Estimating the Unobserved: A Simulation Study on Censoring and Truncation in Race Models of Choice and Response Time

Jeroen E. Timmerman

Department of Psychology, University of Amsterdam

MI2324RM: Research Master's Internship

Dr. Andrew J. Heathcote

May 25th, 2024

DRAFT

Author Note

Jeroen E. Timmerman https://orcid.org/0009-0003-8208-0509

Correspondence concerning this article should be addressed to Jeroen E. Timmerman, Department of Psychology, University of Amsterdam, Nieuwe Achtergracht 129-B, Amsterdam, North-Holland 1018 WS, Email: j.e.timmerman@uva.nl

Abstract

Race models of decision making, such as the linear ballistic accumulator (LBA), log-normal race (LNR) and the racing diffusion model (RDM) are widely used to model response times (RTs) and choices as a race between choice options. These models assume the shapes of the RT distributions to reflect latent parameters, such as a participant's tendency to emphasize accuracy over speed. Since RT data often does not reflect the full RT distribution due to missing data from trial timeouts, outlier removal, or other limitations, which affects the parameter recovery for these models. While missing data is commonly handled by truncation—excluding the missing trials from the analysis—it can also be handled by censoring, which takes the proportion of missing trials into account. In two simulation studies, the parameter recovery for these methods are compared for the LBA, LNR, and RDM. In the first simulation study, we showed how upper truncation biases parameter estimation for all models, while censoring had good parameter identifiability with the large sample size used. In a second simulation with a smaller number of trials, censoring had better parameter recovery in some cases, particularly when the lower tail was missing, while in other cases the parameter recovery of censoring and truncation were not notably different. These results suggest that censoring should be preferred over truncation when the objective is accurate parameter estimation, but since censoring is computationally costly, truncation can be admissible in some analyses with a small number of trials.

Keywords: censoring, truncation, missing data, evidence accumulation, choice response data, diffusion decision model, linear ballistic accumulator, Bayesian hierarchical modeling. decision making

Estimating the Unobserved: A Simulation Study on Censoring and Truncation in Race Models of Choice and Response Time

Many paradigms in experimental psychology involve speeded decision-making. There are two main outcome variables to these tasks: what choice someone made (or whether this matches the corresponding stimulus), and how fast someone made their choice. Researchers are often interested in the conditions that affect these decisions and response times (RTs).

A problem arises when we want to make comparisons about performance on a task. Someone might be quicker to respond—indicating better performance, but at the same time they might be less accurate—indicating worse performance. This is commonly referred to as the speed accuracy trade-off, which complicates inferences on task performance.

A wide range of evidence accumulation models (EAMs) aim to model the cognitive processes behind decision making as noisy accumulation of evidence until a decision threshold is reached. This means that how fast a participant is able to accumulate evidence towards a choice (the evidence accumulation rate or drift rate) is modeled separately from a participants' tendency to value speed over accuracy or vice versa (the threshold or speed-accuracy tradeoff). In addition to these decisional variables, EAMs often estimate the time it takes for non-decisional processes like stimulus encoding or the motor response to occur (non-decision time), and they account for the variability within and between trials, as well as response bias.

This paper will focus on three prominent EAMs that are supported by the EMC2 package (Stevenson et al., 2024): the Linear Ballistic Accumulator (LBA; Brown & Heathcote, 2008), the Racing Diffusion Model (RDM; Tillman et al., 2020), and the Log-Normal Race (LNR; Heathcote & Love, 2012). These models are race models, which model separate accumulators for each choice option, with the first accumulator to reach a threshold resulting in the choice. The time that it takes for this accumulator to reach the threshold is the decision time.

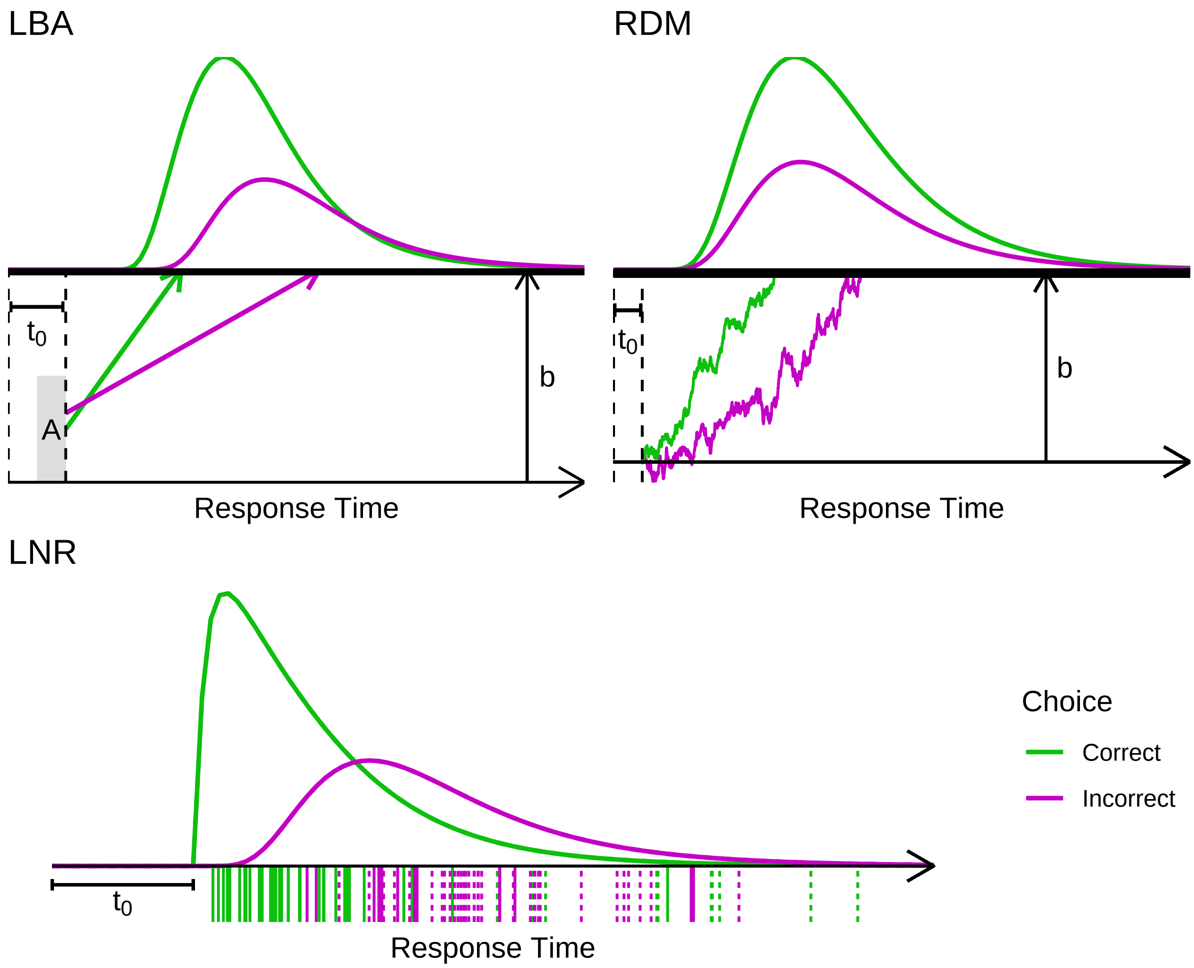
The LBA (Brown & Heathcote, 2008) models RTs and responses as a race between deterministic linear accumulators, with the slope for an accumulator drawn from a normal distribution, $\mathcal{N}(v, s_v^2)$, where v is the mean evidence accumulation rate and s_v^2 the between-trial variance of the evidence accumulation rate. The intercept also has between-trial variation and is drawn from a uniform distribution $\mathcal{U}(0, A)$. The distance from A to the threshold b is called B, and can be understood as caution, or the tradeoff between speed and accuracy. The time it takes for non-decsisional processes like motor response and stimulus encoding are estimated as a sigle non-decision time parameter, t_0. The line that intersects the threshold at the lowest time determines the decision made, and the timepoint of the intersection added to t_0 is the RT (Brown & Heathcote, 2008; see Figure 1).

The RDM (Tillman et al., 2020) has a similar set of parameters, but instead of having between-trial variation in evidence accumulation rate, the RDM has continuous normally distributed variation within each trial. In the simulation studies in this paper, starting point variability A is taken out as it is often not a necessary parameter for the RDM to account for common decisional phenomena (Tillman et al., 2020; see Figure 1).

Lastly, the LNR (Heathcote & Love, 2012) is a race between random samples from log-normal distributions, with the lowest sample determining the choice and decision time, which is then added to non-decision time t_0 to constitute the RT. This means that the LNR has three main parameters: the scale m of the lognormal, the shape s of the lognormal, and the non-decision time t_0. The LNR can not distinguish between the threshold and the accumulation rate, so it does not rest on latent assumptions about the accumulation process (Heathcote & Love, 2012; see Figure 1).

Figure 1

Simple Race Model Illustrations



Note. The dynamics of the race models in this paper are illustrated here. The Linear Ballistic Accumulator (LBA) accumulates evidence in a linear deterministic fashion, with normally distributed variability s_v^2 in slope v and uniformly distributed variability A in starting point between trials. The Racing Diffusion Model (RDM) has continuous normally distributed variance s_v^2 within the trial instead (diffusion), and can include starting point variability between trials like the LBA, although this is not the case for simulations in this paper. The Log-Normal Race (LNR) samples accumulation times (vertical lines) from log-normal distributions, with the shortest accumulation time (solid vertical lines) determining the choice and the decision time. For all models, the duration of non-decision processes like stimulus encoding and responding are captured by t_0.

Each of these race models have defined probability density functions $p(t \mid \boldsymbol{\theta})$ and cumulative density functions $P(t \mid \boldsymbol{\theta})$ for the finishing time of a single accumulator. To compute the likelihood of parameter vector $\boldsymbol{\theta}$ with winning accumulator i at time t, we can take the probability density for the winning accumulator $p(t \mid \boldsymbol{\theta}_i)$, and multiply it by the probability that none of the other accumulators j finish before time t:

$$\mathcal{L}(\boldsymbol{\theta} \mid t, i) = p(t \mid \boldsymbol{\theta}_i) \prod_{j \neq i} 1 - P(t \mid \boldsymbol{\theta}_j), \tag{1}$$

which is the "defective distribution" for i when written as a function of t, meaning that it integrates to the probability of response i.

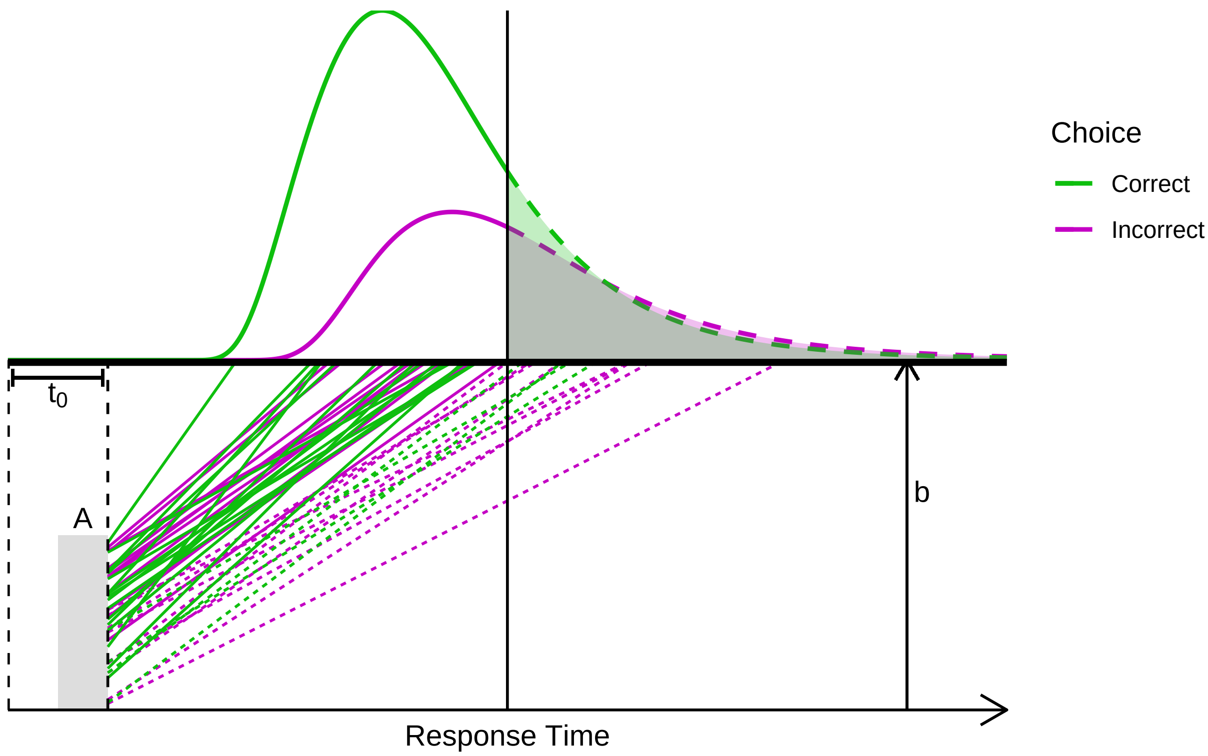
Complicating the estimation of speeded decision making models, RTs and choices are often missing on certain trials by design. A researcher may want to limit slow RTs in their experiment design, for example to reduce slow type II thinking (Evans, 2003), or to emphasise speed. Alternatively, researchers may want to remove outlying responses that cannot have come from the process of interest (e.g., a response 0.05 seconds after stimulus onset, which is too fast for a decision making process to occur).

There are two main ways to handle missing values: truncation, which discards missing values with no assumption of the underlying distribution, and censoring, which assumes the proportion of missing values to reflect the true distribution, and takes this into account in the model estimation.

Although truncation could potentially improve parameter estimates by eliminating irrelevant outliers, outlier removal often increases estimation bias by excluding extreme but valid RTs (Dolan et al., 2002; Miller, 2023; Ratcliff, 1993; Ulrich & Miller, 1994). Even less extreme values might be truncated in experiments with short time windows, leading to even worse estimates. Figure 2 illustrates how upper truncation might distort parameter estimation in a simple two forced choice decision task with slow errors. Trials with slower accumulation rates—which tend to result in more incorrect trials—get discarded, while trials with higher accumulation rates are used to estimate the underlying parameters.

Figure 2

Illustration of Missing Upper Response Times in the Linear Ballistic Accumulator



Note. This figure illustrates how upper censoring and truncation relate to the LBA and the defective density. Dashed lines represent the cognitive dynamics behind missing RT trials and the missing tails of the defective densities. Truncation discards the corresponding data completely, while censoring uses the shaded areas under the defective density curves when computing the likelihood of the parameter estimates.

Despite the issues, researchers often use truncation for missing RTs, perhaps because of common practice, or because computing the likelihood for a censored RT requires computing the area under the curve of the likelihood function over the censored range. For race models, this means that we integrate Equation 1 over the censored time range R:

$$\mathcal{L}(\boldsymbol{\theta} \mid t \in R, i) = \int_R p(t \mid \boldsymbol{\theta}_i) \prod_{j \neq i} 1 - P(t \mid \boldsymbol{\theta}_j)\, dt. \tag{2}$$

Since censoring in maximum likelihood estimation and Markov chain Monte Carlo methods require numerically computing this integral for many iterations, censoring is often slow.

Equation 2 can be extended to accomodate missing responses in addition to missing RTs by summing Equation 2 for each accumulator, resulting in the total probability of any response in time range R. Similarly, if we censored both fast and slow responses with no distinction between the two, we can sum the integrals over the lower and the upper range. We can even combine censoring and truncation with

$$\mathcal{L}(\boldsymbol{\theta} \mid t \in R, i) \frac{\mathcal{L}(\boldsymbol{\theta} \mid 0 \leq t < \infty, i)}{\mathcal{L}(\boldsymbol{\theta} \mid t \in S, i)}, \tag{3}$$

where S is the range of all untruncated values.

Although censoring has been shown to result in better parameter recovery than truncation in common RT distributions (Dolan et al., 2002), the difference in parameter recovery has not yet been established for race models. As data are routinely censored or truncated, this study aims to compare parameter recovery for different levels of censoring and truncation for the LBA, LNR, and RDM. We will examine differences in sensitivity to censoring and truncation between the models, and to what extent different levels of censoring and/or truncation cause estimation bias and/or imprecision in designs with varying numbers of trials. Parameter estimates are expected to deteriorate at a higher rate for truncation than for censoring with increases in the proportion of missing RTs.

Methods

We compared censoring and truncation in two simulation studies. The first study compared upper censoring and truncation with the responses known with a large number of samples to assess asymptotic parameter identifiability. The second study assessed parameter recovery on a smaller number of trials to evaluate the practical differences between censoring and truncation, comparing a larger number of missing data scenarios.

To assess parameter identifiability, we first simulated data using known parameters, which we then fit the same model to. This allows us to compare the known, “true” parameter values with our fitted parameter values. The code for all analyses and data generation are available on https://github.com/timmerj1/censoring-truncation-study-EAMs.

For both simulation studies, a simple model with two stimuli and two racers was used to generate the data. Conventional constants for the LBA's ($s_v = 1$) and the RDM's drift rate variance ($s = 1$) were used to generate and fit the data. Parameter values were chosen to reflect common RT and choice distributions in simple two forced choice tasks (see Table 1 for an overview of parameters used). For the LBA, decision thresholds were defined with $B = b - A$. As is the default in EMC2, parameters on the positive real line were log transformed. Data was generated using the `make_data` function from EMC2 (Stevenson et al., 2024).

Table 1

Race Model Simulation Parameters

LBA		LNR		RDM	
Parameter	Value	Parameter	Value	Parameter	Value
B	2	m	0.75	B	3
v	3	m_{true}	0.65	v	1
v_{true}	1	s	0.5	v_{true}	4
A	2	s_{true}	0.8	s_{true}	0.75
$s_{v_{true}}$	0.75	t_0	0.4	t_0	0.2
B	2				

Note. The parameters that were used to simulate the response time and choice data. Parameters on the real positive line are estimated on a log scale in EMC2. Subscript 'true' relates to a match between stimulus and choice (a correct choice is made), and these parameters are added to the corresponding parameter for the non-matching racer.

For the first simulation, upper censoring and truncation were compared using a large number of trials (20,000 trials, 10,000 per stimulus) to investigate the differences between censoring and truncation without fits being affected by random error. RTs were cut off at three different levels: at 2.5%, 10%, and 30% of upper values, with cutoff points estimated in a separate simulation with the same number of trials. Responses were not missing for any of the censored values. For each combination of model and missing level, new data was simulated, but censoring and truncation were compared on the same datasets to ensure differences cannot be caused by random sampling error.

