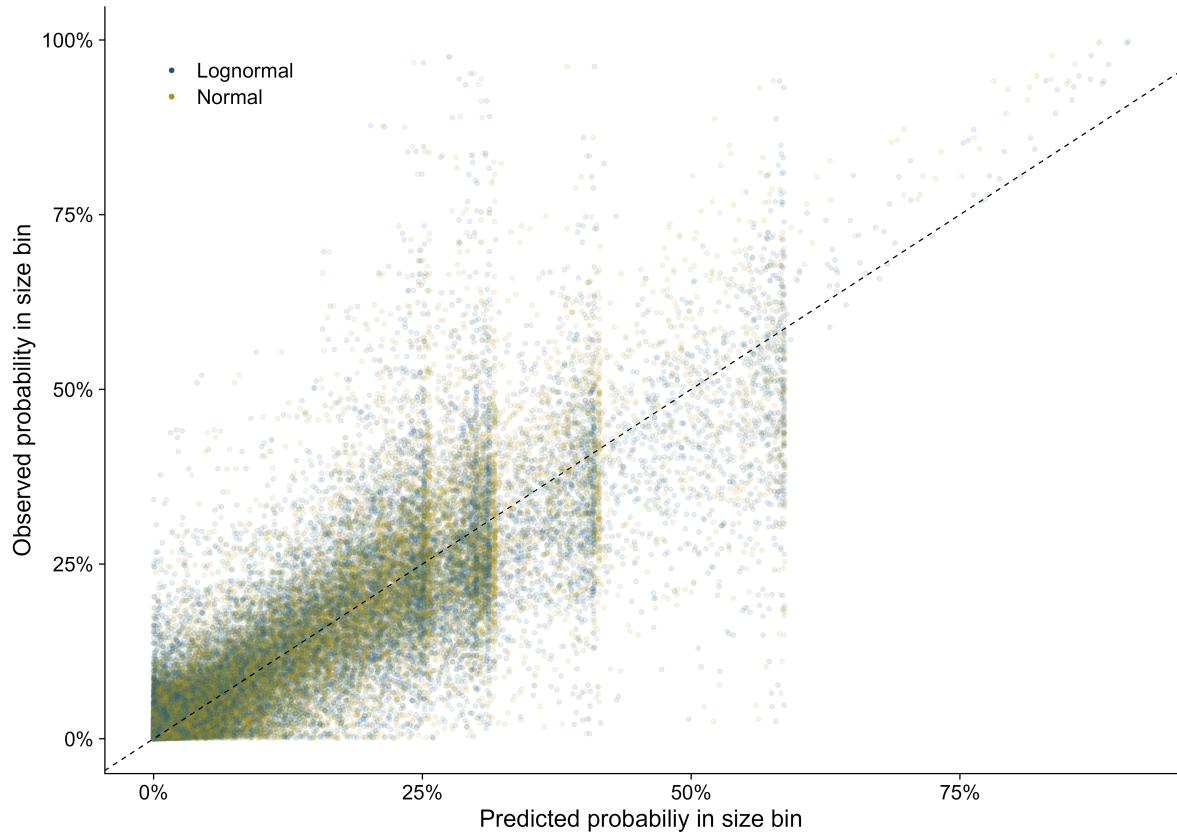


Figure 3: Mean body size and the assumption of a constant coefficient of variation, are sufficient to explain about 70% of the variation in observed body size within body size classes. Lognormal and normal distribution fits perform very similar in their ability to estimate the probability within body size classes.

per population to fit a distribution, for the binned data, this minimum count was 100. We performed a sensitivity analysis to these filtering parameters to show that across a range of filters for minimum count, the overall result did not substantially change.

A population was defined as a species within a location. For the Reef Life Survey dataset, to avoid biases associated with clumped we determined a ‘location’ to be all surveys within a $1^\circ \times 1^\circ$ latitude-longitude gridcell ($n = 3089$ populations,). For the cryptobenthic fish data, a location was defined at the reef outcrop that was netted ($n = 139$ populations).

XX body size distributions were excluded from the final analysis due to model-fitting issues as a result of either a large proportion of individuals in a single body size bin ($n = XX$ populations) or due to bimodal body size distributions ($n = XX$ populations). See Supplementary Material XX for these distributions.

Statistical analysis

All statistical analyses were performed using the statistical programming language R¹³, and in combination with the bayesian statistical modelling language Stan¹⁴.

A lognormal and normal distribution were fitted to each of the cryptobenthic population’s continuous body size distributions.

$$L_i \sim \mathcal{N}(\mu_i, \sigma_i)$$

$$L_i \sim \mathcal{LN}(\nu_i, \tau_i)$$

Where L_i is the observed body length of population, i . Population body length, L_i , is fitted to either a normal, \mathcal{N} , or lognormal distribution, \mathcal{LN} . We have used ν and τ as the location and scale parameters of the lognormal distribution to avoid confusion with μ and σ , of the normal distribution. Each population has its own location (μ_i and ν_i) and scale (σ_i and τ_i) parameter for each of the two distributions, normal and lognormal, respectively.