In the second simulation, a factorial design was used to vary whether responses were known or unknown, which tail missed response times (lower, upper, or both tails), the percentage of missing responses (2%, 10%, 30%, or 50%), and whether missing responses were censored or truncated. For each condition, ten samples with a small number of trials (400 trials) were simulated and fit. Missing level cutoffs were chosen using the quantiles on a non-missing simulation with 4000 trials.

Parameter posteriors were sampled using the particle Metropolis within Gibbs (PMwG; Kuhne et al., 2024) sampler in EMC2 (Stevenson et al., 2024) using its default (standard normal) priors for non-hierarchical estimation. Three PMwG chains were sampled for each parameter, with 50 particles per parameter (Kuhne et al., 2024).

To assess parameter recovery, we used the Root Mean Squared Errors (RMSEs) between the posterior medians $\hat{\boldsymbol{\theta}}$ and the true parameters $\boldsymbol{\theta}$:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}_i - \theta_i)^2}. \tag{4}$$

Since we used Bayesian methods, the parameter estimates are not restricted to point estimates. We will compute the quadratic Wasserstein distance between the full posterior P samples $X_1, ..., X_n$ and Dirac point mass distribution $\delta_{\boldsymbol{\theta}}$ at the true parameters $\boldsymbol{\theta}$:

$$W_2(\delta_{\boldsymbol{\theta}}, P) = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\|X_i - \boldsymbol{\theta}\|^2}. \tag{5}$$

The Wasserstein distance was taken for the full model using the Euclidean distance, as well as for the separate parameters with unidimensional distance. The quadratic Wasserstein distance simplifies to the RMSE if point estimates were used. Moreover, the quadratic Wasserstein over a single dimension becomes an unbiased estimate of the standard deviation when θ is the mean of P. This makes the quadratic Wasserstein distance an interesting Bayesian alternative to the RMSE.

For completeness, additional Pearson's correlation coefficients and mean absolute errors were computed, but since these did not differ substantially from the RMSE and W_2, these statistics were plotted in the appendix. Parameters that were log transformed to be estimated on the real scale were not transformed back when computing these statistics.

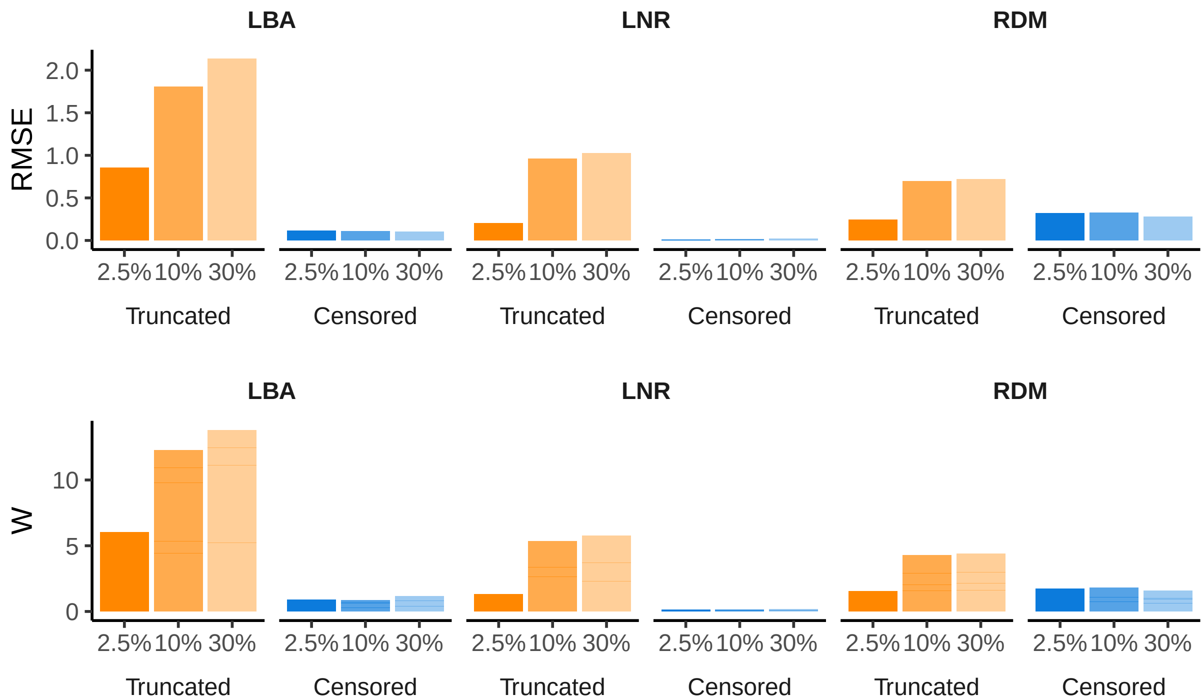
Results

Study 1

Figure 3 shows the RMSE between the true parameter values and the medians of the posterior distribution for each fit. As expected, both distance measures RMSE and W_2 increased with an increase of missing RTs for truncation. For censoring, the distance measures stay closer to 0 and do not clearly increase when data is censored rather than truncated. This indicates that censoring improved the parameter recovery in an asymptotic fit as expected. The only case where censoring does not seem to outperform truncation is for the RDM, where 2.5% truncation did not perform worse than 2.5% censoring. With higher missing percentages, censoring still outperformed truncation for the RDM.

Figure 3

RMSE and W_2 for Upper Censoring

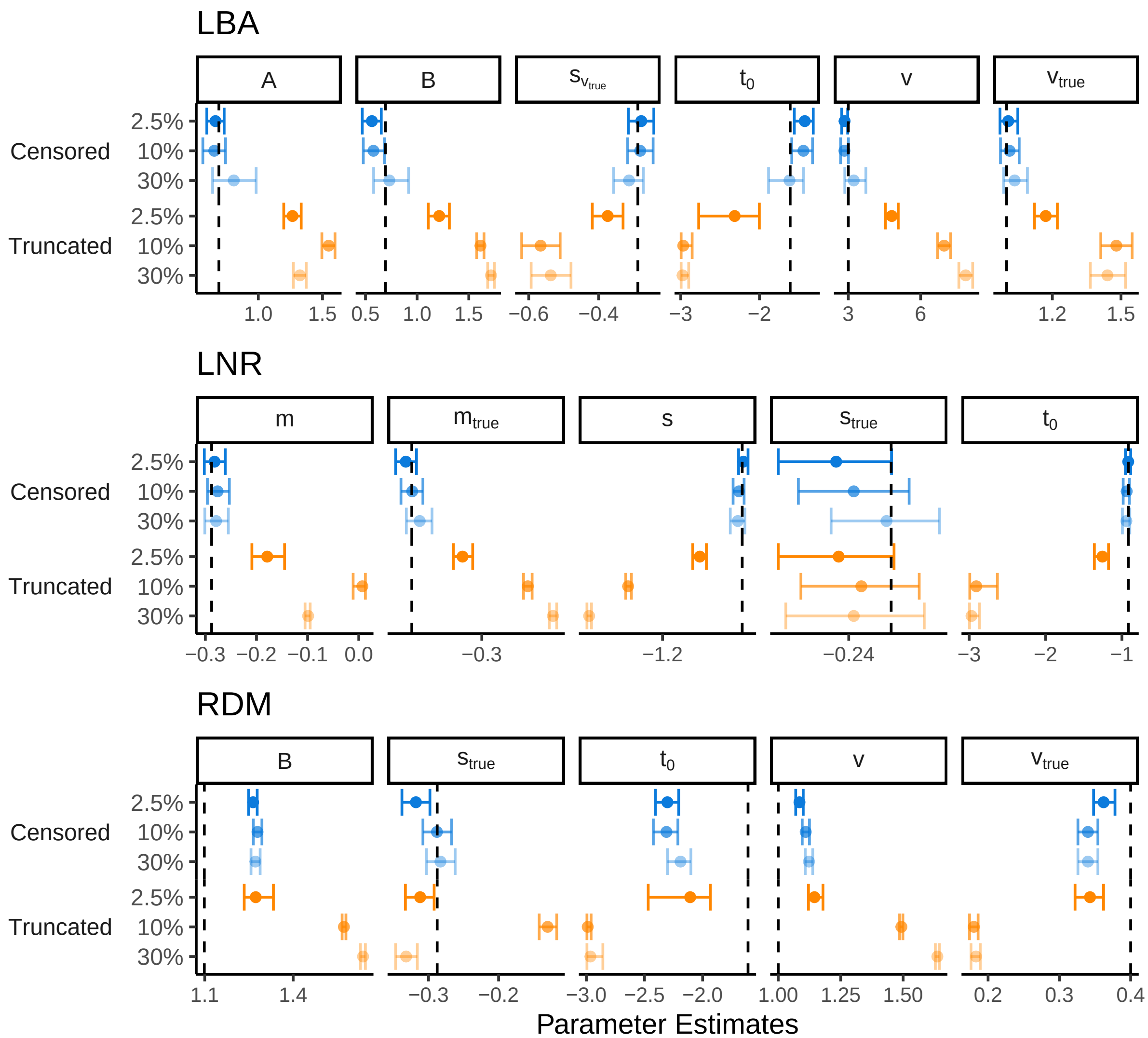


Note. For each model and each level of missingness, the root mean squared errors (RMSE) and quadratic Wasserstein distances (W_2) were computed and plotted. Lower RMSE and W_2 indicate better parameter recovery.

Looking at the credible intervals for each parameter in Figure 4, the generally higher RMSE for RDMs is explained by a general tendency for B and v to be overestimated, while v_{win} and t_0 are underestimated. Overall, these parameters still seem to be recovered better with censoring than with truncation, except that t_0 and s_{win} were recovered slightly better for truncation at a low percentage in this simulation. For the other models, censoring clearly performs better than truncation, with true parameters included by most credible intervals for censoring, while most truncation credible intervals exclude the true parameters. The main exception to this is the s_{win} parameter for the LNR, where the credible interval for 30% censoring does not include the true parameter value.

Figure 4

Parameter Recoveries and 95% CIs for Upper Censoring and Truncation



Note. Race model recoveries for the linear ballistic accumulator model (LBA), the log-normal race model (LNR) and the racing diffusion model (RDM) with 2%, 10%, or 30% missing. Dots are placed at the posterior median, and error bars denote the 95% equal-tailed credible intervals of the parameter posteriors. The dashed vertical lines represent the true parameter values.

Overall, these results indicate that upper censoring results in better parameter recovery than upper truncation in an asymptotic sample, i.e. with a large number of trials. This cannot yet be generalized to smaller sample sizes, as random error and parameter tradeoffs might affect parameter recovery more than censoring versus truncation does. Moreover, one might want to use lower censoring or truncation instead, or a combination of lower and upper censoring or truncation. Lastly, in this simulation the responses were known. Often censoring or truncation is implemented when there is a response time window, where neither the RTs or the choices are recorded. To account for these issues, the second simulation study takes these factors into account by using a smaller number of trials (200 instead of 1000 trials), in a $2 \times 2 \times 3 \times 3 \times 4$ design: censoring versus truncation, known versus unknown, lower versus upper versus both tails missing, LBA versus LNR versus RDM, and 2%, 10%, 30%, and 50% censoring or truncation.

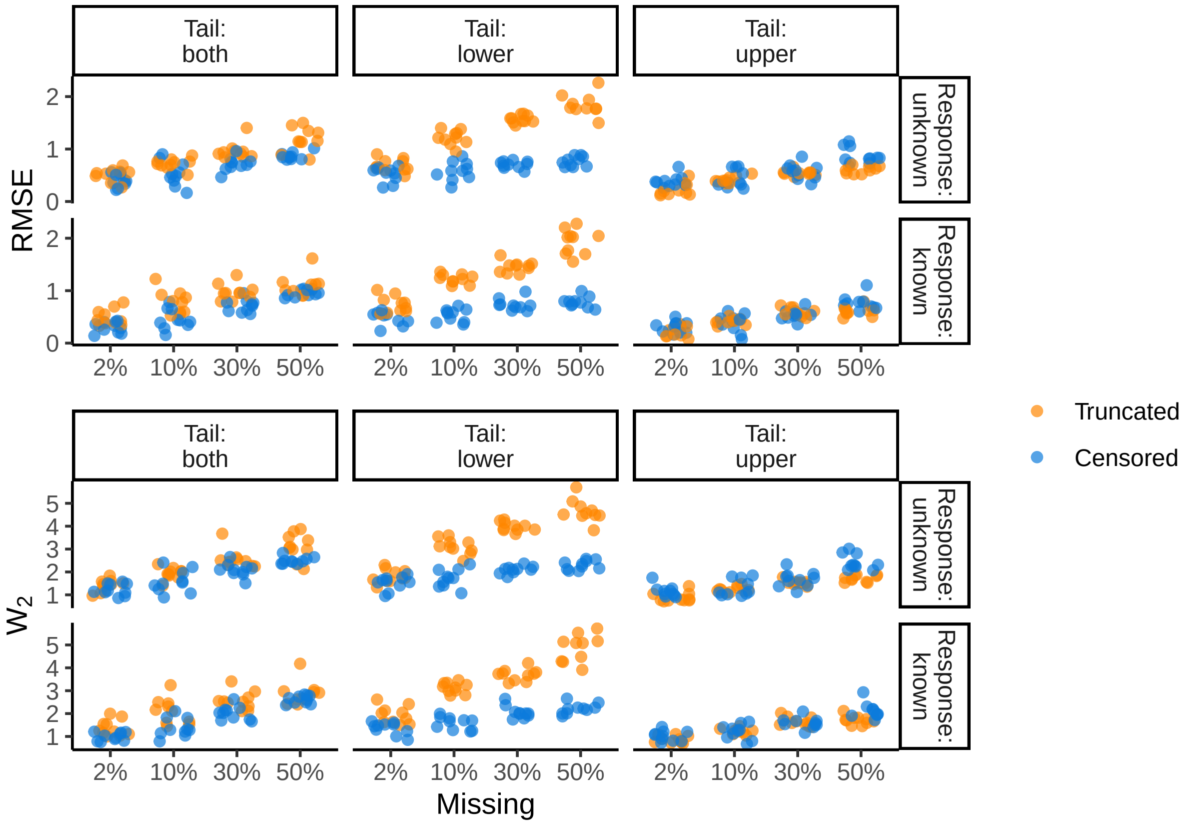
Study 2

Linear Ballistic Accumulator Model

Contrary to the first study, upper censoring with responses known did not have better parameter recovery than upper truncation for the LBA. Figure 5 shows similar RMSEs and W_2s for all upper censoring and truncation, both getting worse with higher missing percentages. Contrary to this, fits for the lower tail clearly diverge with increasing missing percentages, with censoring outperforming truncation. Interestingly, upper tail censoring and truncation both seem to perform on par with lower tail censoring, with relatively low RMSE and W_2. Censoring on both tails resulted in similar RMSE and W_2 compared to truncation, but censored RMSEs and W_2s were generally lower.

Figure 5

Model RMSE and W_2 for the LBA

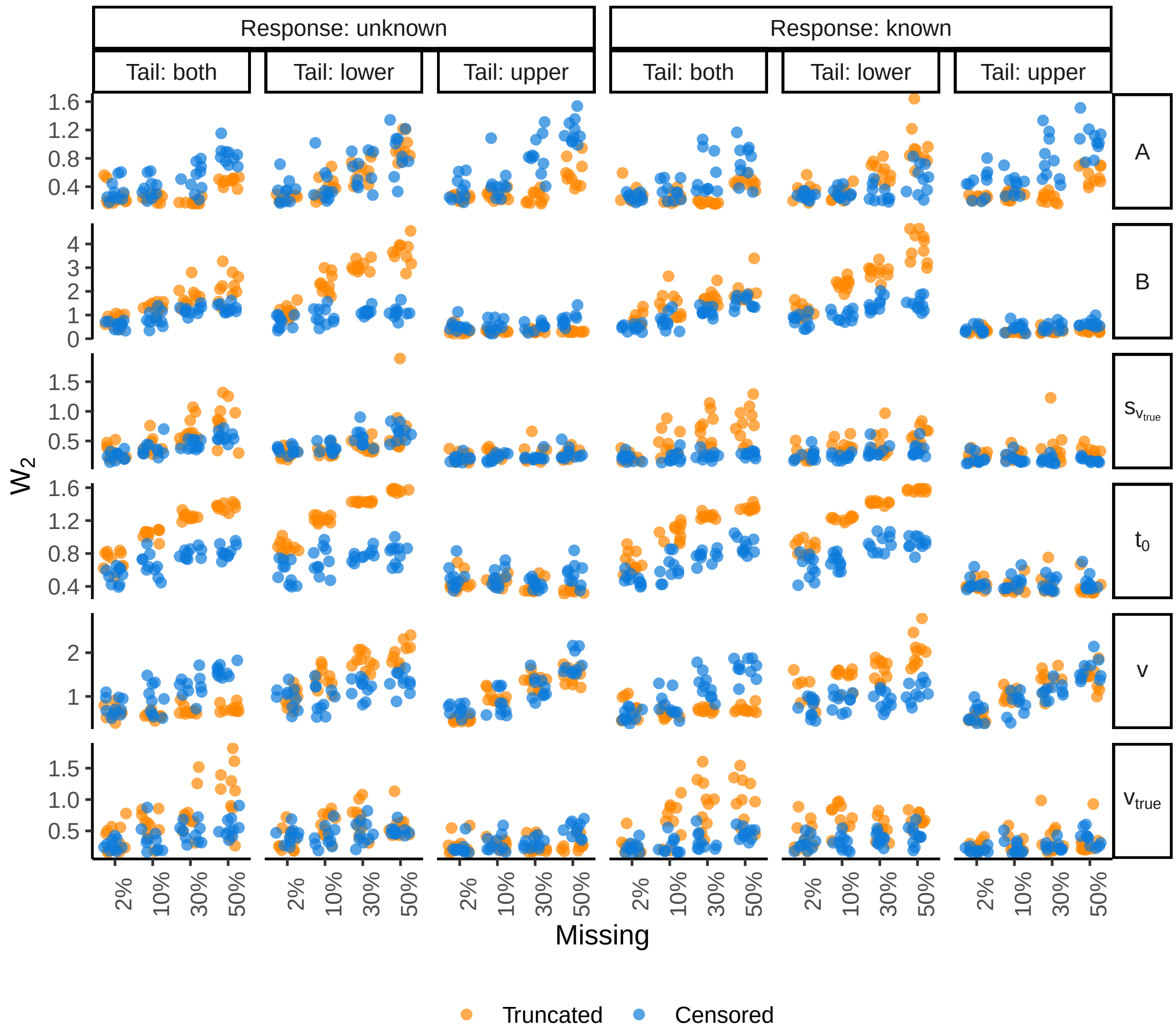


Note. Root mean squared errors (RMSE) and quadratic Wasserstein distances (W_2) over all parameters for the linear ballistic accumulator model (LBA). Colors denote whether missing RTs were censored or truncated. Points were randomly horizontally shifted to avoid overlap. Lower RMSE and W_2 indicate better parameter recovery.

Figure 6 shows the W_2 for each LBA parameter. In accordance with Figure 5, lower censoring resulted in similar or lower W_2s than lower truncation. Especially the boundary B, non-decision time t_0, and mean evidence accumulation rate v showed better parameter recovery for censoring compared to truncation when responses are unknown, while lower censoring with responses known improved parameter recovery for all parameters. With both tails missing, censoring showed better parameter recovery than truncation for most parameters, but worse recovery for starting point variability A and mean evidence accumulation rate v, explaining the similar RMSE and W_2 in Figure 5. Lastly, when the upper tail was missing, truncation recovered the parameters similarly to censoring, with truncation showing better parameter recovery of starting point variability A.

Figure 6

W_2 by Parameter for the LBA



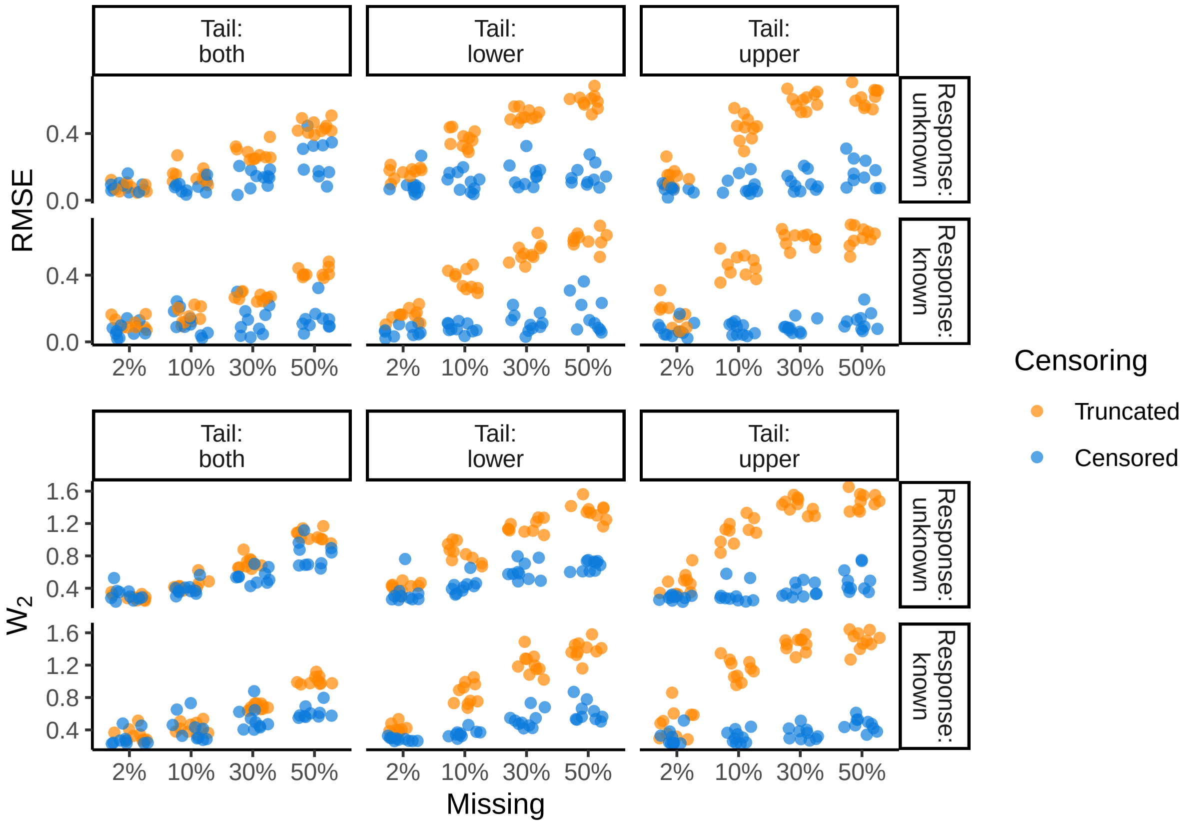
Note. Quadratic Wasserstein distances (W_2) between the true linear ballistic accumulator (LBA) parameters and each parameter posterior for each dataset. Each LBA parameter is represented by a different plot row labeled on the right, while each combination of tail and response condition is represented by the columns. Colors denote whether missing RTs were censored or truncated. Points were randomly horizontally shifted to avoid overlap. Lower W_2 indicates better parameter recovery.

Log-Normal Race Model

For the LNR, both upper and lower censoring showed better parameter recovery as expected. For the lower and upper tail simulation, Figure 7 shows the expected pattern of increasing RMSEs and W_2s for truncation while censoring RMSEs and W_2s remain low. Unexpectedly, missing values in both tails did not result in the same clear pattern, especially for W_2. This difference between RMSEs and W_2 indicate that although the medians of censoring posterior distributions were closer to the true parameter values, the complete posteriors had similar distances to the true parameter values for censoring compared to truncation.

Figure 7

Model RMSE and W_2 for the LNR

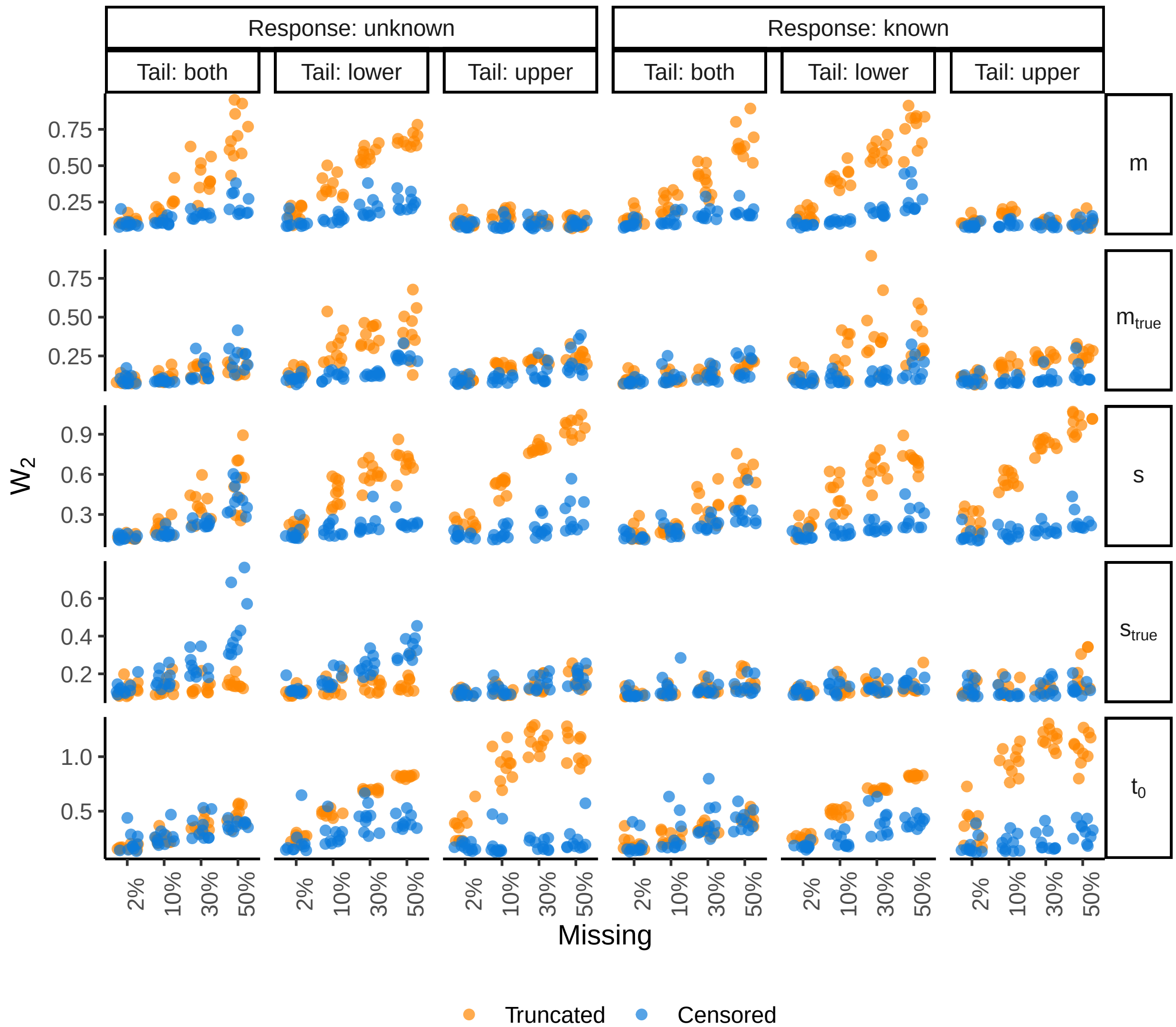


Note. Root mean squared errors (RMSE) and quadratic Wasserstein distances (W_2) over all parameters for the log-normal race model (LNR). Colors denote whether missing RTs were censored or truncated. Points were randomly horizontally shifted to avoid overlap. Lower RMSE and W_2 indicate better parameter recovery.

Looking at the parameter W_2s in Figure 8, we see that lower truncation particularly deteriorated the estimates for the log-normal μ estimates m and m_{true}, and non-decision time t_0 compared to lower censoring, whereas upper truncation deteriorated the estimates for log-normal σ estimate s and t_0 compared to upper censoring. Surprisingly, W_2s for t_0 estimates were worse for upper truncation than lower truncation, even though non-decision time is largely reflected by the minimum RT. When responses were unknown, lower and two-tailed censoring had worse parameter recovery for s_{true} than truncation. Two-tailed censoring recovered m and s better than two-tailed truncation, but resulted in recoveries that were similar to truncation for the other parameters.

Figure 8

W_2 by Parameter for the LNR



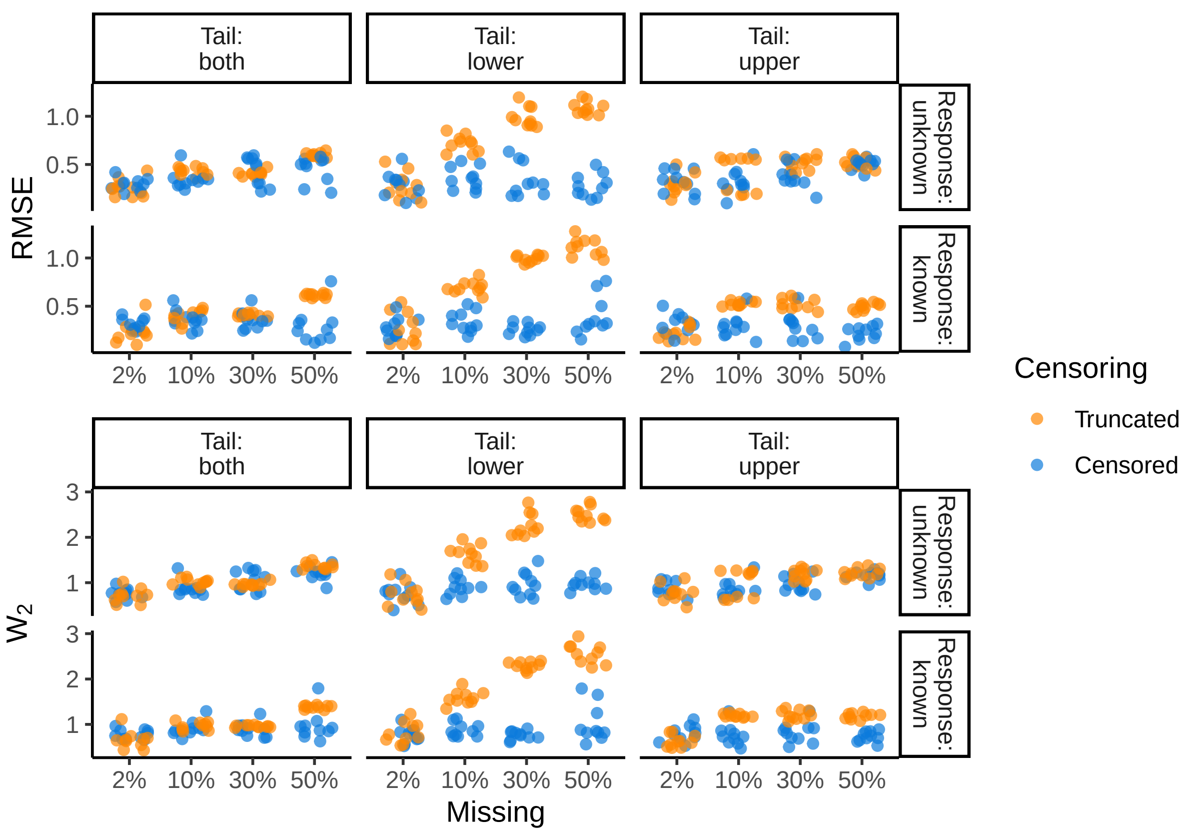
Note. Quadratic Wasserstein distances (W_2) between the true log-normal race (LNR) parameters and each parameter posterior for each dataset. Each LNR parameter is represented by a different plot row labeled on the right, while each combination of tail and response condition is represented by the columns. Colors denote whether missing RTs were censored or truncated. Points were randomly horizontally shifted to avoid overlap. Lower W_2 indicates better parameter recovery.

Racing Diffusion Model

Like the LBA, Figure 9 shows higher RMSEs and W_2s for lower truncation compared to lower censoring, while upper and two-tailed censoring show similar, lower RMSEs and W_2s compared to truncation. Only when responses were known, upper tail censoring resulted in lower RMSE and W_2 than truncation, but without a clear separation.

Figure 9

Model RMSE and W_2 for the RDM

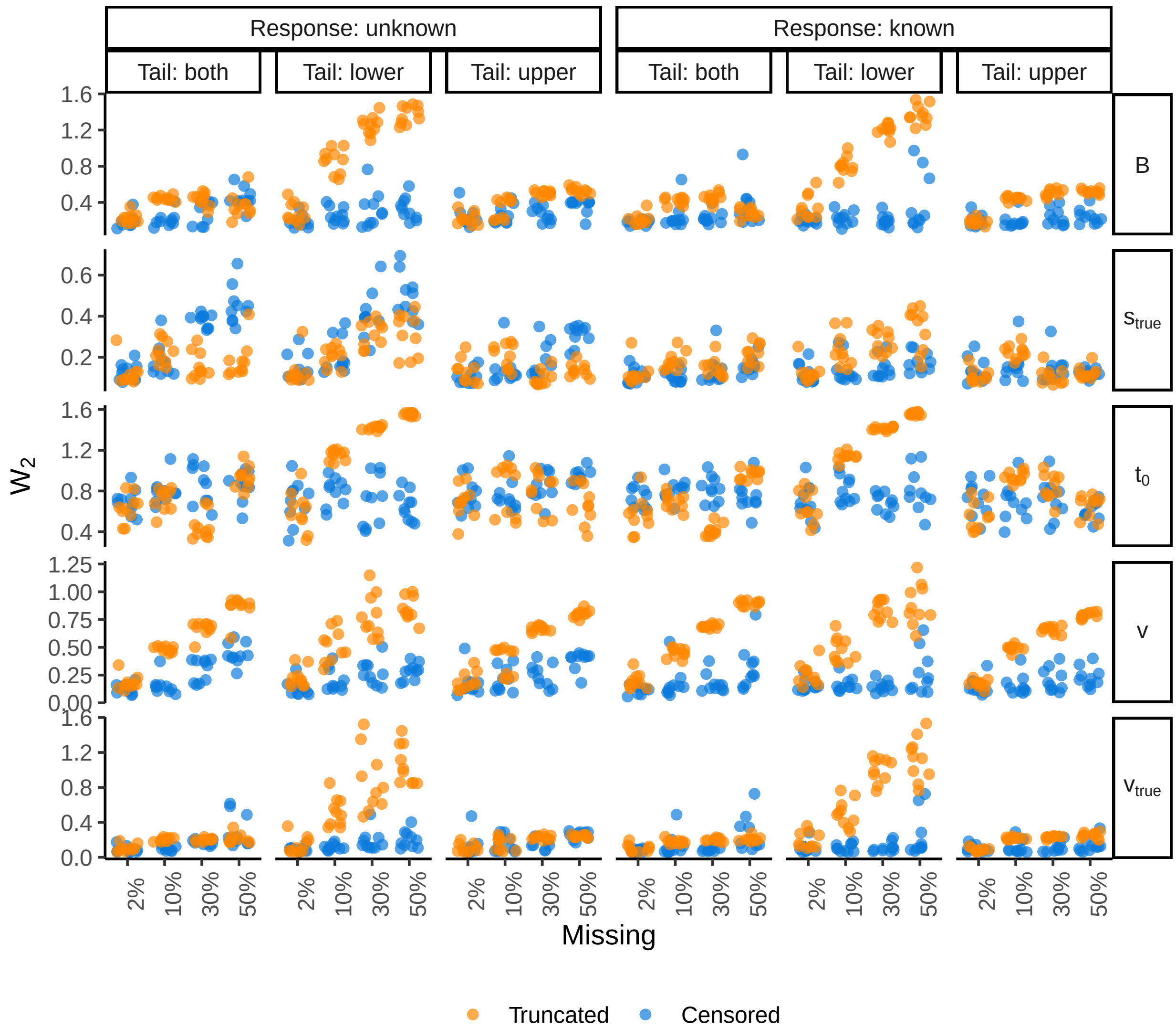


Note. Root mean squared errors (RMSE) and quadratic Wasserstein distances (W_2) over all parameters for the racing diffusion model (RDM). Colors denote whether the missing RTs were censored or truncated. Points were randomly horizontally shifted to avoid overlap. Lower RMSE and W_2 indicate better parameter recovery.

The parameter W_2s shown in Figure 10 also showed a similar pattern to the LBA. Lower tail truncation mainly affected the boundary parameter B, the non-decision time t_0, and the drift rates v and v_{true} compared to lower censoring. The W^2s for the drift rate of the incorrect option, v, was increasing for truncation compared to censoring, regardless of the tail or whether responses were known. For the other parameters, upper tail and two-tailed censoring was not particularly better or worse than truncation, with an exception of the recovery of B and v_{true} for upper tail censoring with responses known.

Figure 10

W_2 by Parameter for the RDM



Note. Quadratic Wasserstein distances (W_2) between the true racing diffusion model (RDM) parameters and each parameter posterior for each dataset. Each RDM parameter is represented by a different plot row labeled on the right, while each combination of tail and response condition is represented by the columns. Colors denote whether missing RTs were censored or truncated. Points were randomly horizontally shifted to avoid overlap. Lower W_2 indicates better parameter recovery.

Discussion

We compared censoring and truncation for three different race models: the linear ballistic accumulator model, the log-normal race model, and the racing diffusion model. To this end, we simulated data using pre-specified parameter values, with missing data cutoffs based on quantiles. The first simulation study compared asymptotic parameter recovery upper censoring and truncation with responses known. The second simulation used several smaller datasets for each condition in a factorial design. It compared lower, upper and two-tailed censoring or truncation, with either results known or unknown.