Reef Life Survey body size data, in body size bins, were also fit using Stan, but using a method to account for the body size binning approach. The probability of being within each RLS body size bin (e.g. 5cm) was calculated by integrating the probability density function of the distribution between the upper (b_u) and lower (b_l) bounds of the size bin, b (e.g. between 3.75cm and 6.25cm for the 5cm size bin). For example, the probability of an individual within the population i , being in size bin b , given body size is normally distributed is:

$$Pr(b, i) = \frac{\phi_i(b_u) - \phi_i(b_l)}{1 - \phi_i(1.25)}$$

Where ϕ_i refers to the cumulative density function of the normal (or lognormal) distribution, with population-level parameters. The denominator is a normalising constant, defined by the probability of being less than 1.25cm; accounting for the body sizes not recorded in the RLS method.

This method was run for both the normal and lognormal distributions. Reasonable starting values were used in the Markov Chain Monte Carlo (MCMC) sampling process of fitting normal distributions to avoid numerical computational limits (i.e., near-zero probability estimates) of the integration of unreasonable distribution parameters. MCMC sampling was run for three chains for 10,000 iterations, with 5,000 iterations discards as part of the burn-in, leaving a sample of 15,000 iterations of the posterior distributions.

The likelihood of the model was defined as the summation of the normalised probability of the abundance within each size bin given the model parameters. Since the lognormal and normal distributions have the same number of parameters, the log-likelihood of the fits of the distributions could be directly compared. For each population the preferred distribution (normal or lognormal) was determined as the distribution with the lowest median log-likelihood value of the posterior distribution.

Coefficient of variation

The coefficient of variation is a measure of the variability in the data relative to the mean. It is calculated as the standard deviation divided by the mean. For the normal distribution this is simply calculated as σ/μ . For the lognormal distribution, the coefficient of variation can be calculated as $\sqrt{e^{\tau^2} - 1}$, where τ is the scale parameter for the lognormal distribution.

Conclusion

Acknowledgements

- RLS
- CBF divers
- ARC Discovery funding

References

1. Jennings, S., Pinnegar, J. K., Polunin, N. V. & Boon, T. W. [Weak cross-species relationships between body size and trophic level belie powerful size-based trophic structuring in fish communities](#). *Journal of Animal Ecology* 934–944 (2001).

2. Goatley, C. H. R. & Bellwood, D. R. Body size and mortality rates in coral reef fishes: A three-phase relationship. *Proceedings of the Royal Society B: Biological Sciences* **283**, 20161858 (2016).
3. Audzijonyte, A. *et al.* Fish body sizes change with temperature but not all species shrink with warming. *Nature Ecology & Evolution* **4**, 809–814 (2020).
4. Mangel, M. *et al.* A perspective on steepness, reference points, and stock assessment. *Canadian Journal of Fisheries and Aquatic Sciences* **70**, 930–940 (2013).
5. Froese, R. *et al.* A new approach for estimating stock status from length frequency data. *ICES Journal of Marine Science* **75**, 2004–2015 (2018).
6. Hordyk, A., Ono, K., Valencia, S., Loneragan, N. & Prince, J. A novel length-based empirical estimation method of spawning potential ratio (SPR), and tests of its performance, for small-scale, data-poor fisheries. *ICES Journal of Marine Science* **72**, 217–231 (2015).
7. Froese, R. *et al.* A new approach for estimating stock status from length frequency data. *ICES Journal of Marine Science* **75**, 2004–2015 (2018).
8. Clauset, A. & Erwin, D. H. The Evolution and Distribution of Species Body Size. *Science* **321**, 399–401 (2008).
9. Giometto, A., Altermatt, F., Carrara, F., Maritan, A. & Rinaldo, A. Scaling body size fluctuations. *Proceedings of the National Academy of Sciences* **110**, 4646–4650 (2013).
10. Rogers, A., Blanchard, J. L. & Mumby, P. J. Vulnerability of Coral Reef Fisheries to a Loss of Structural Complexity. *Current Biology* **24**, 1000–1005 (2014).
11. Edgar, G. J. & Stuart-Smith, R. D. Systematic global assessment of reef fish communities by the Reef Life Survey program. *Scientific Data* **1**, 140007 (2014).
12. Brandl, S. J. *et al.* Demographic dynamics of the smallest marine vertebrates fuel coral reef ecosystem functioning. *Science* **364**, 1189–1192 (2019).
13. R Core Team. *R: A language and environment for statistical computing*. (R Foundation for Statistical Computing, 2022).
14. Stan Development Team. *Stan modeling language users guide and reference manual, v2.21.0.* (2023).