In the first simulation, parameters were recovered well for censoring, while parameters estimated with truncation were considerably off, even at lower percentages of missingness. Only for the RDM, 2.5% truncation was on par with censoring. The RDM had worse parameter recovery than the other two models in general, with credible intervals for some parameters missing the true parameter for each level of censoring or truncation. Censoring still outperformed truncation in the RDM for higher percentages of missing RTs.

The second simulation showed that the previous results did not hold with smaller sample sizes for the LBA and RDM. However, lower truncation did result in worse parameter recovery than censoring. For the LNR, upper and lower truncation both resulted in better recovery than truncation, but two-tailed truncation was closer to censoring in terms of parameter recovery.

The results indicate that censoring improves parameter recovery asymptotically, but in smaller numbers of trials, differences between censoring and truncation can become inconsequential compared to random sampling error.

Since parameter recovery for censoring was not worse than for truncation, censoring should be preferred if parameter recovery is the only consideration. When computation time and resources are relevant factors, however, censoring might not always be worth the cost. When responses were unknown, resulting in a higher number of numerical integrals for each likelihood calculation, the PMwG sampler took more than 28 hours in an extreme case in the second simulation. In less extreme cases with responses known, this time was closer to 15 minutes.

To speed up the MCMC sampling for censoring, the likelihood function for censoring was translated from R to Rcpp. This approximately doubled the speed. In the Rcpp likelihood function, numerical integration was implemented with RcppNumerical Qiu et al. (2023), an Rcpp library for numerical integration and optimization. Although this results in faster likelihood estimations than with the `integrate` function in R, the computation time is still considerable. To make censoring more worthwhile, changing to a faster quadrature rule that does not have to work for complicated multimodal functions might be beneficial.

The simulations in this paper all used non-hierarchical models. Since the second simulation seems to have been more affected by random error than by truncation, future simulations might use hierarchical modeling, which is less affected by random error for single participants, but instead shrinks more extreme values to group level means (Efron & Morris, 1977), making the inference more robust against random sampling error. Hierarchical models are increasingly used and it remains unclear how censoring and truncation affects hierarchical model estimation.

Another avenue that remains uninvestigated is how contamination ties into censoring and truncation. Since truncation can remove outliers, truncation could potentially lead to better parameter recovery than censoring would in the presence of contaminant RTs. On the other hand, MCMC methods also allow for the explicit modeling of contaminants. The likelihood function that was implemented in EMC2 allows for combinations of censoring, truncation, and explicit contaminant modeling, and future simulations could investigate how combinations of these methods might lead to better parameter recovery.

How researchers handle outliers and missing data can greatly affect the outcomes of their research. Our results suggest that ignoring values outside of a prespecified response window or outside of common outlier thresholds can bias race model estimates. Although censoring can be computationally costly, it should be the default over truncation when the objective is accurate estimation of latent cognitive parameters.

References

Brown, S. D., & Heathcote, A. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*(3), 153–178. https://doi.org/10.1016/j.cogpsych.2007.12.002

Dolan, C. V., Van der Maas, H. L., & Molenaar, P. C. (2002). A framework for ML estimation of parameters of (mixtures of) common reaction time distributions given optional truncation or censoring. *Behavior Research Methods, Instruments, & Computers*, *34*, 304–323. https://doi.org/10.3758/BF03195458

Efron, B., & Morris, C. (1977). Stein's paradox in statistics. *Scientific American*, *236*(5), 119–127.

Evans, J. St. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences*, *7*(10), 454–459. https://doi.org/10.1016/j.tics.2003.08.012

Heathcote, A., & Love, J. (2012). Linear deterministic accumulator models of simple choice. *Frontiers in Psychology*, *3*. https://doi.org/10.3389/fpsyg.2012.00292

Kuhne, C., Innes, R. J., Cooper, G., Stevenson, N., Cavallaro, J.-P., Brown, S., & Hawkins, G. (2024). *Hierarchical bayesian estimation for cognitive models using particle metropolis within gibbs (PMwG): A tutorial*. PsyArXiv. https://doi.org/10.31234/osf.io/xr37a

Miller, J. (2023). Outlier exclusion procedures for reaction time analysis: The cures are generally worse than the disease. *Journal of Experimental Psychology: General*. https://doi.org/10.1037/xge0001450

Qiu, Y., Balan, S., Beall, M., Sauder, M., Okazaki, N., & Hahn, T. (2023). *RcppNumerical: 'Rcpp' integration for numerical computing libraries*. https://CRAN.R-project.org/package=RcppNumerical

Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin*, *114*(3), 510. https://doi.org/10.1037/0033-2909.114.3.510

Stevenson, N., Donzallaz, M. C., Innes, R. J., Forstmann, B., Matzke, D., & Heathcote, P., Andrew. (2024). *EMC2: An r package for cognitive models of choice*.

https://doi.org/10.31234/osf.io/2e4dq

Tillman, G., Van Zandt, T., & Logan, G. D. (2020). Sequential sampling models without random between-trial variability: The racing diffusion model of speeded decision making. *Psychonomic Bulletin & Review*, *27*(5), 911–936. https://doi.org/10.3758/s13423-020-01719-6

Ulrich, R., & Miller, J. (1994). Effects of truncation on reaction time analysis. *Journal of Experimental Psychology: General*, *123*(1), 34. https://doi.org/10.1037//0096-3445.123.1.34

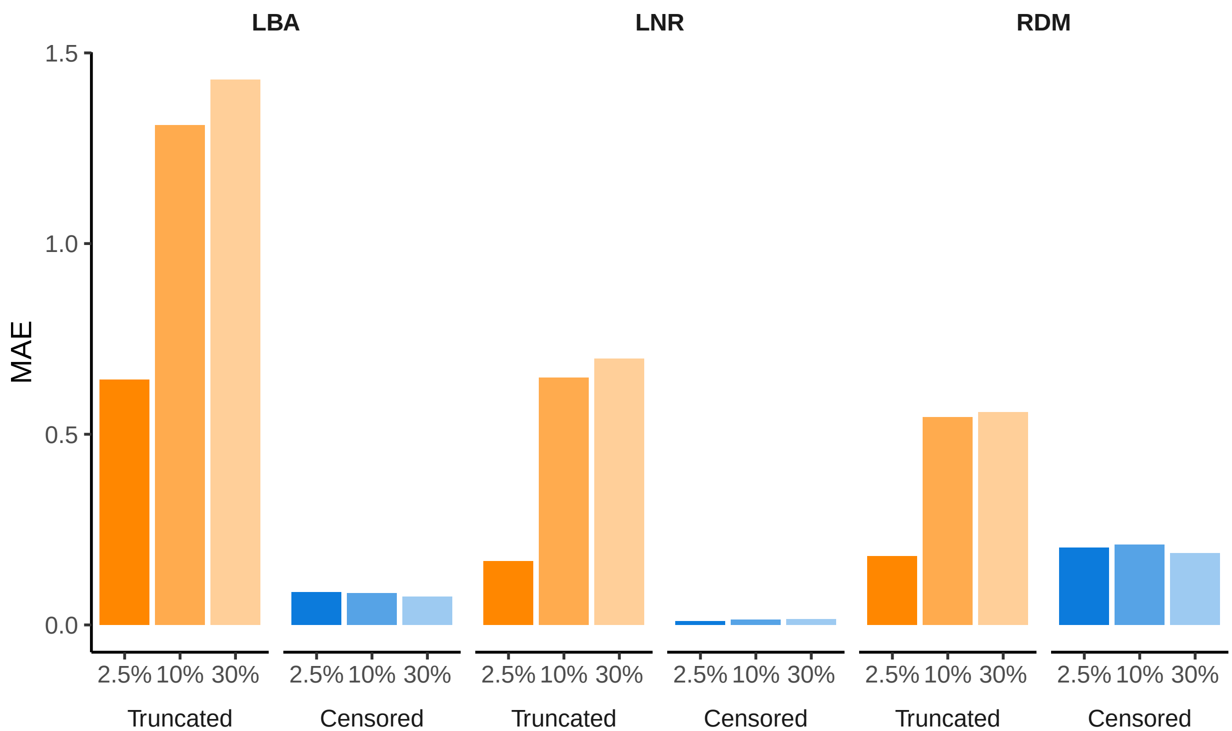
Appendix

Supplementary Materials

Study 1

Figure A1

MAE for Upper Censoring



Note. For each model and each level of missingness, the mean absolute error (MAE) is plotted. Lower MAE indicates better parameter recovery.

Study 2

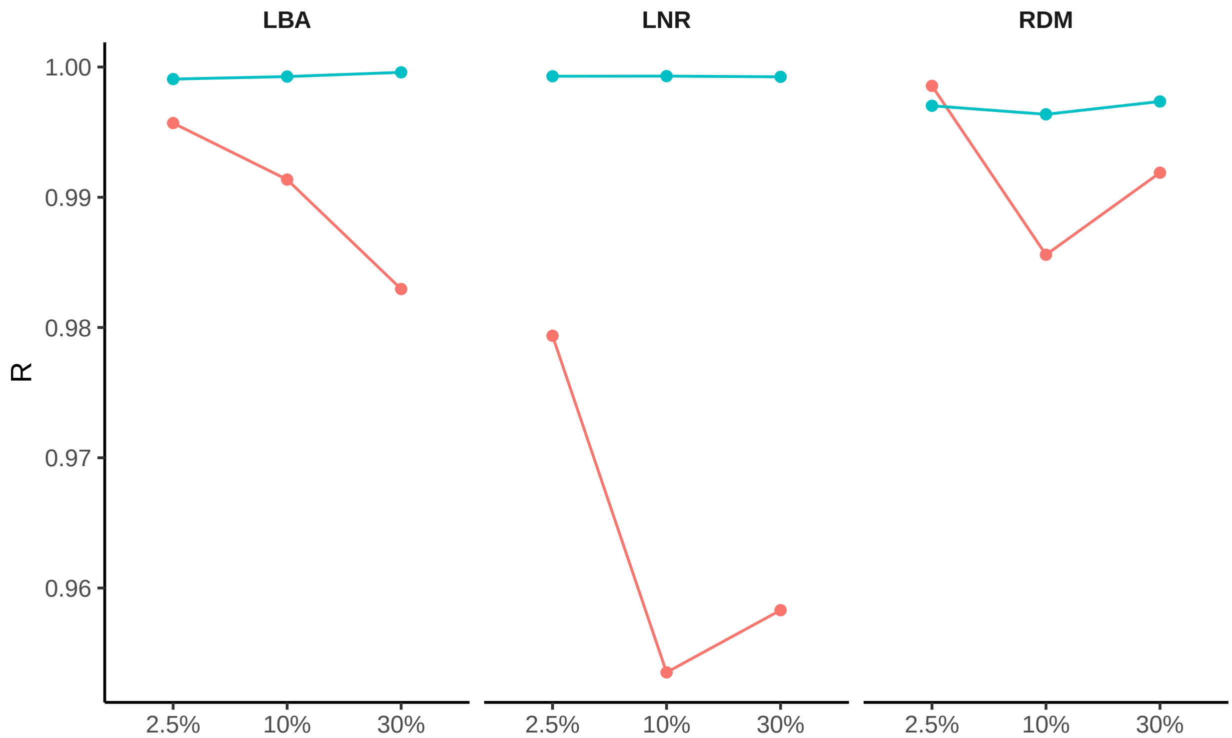
Linear Ballistic Accumulator Model

Model Distances.

Credible Intervals.

Figure A2

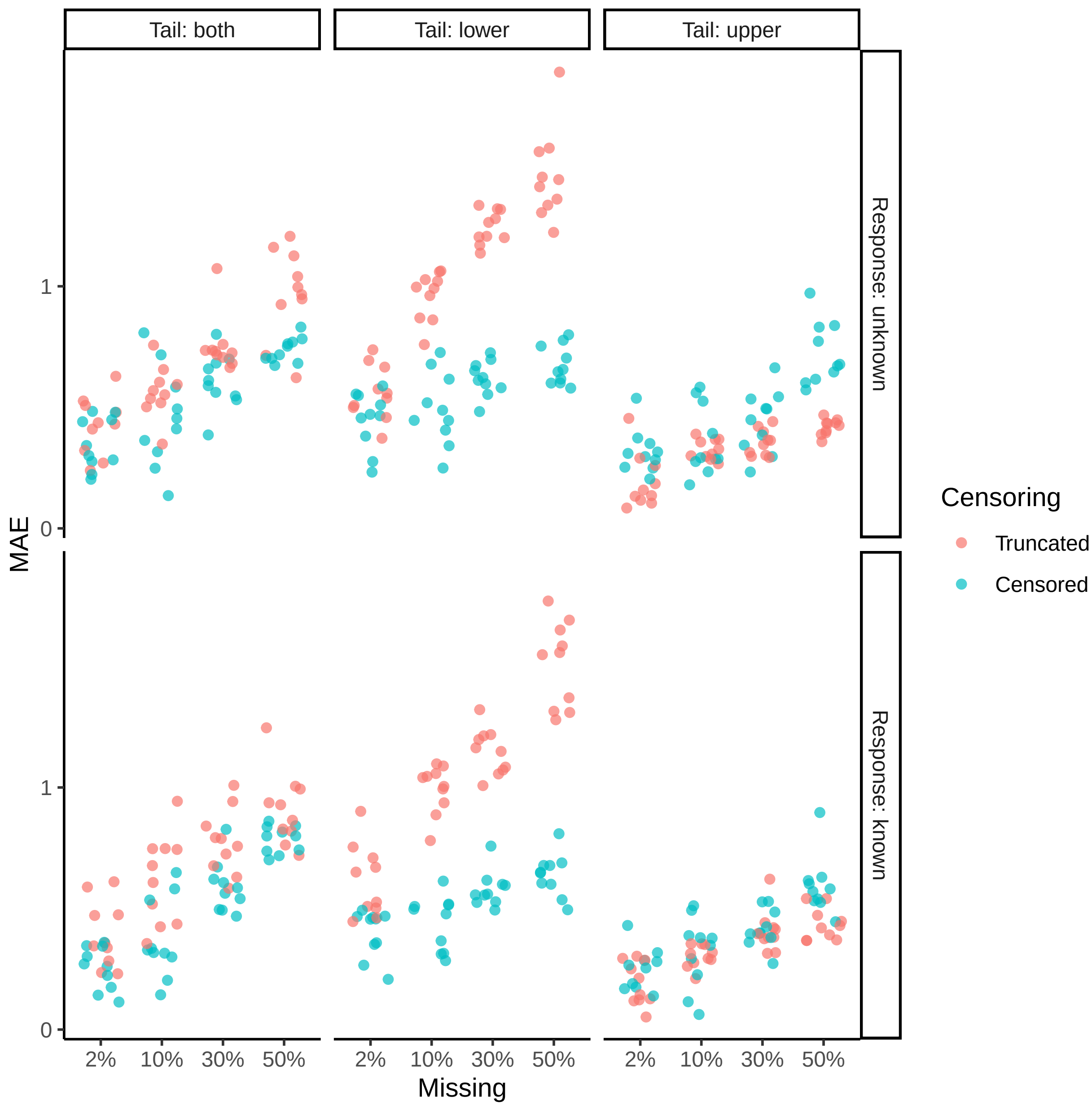
Pearson Correlations for Upper Censoring



Note. For each model and each level of missingness, the Pearson correlation (R) is plotted. Higher R indicates better parameter recovery.

Figure A3

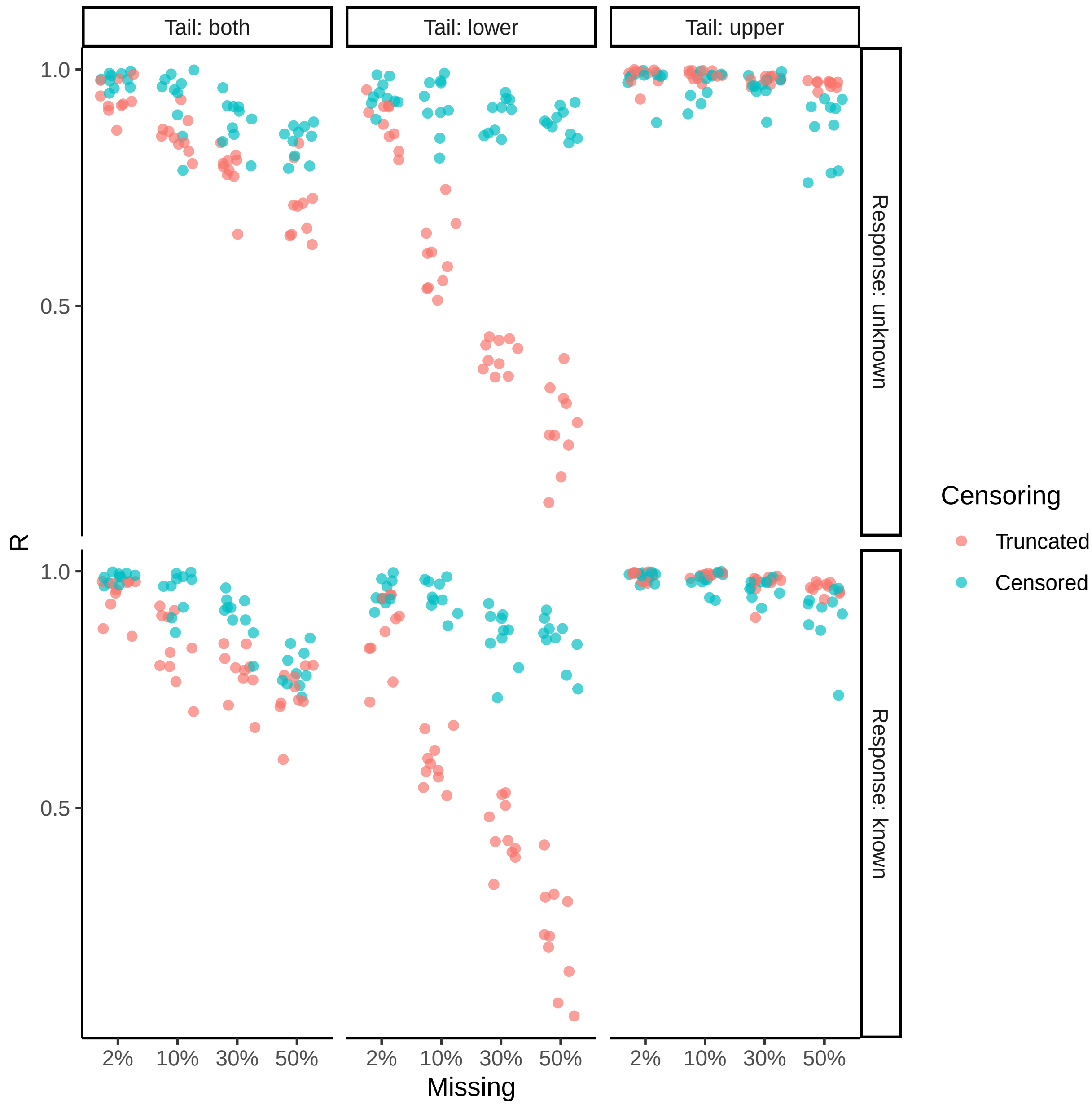
MAEs for the LBA



Note. Mean absolute errors over all parameters for the linear ballistic accumulator model (LBA). Colors denote whether missing RTs were censored or truncated. Points were randomly horizontally shifted to avoid overlap. Lower MAE indicates better parameter recovery.

Figure A4

Pearson correlations for the LBA



Note. Pearson correlations (R) between the true linear ballistic accumulator model (LBA) parameters and the posterior medians. Colors denote whether missing RTs were censored or truncated. Points were randomly horizontally shifted to avoid overlap. High correlation indicates better parameter recovery.

Figure A5

Linear Ballistic Accumulator Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Upper Tail and Responses Known

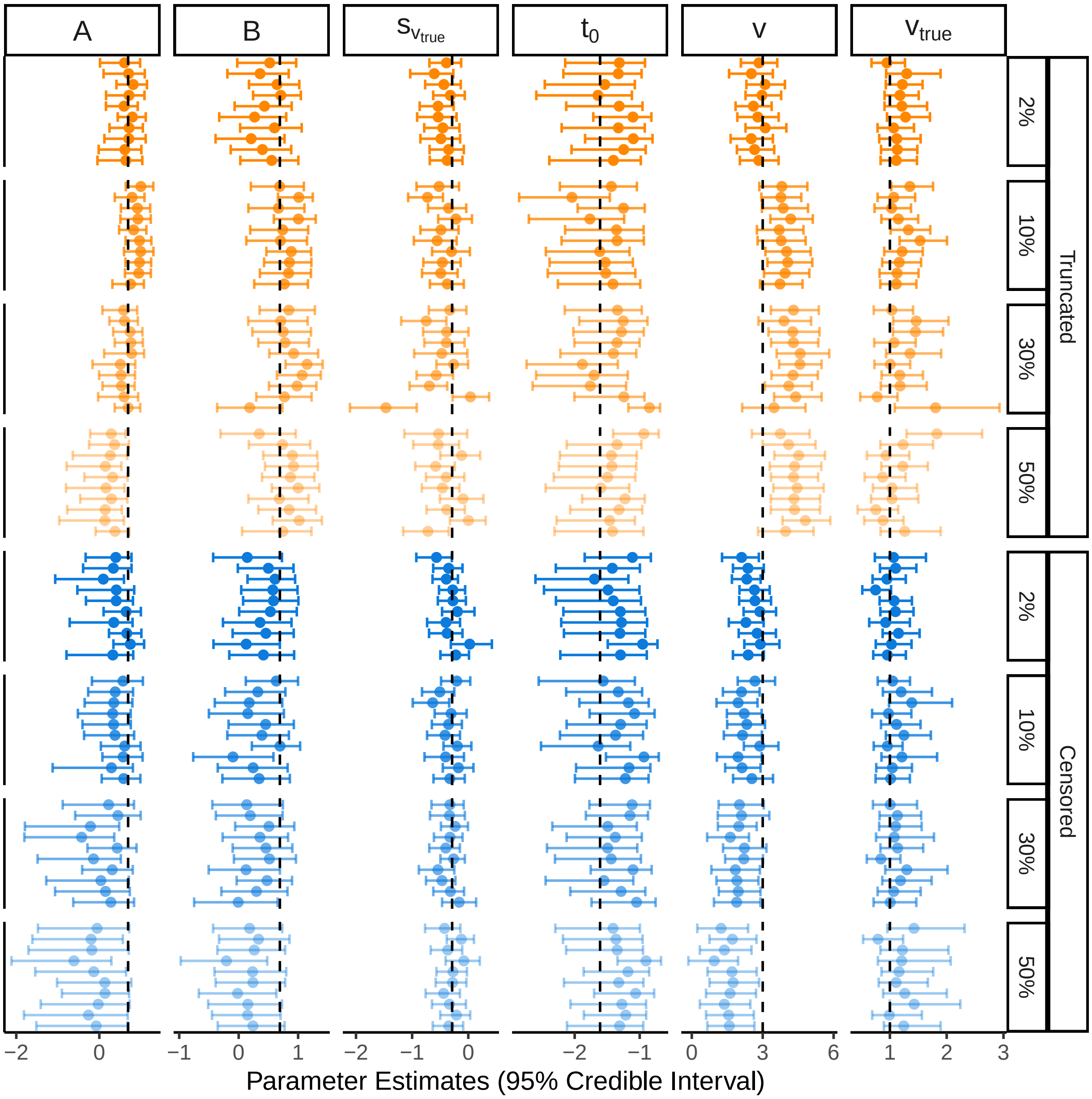
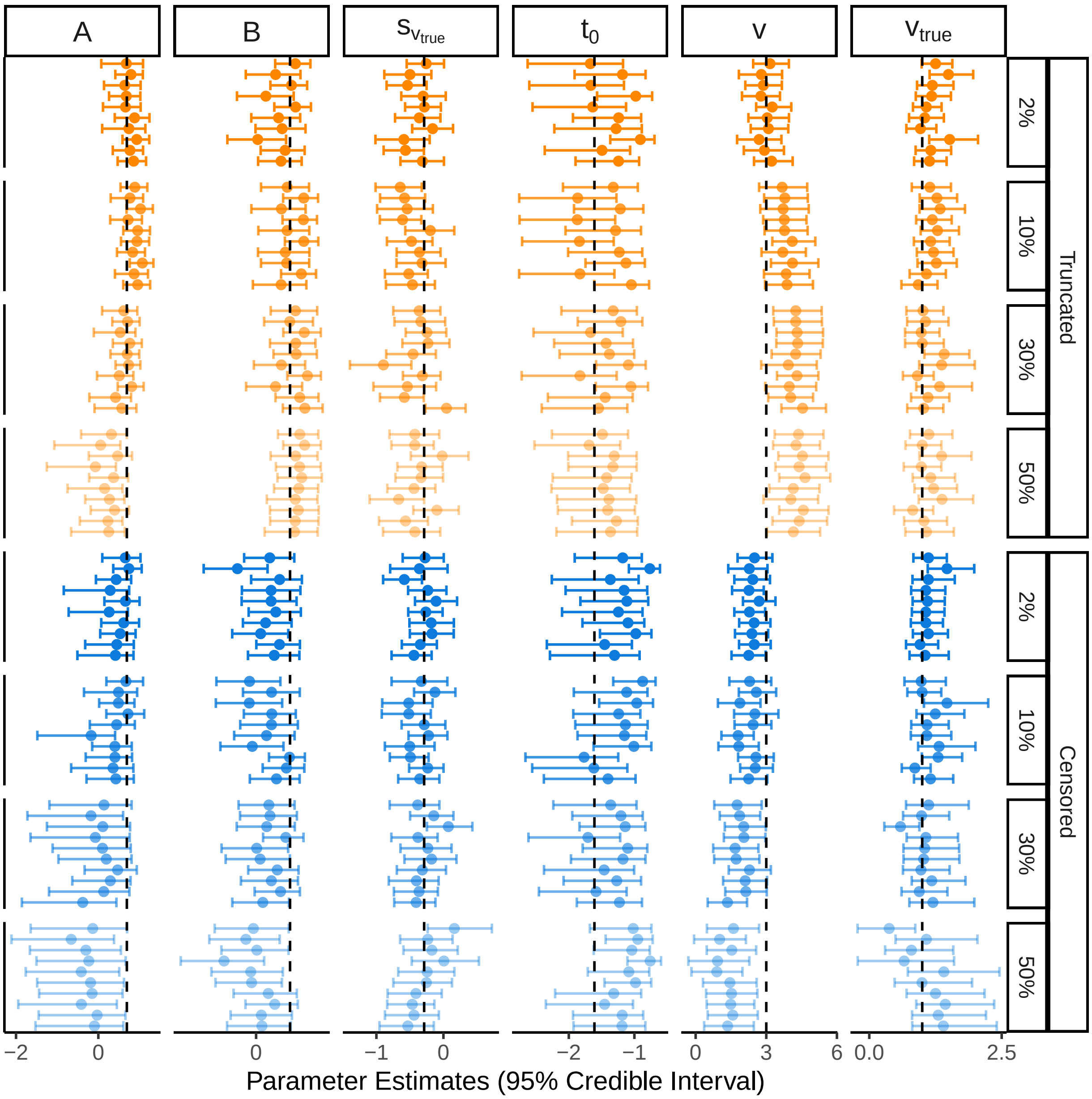


Figure A6

Linear Ballistic Accumulator Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Upper Tail and Responses Unknown



Upper Tail.

Figure A7

Linear Ballistic Accumulator Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Lower Tail and Responses Known

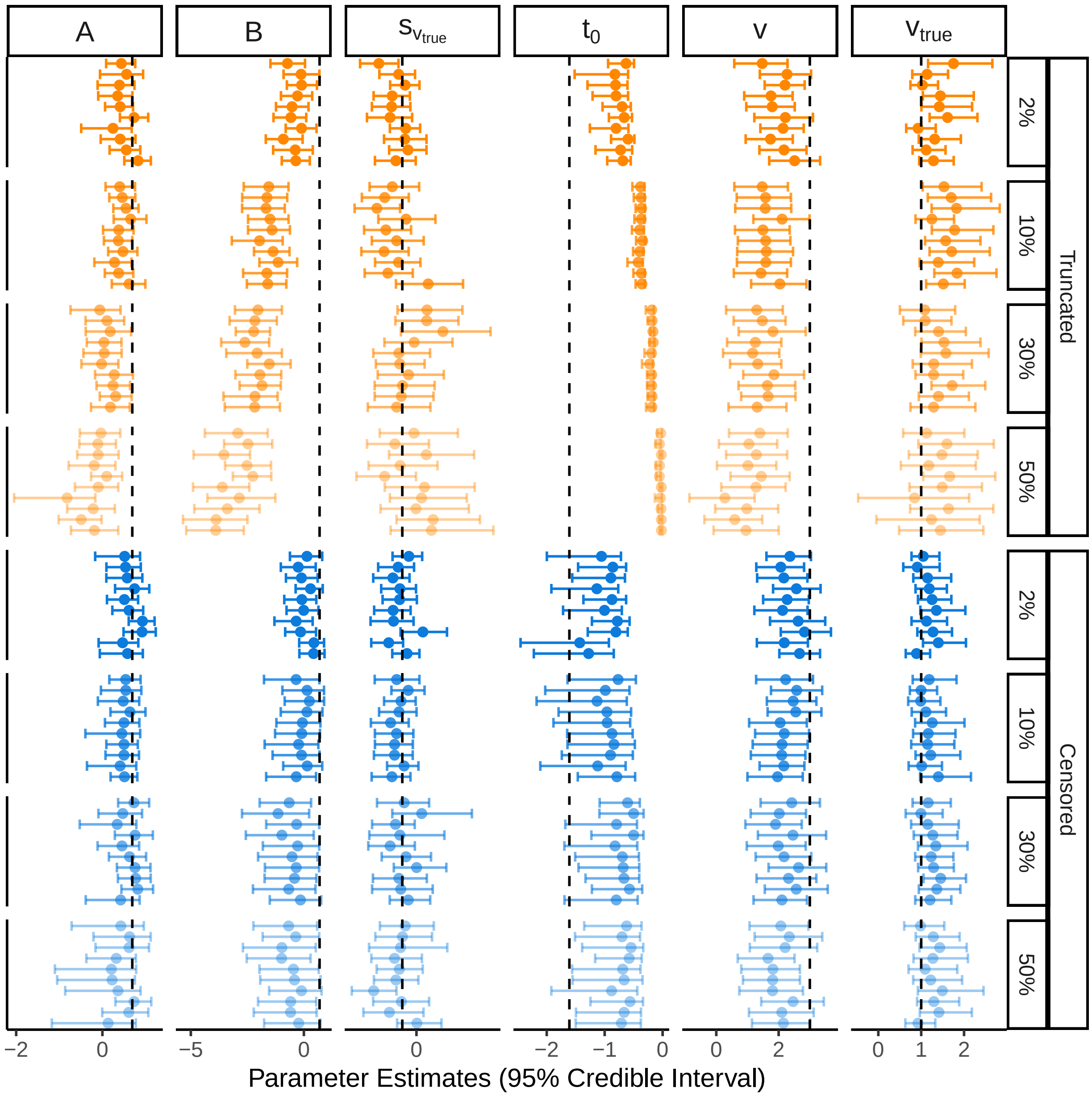
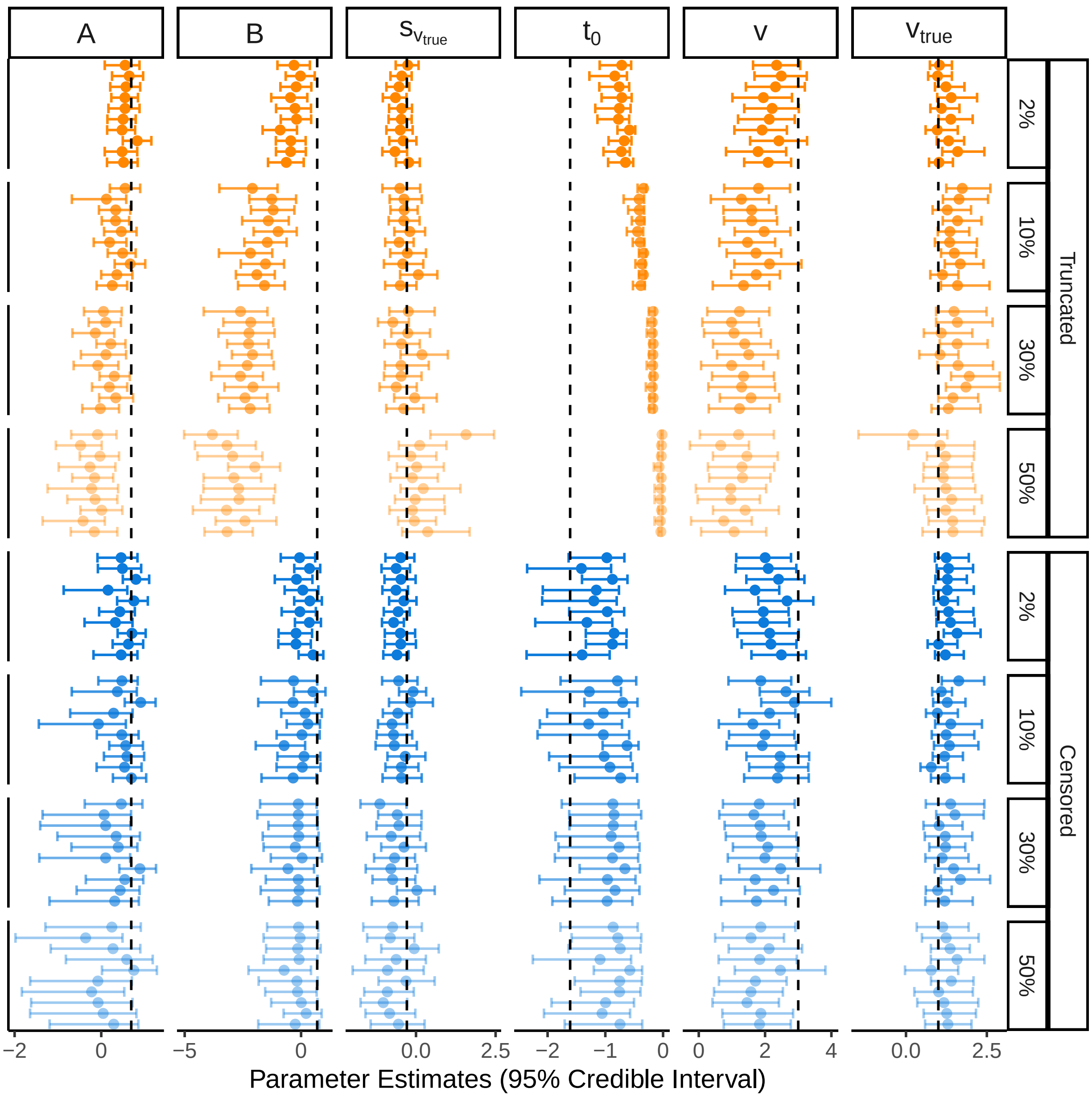


Figure A8

Linear Ballistic Accumulator Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Lower Tail and Responses Unknown



Lower Tail.

Figure A9

Linear Ballistic Accumulator Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Upper and Lower Tail and Responses Known

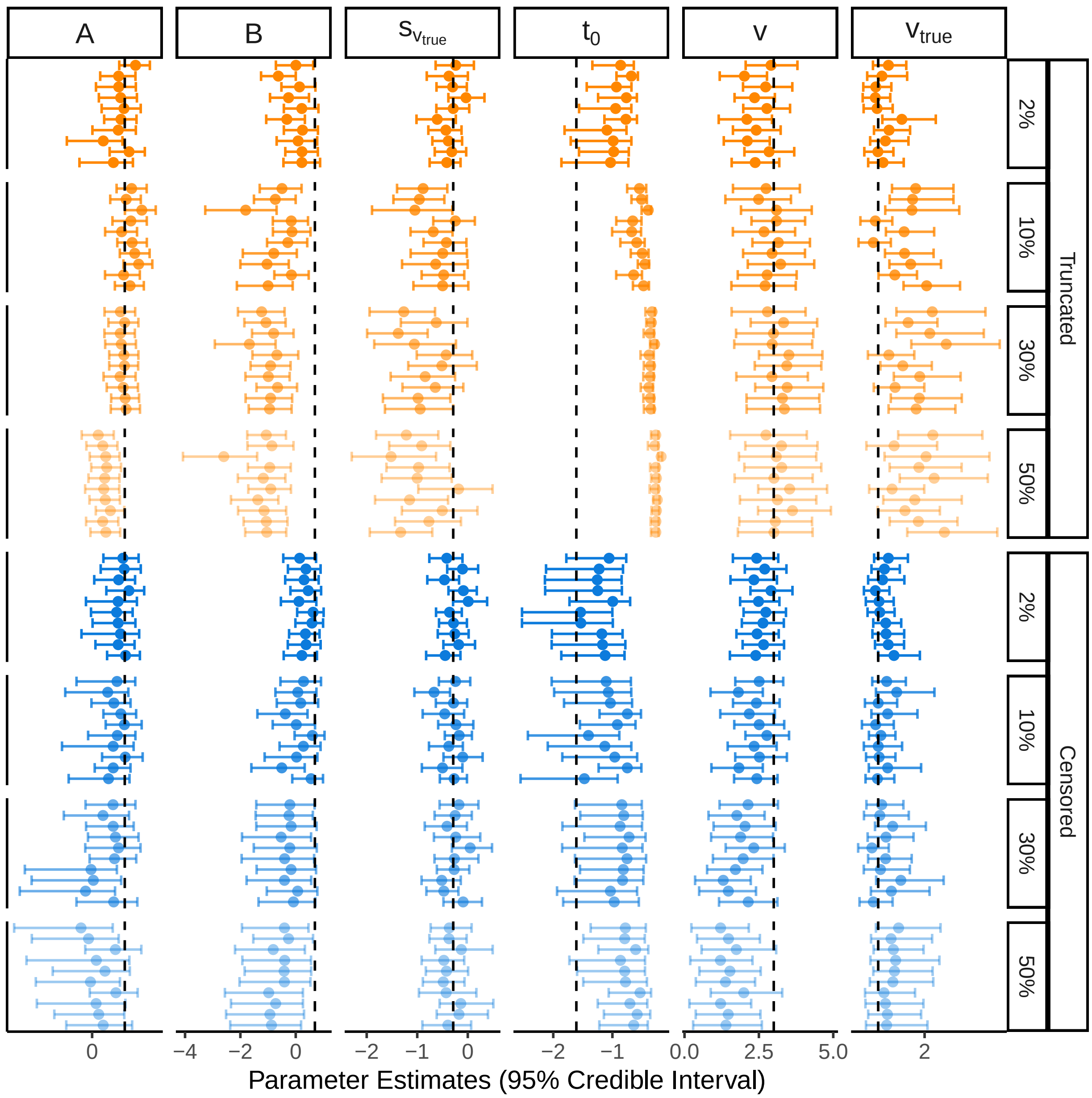
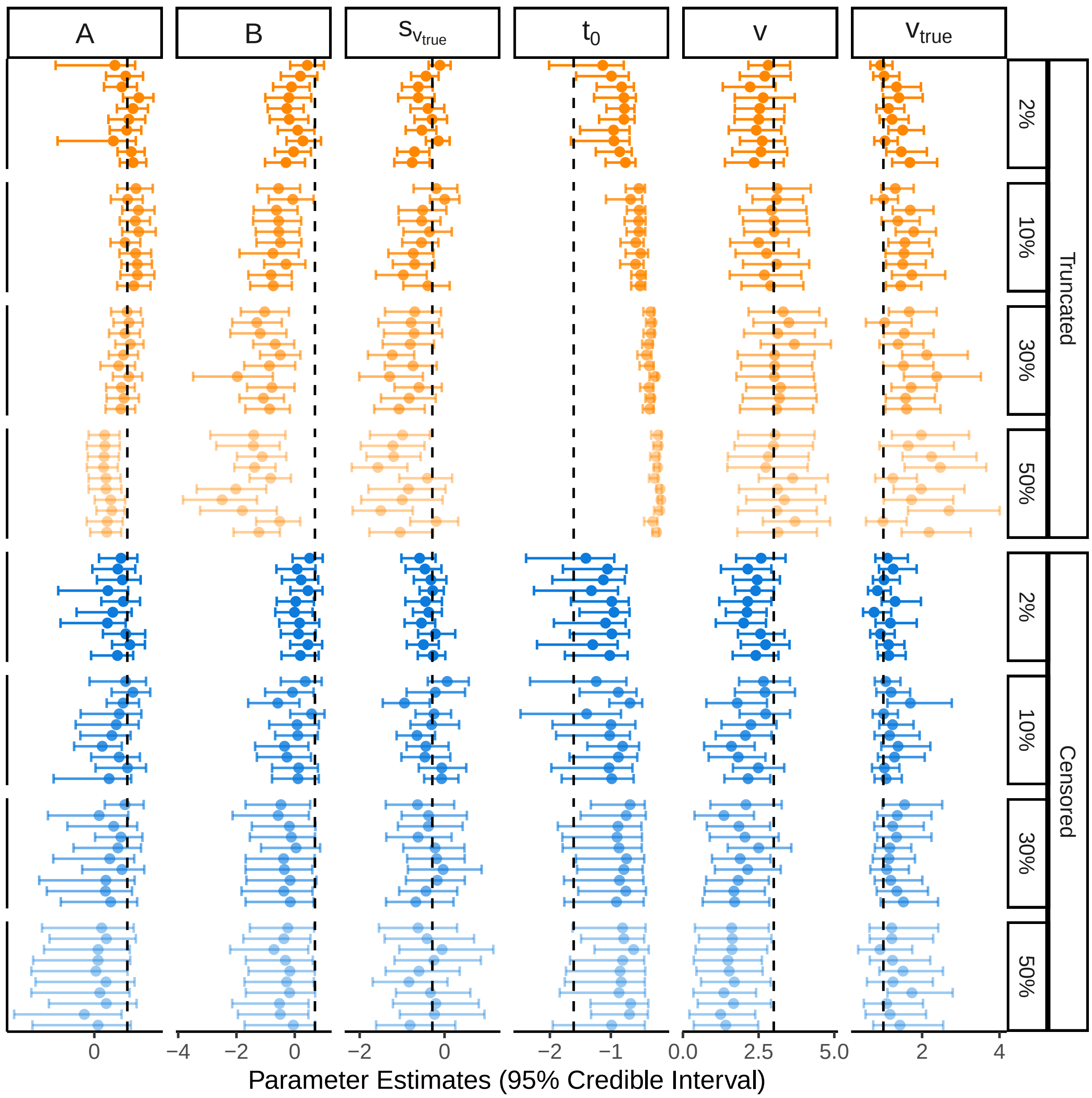


Figure A10

Linear Ballistic Accumulator Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Upper and Lower Tail and Responses Unknown



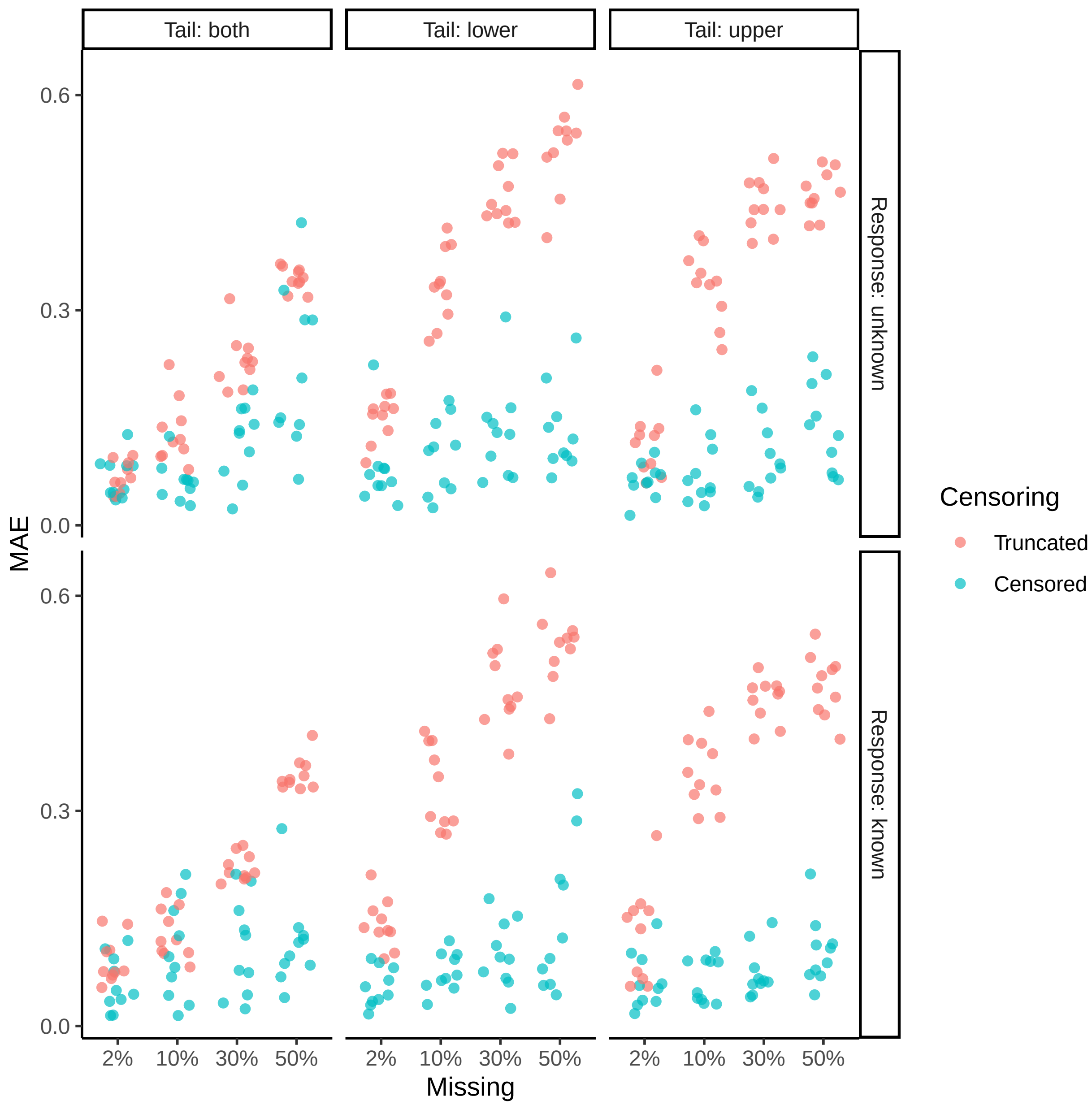
Both Tails.

Log-Normal Race Model

Model Distances.

Figure A11

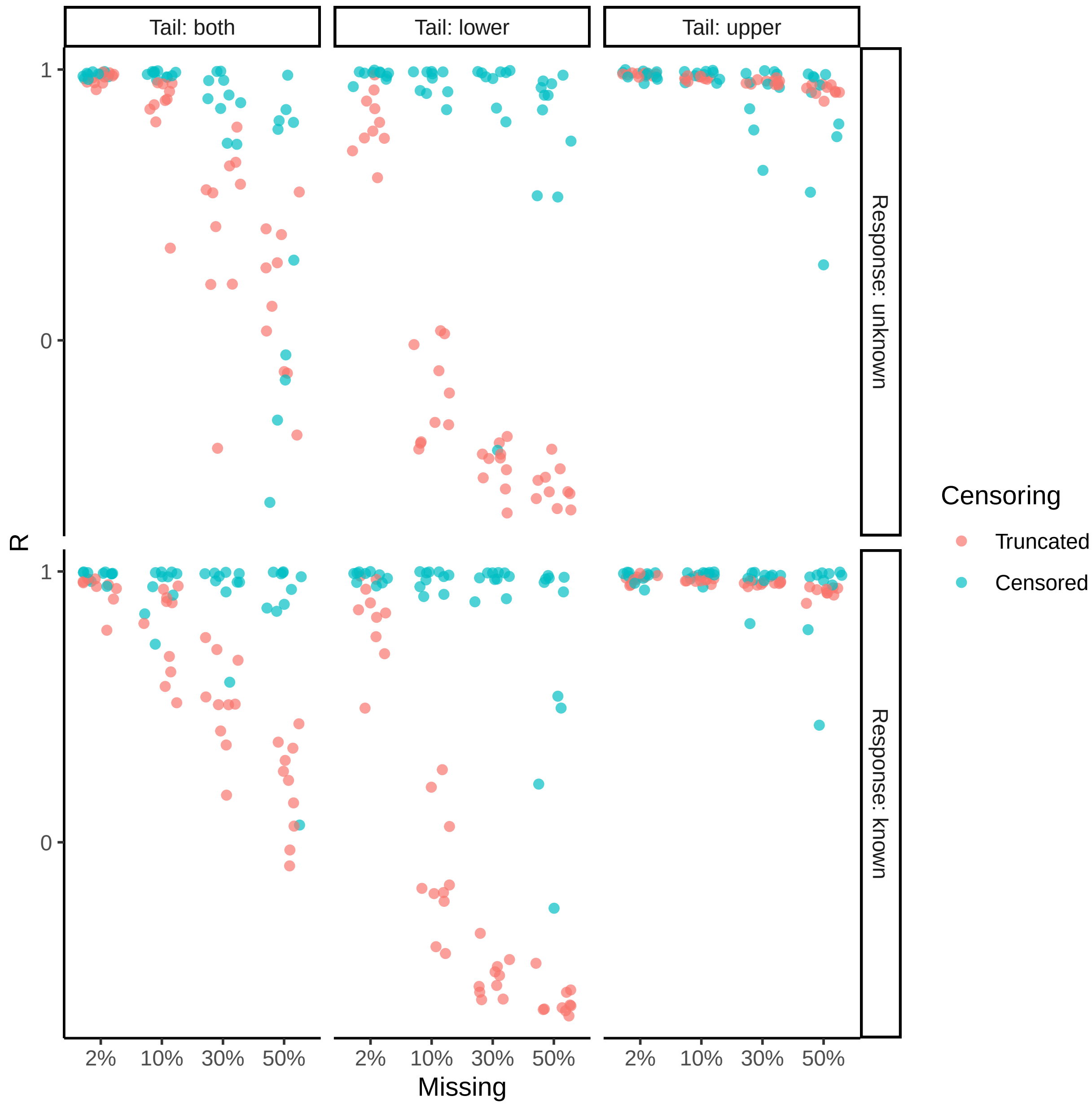
MAEs for the LNR



Note. Mean absolute errors over all parameters for the log-normal race model (LNR). Colors denote whether missing RTs were censored or truncated. Points were randomly horizontally shifted to avoid overlap. Lower MAE indicates better parameter recovery.

Figure A12

Pearson correlations for the LNR



Note. Pearson correlations (R) between the true log-normal race model (LNR) parameters and the posterior medians. Colors denote whether missing RTs were censored or truncated. Points were randomly horizontally shifted to avoid overlap. High correlation indicates better parameter recovery.

Credible Intervals.

Figure A13

Log-Normal Race Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Upper Tail and Responses Known

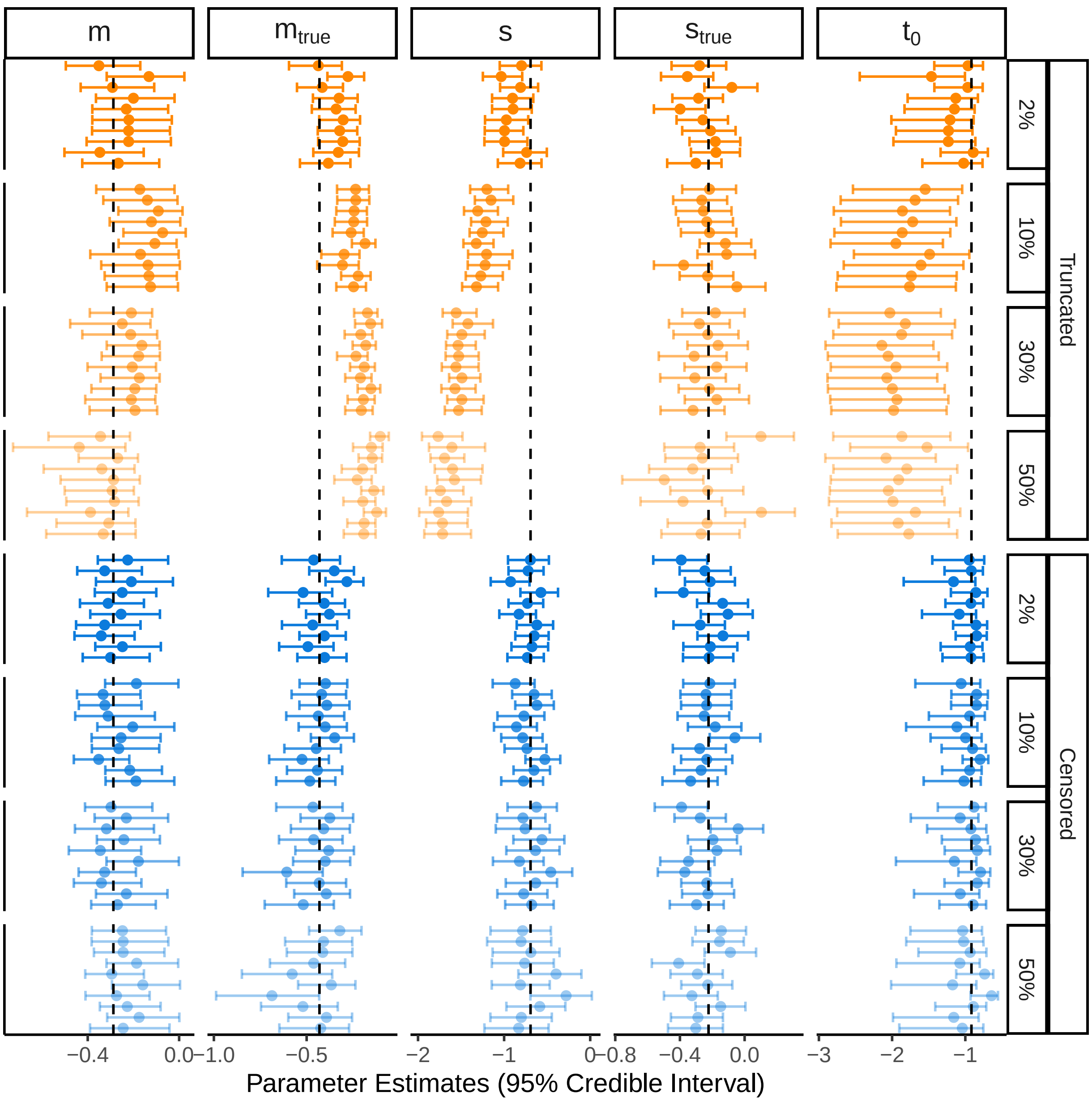
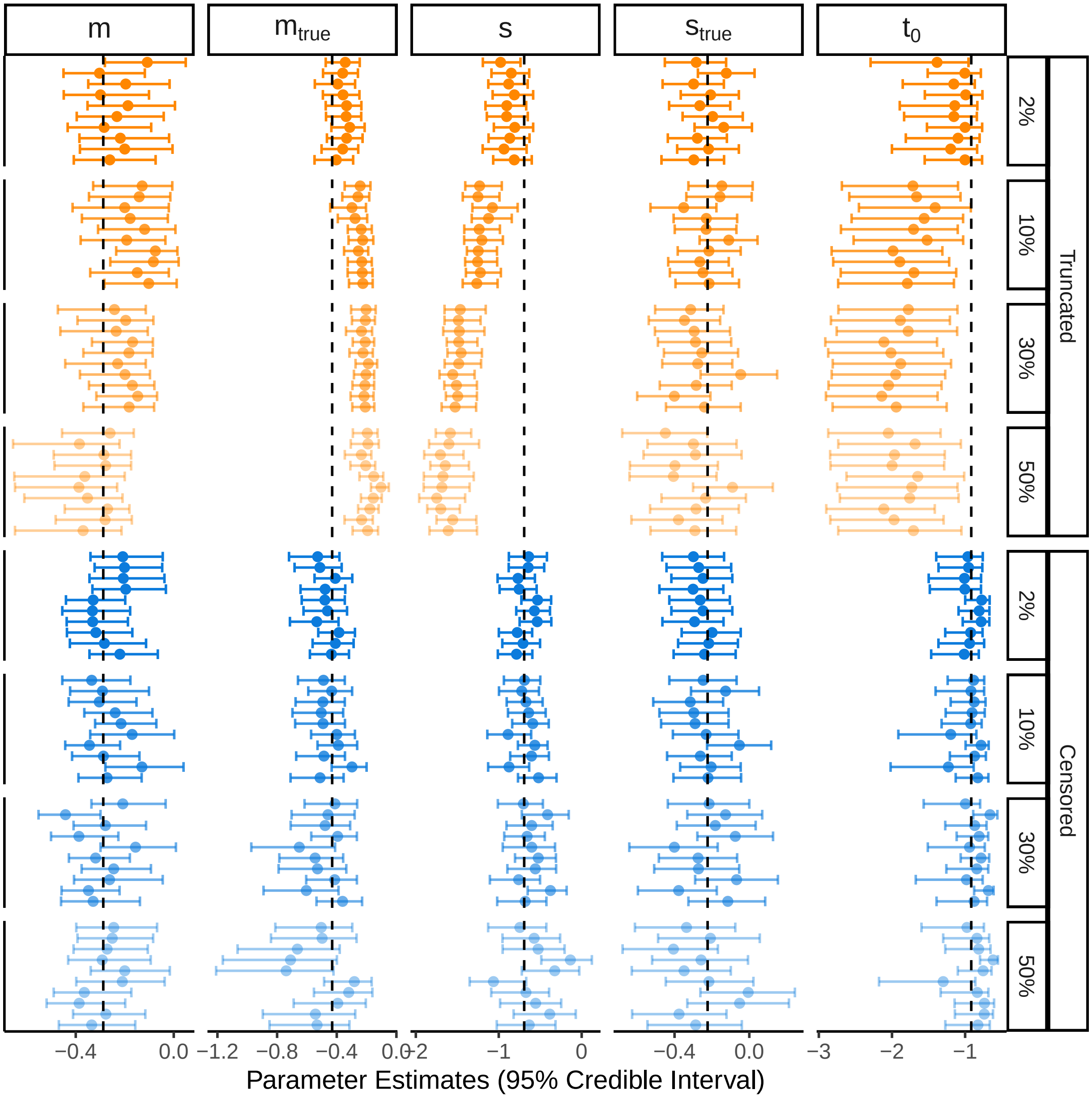


Figure A14

Log-Normal Race Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Upper Tail and Responses Unknown



Upper Tail.

Figure A15

Log-Normal Race Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Lower Tail and Responses Known

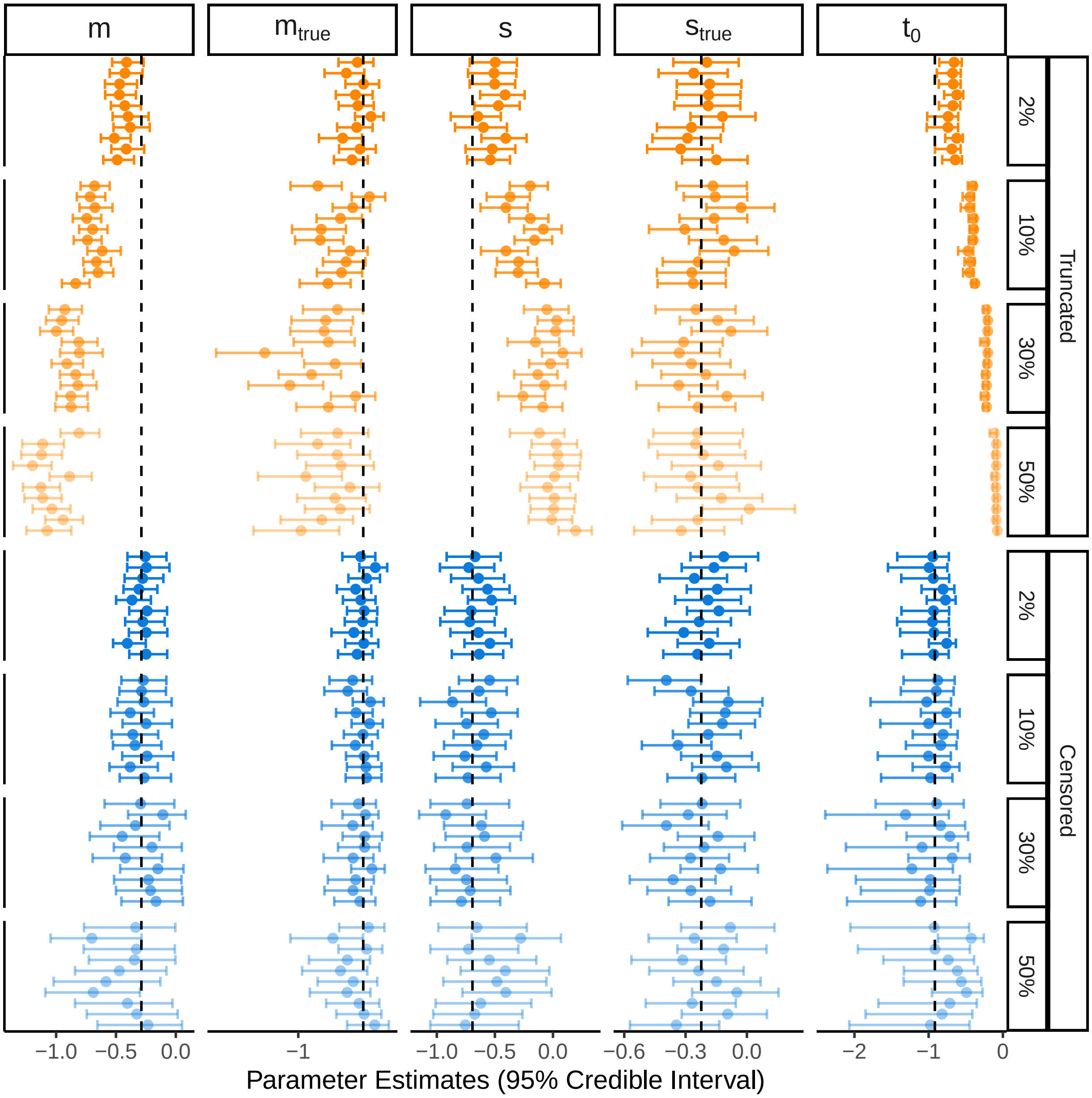
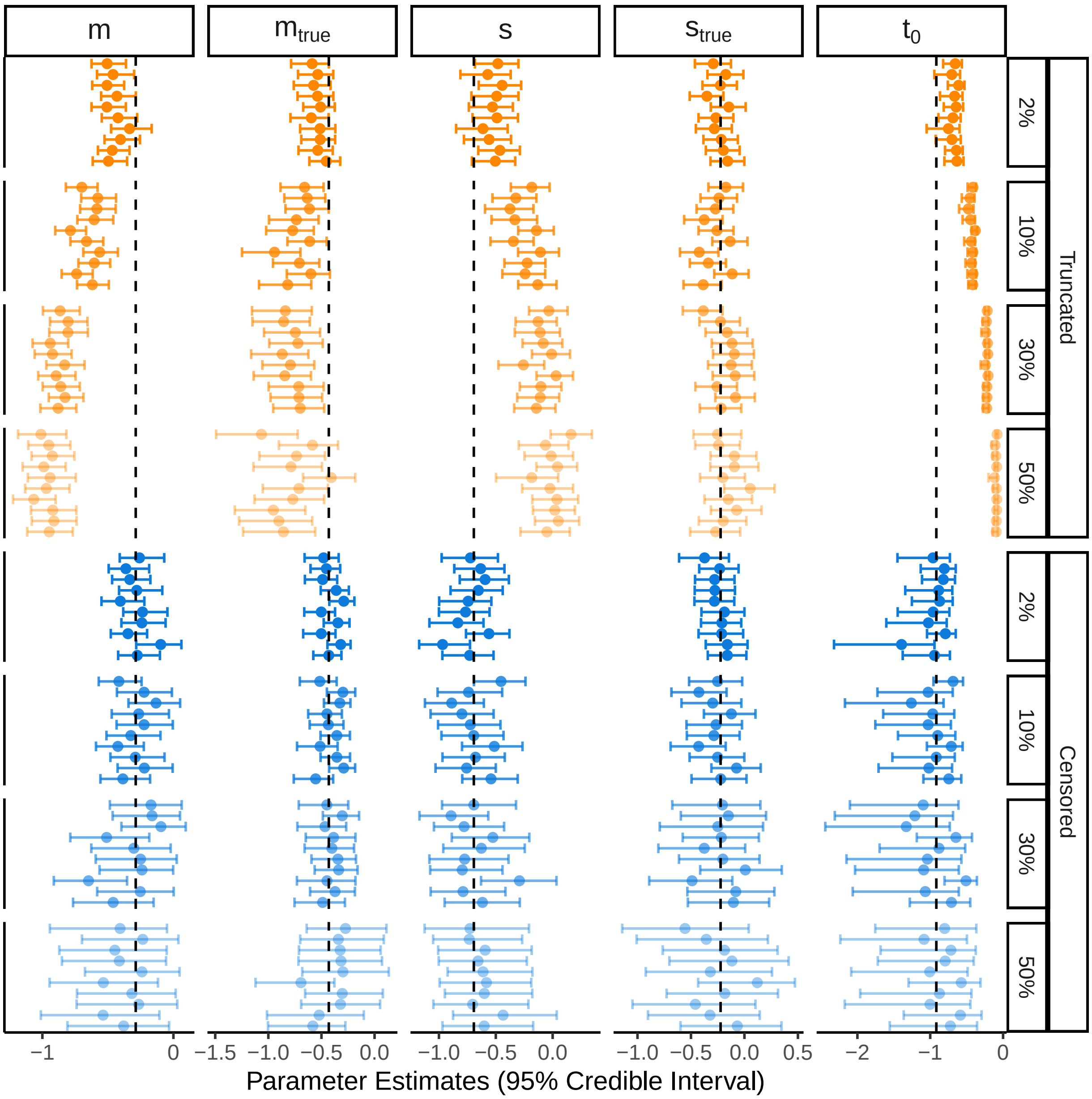


Figure A16

Log-Normal Race Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Lower Tail and Responses Unknown



Lower Tail.

Figure A17

Log-Normal Race Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Upper and Lower Tail and Responses Known

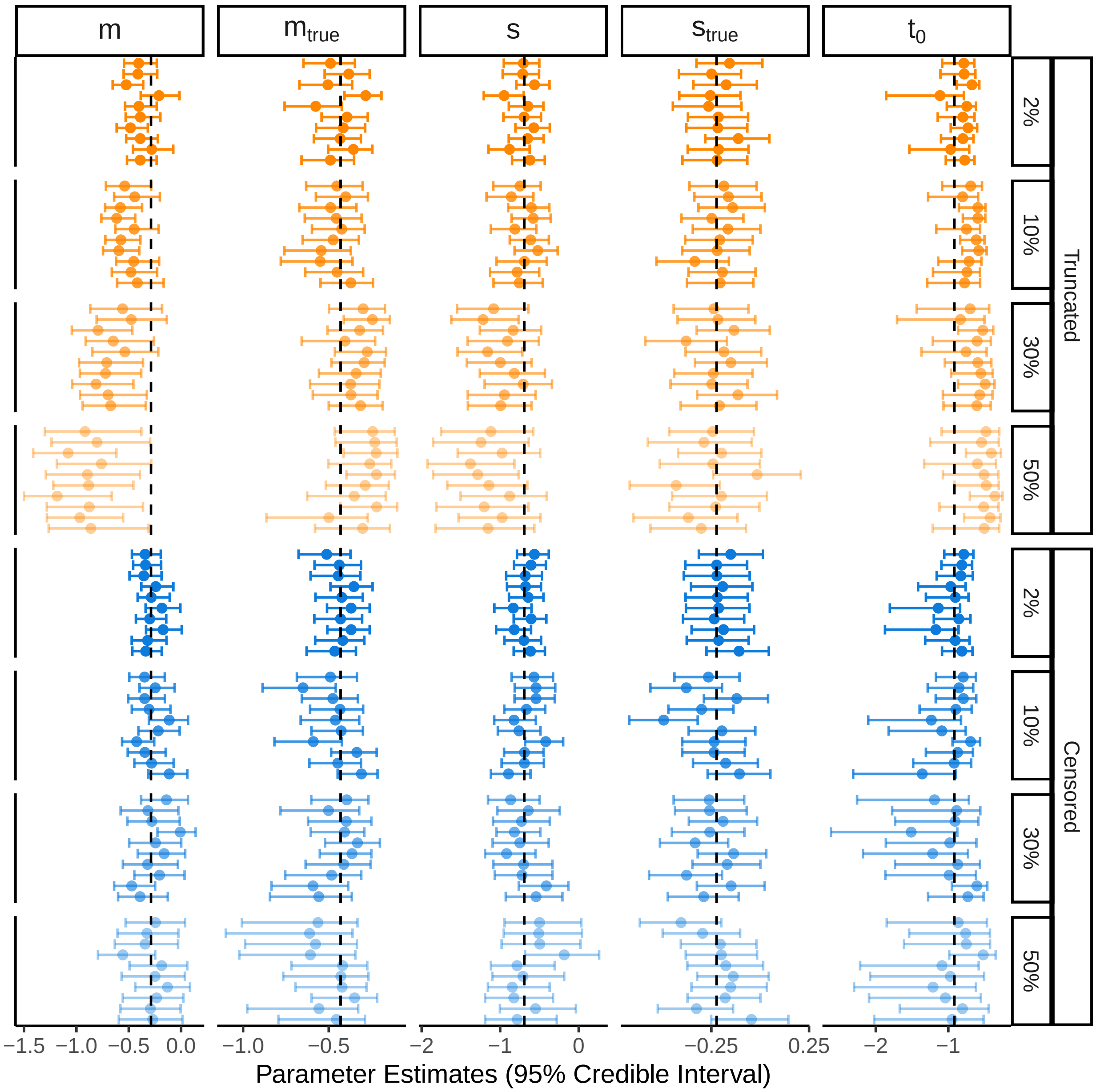
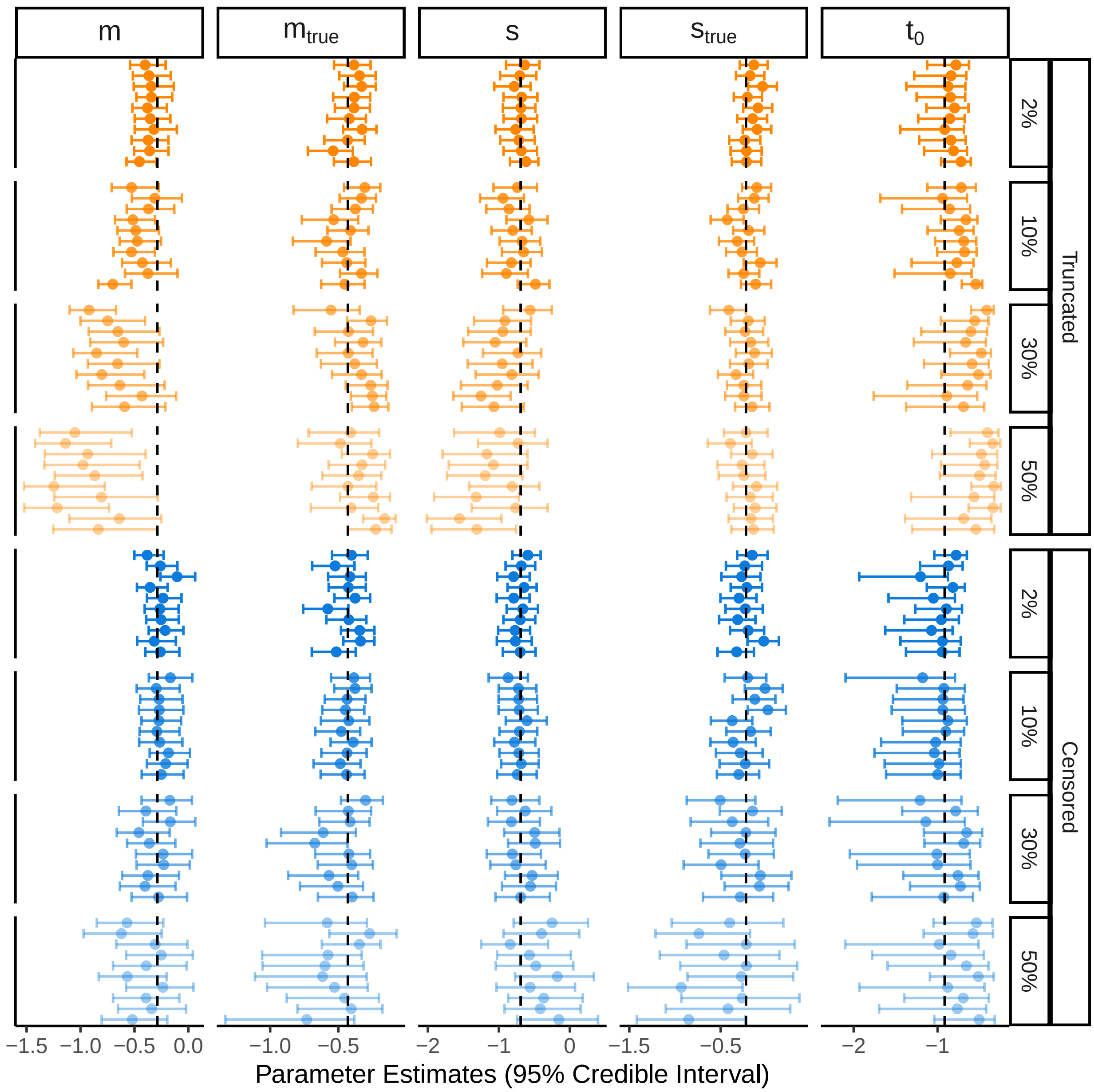


Figure A18

Log-Normal Race Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Upper and Lower Tail and Responses Unknown



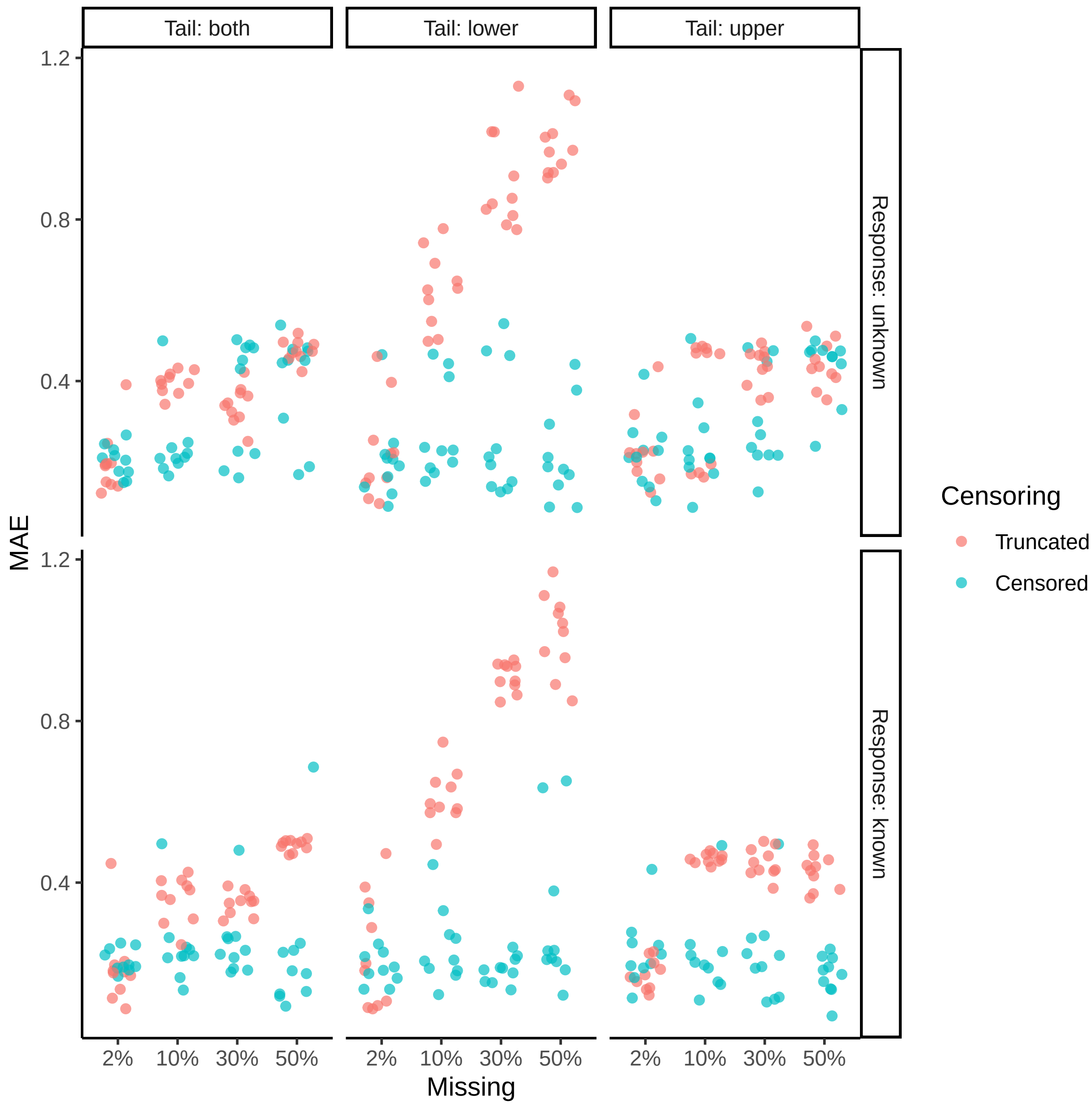
Both Tails.

Racing Diffusion Model

Model Distances.

Figure A19

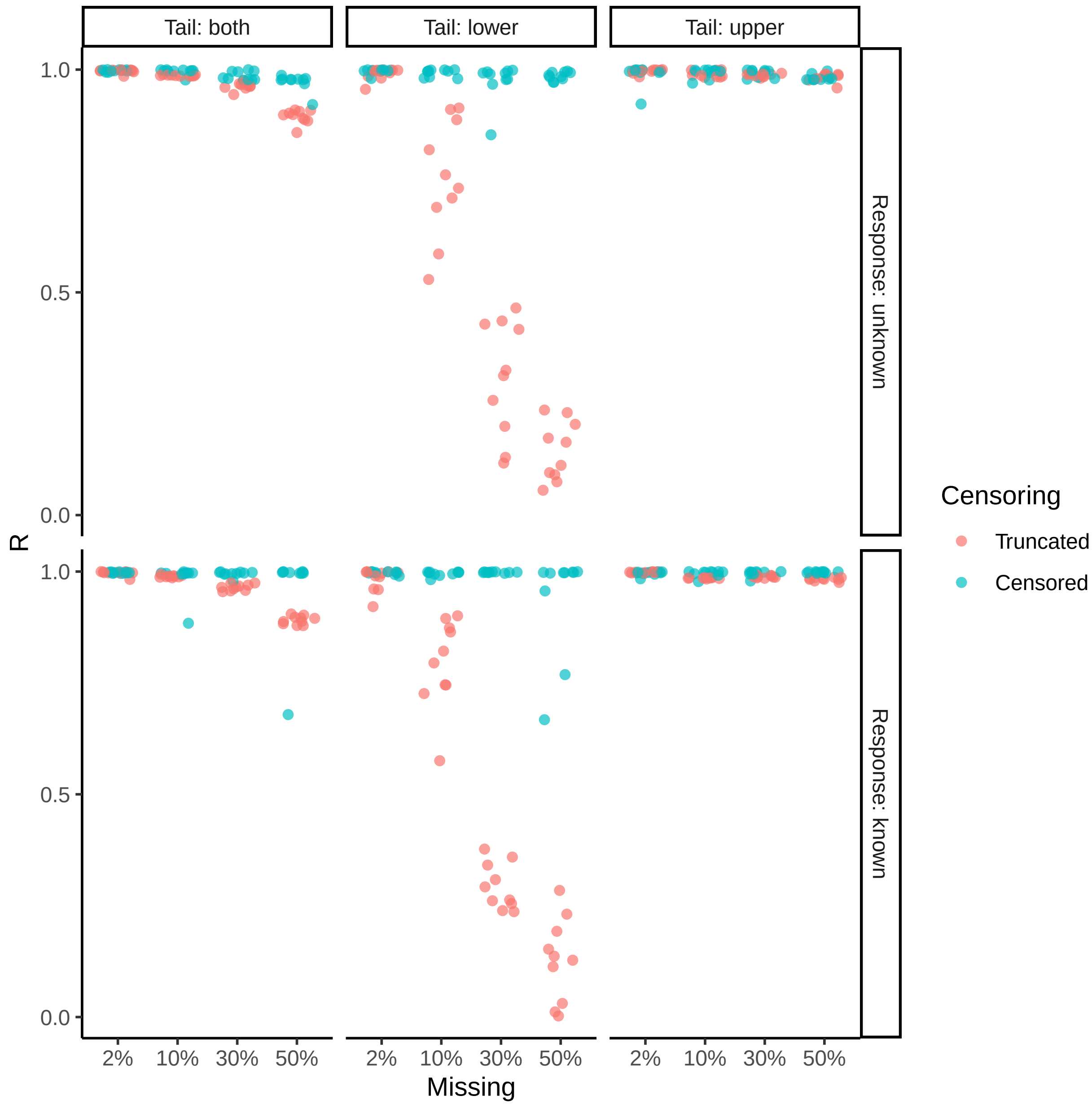
MAEs for the RDM



Note. Mean absolute errors over all parameters for the racing diffusion model (RDM). Colors denote whether missing RTs were censored or truncated. Points were randomly horizontally shifted to avoid overlap. Lower MAE indicates better parameter recovery.

Figure A20

Pearson correlations for the RDM



Note. Pearson correlations (R) between the true racing diffusion model (RDM) parameters and the posterior medians. Colors denote whether missing RTs were censored or truncated. Points were randomly horizontally shifted to avoid overlap. High correlation indicates better parameter recovery.

Credible Intervals.

Figure A21

Racing Diffusion Model Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Upper Tail and Responses Known

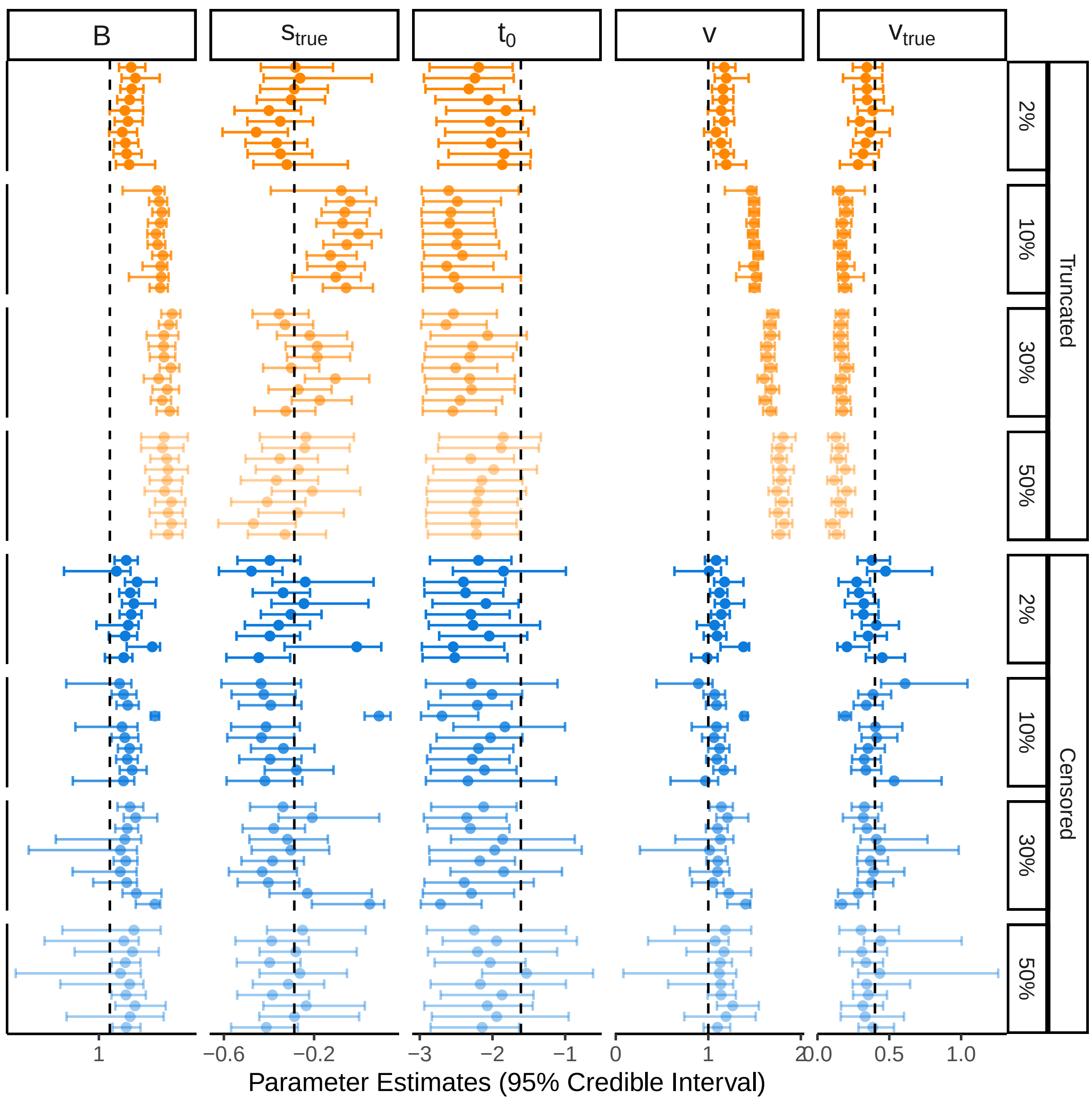
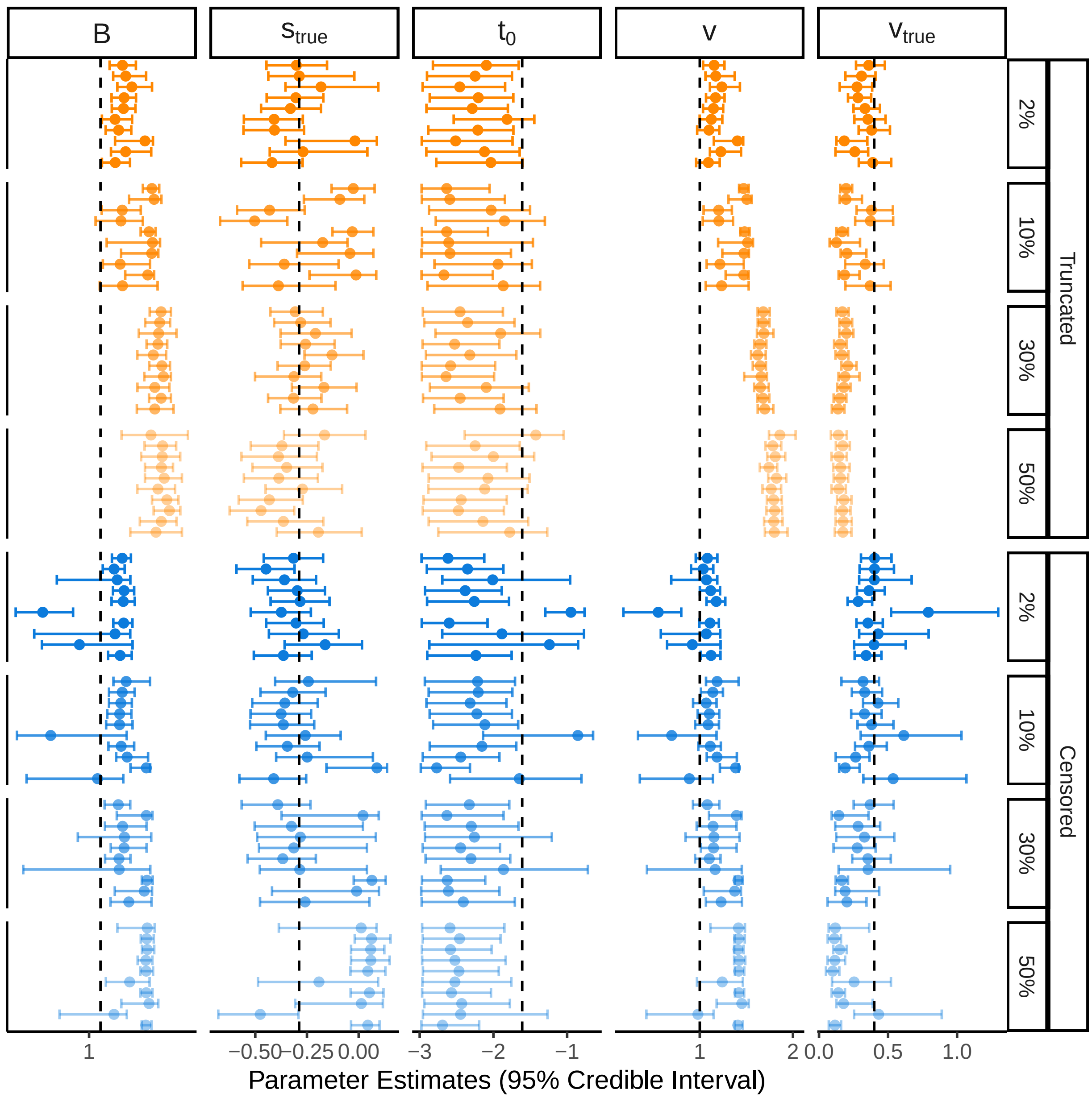


Figure A22

Racing Diffusion Model Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Upper Tail and Responses Unknown



Upper Tail.

Figure A23

Racing Diffusion Model Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Lower Tail and Responses Known

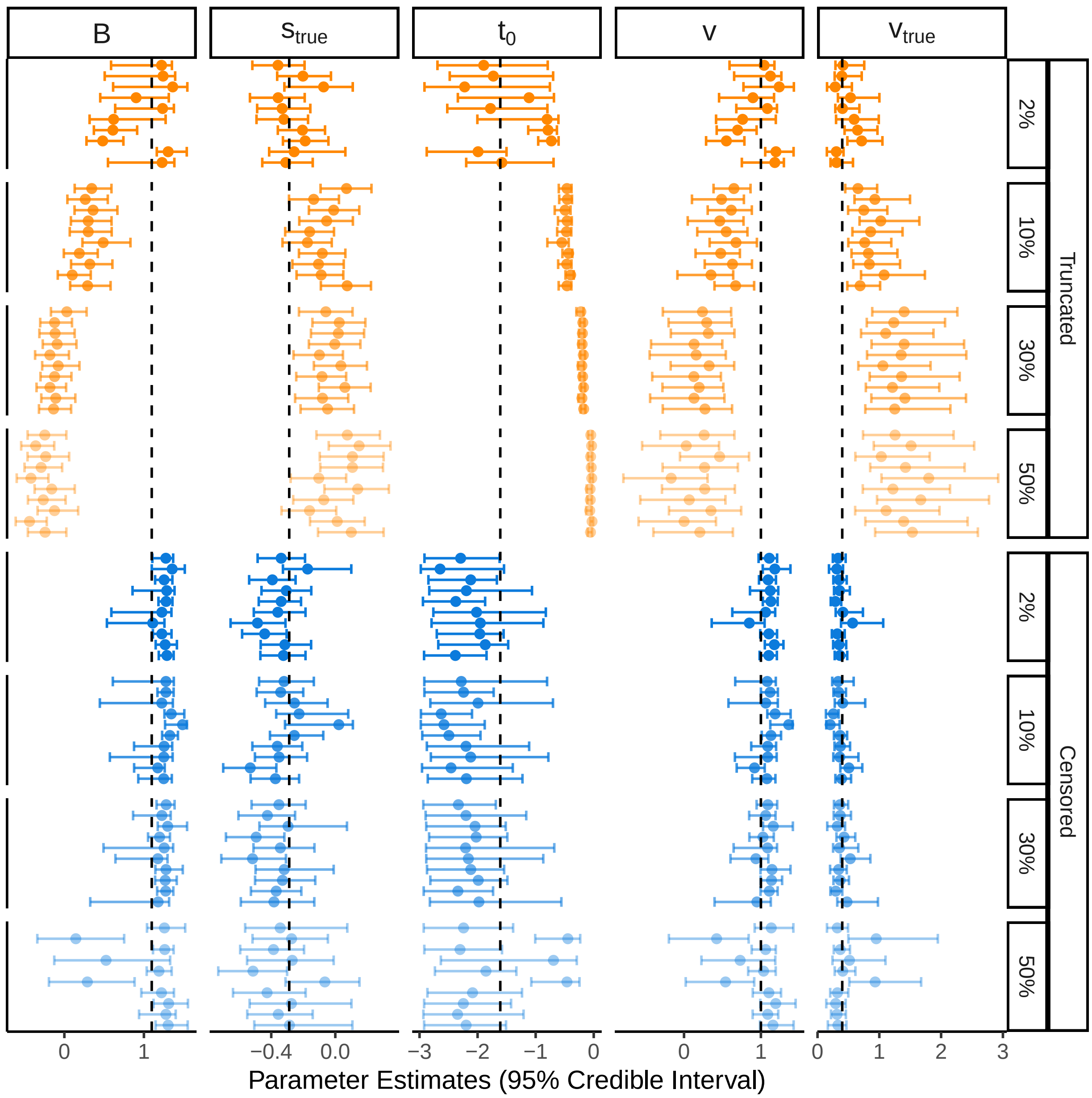
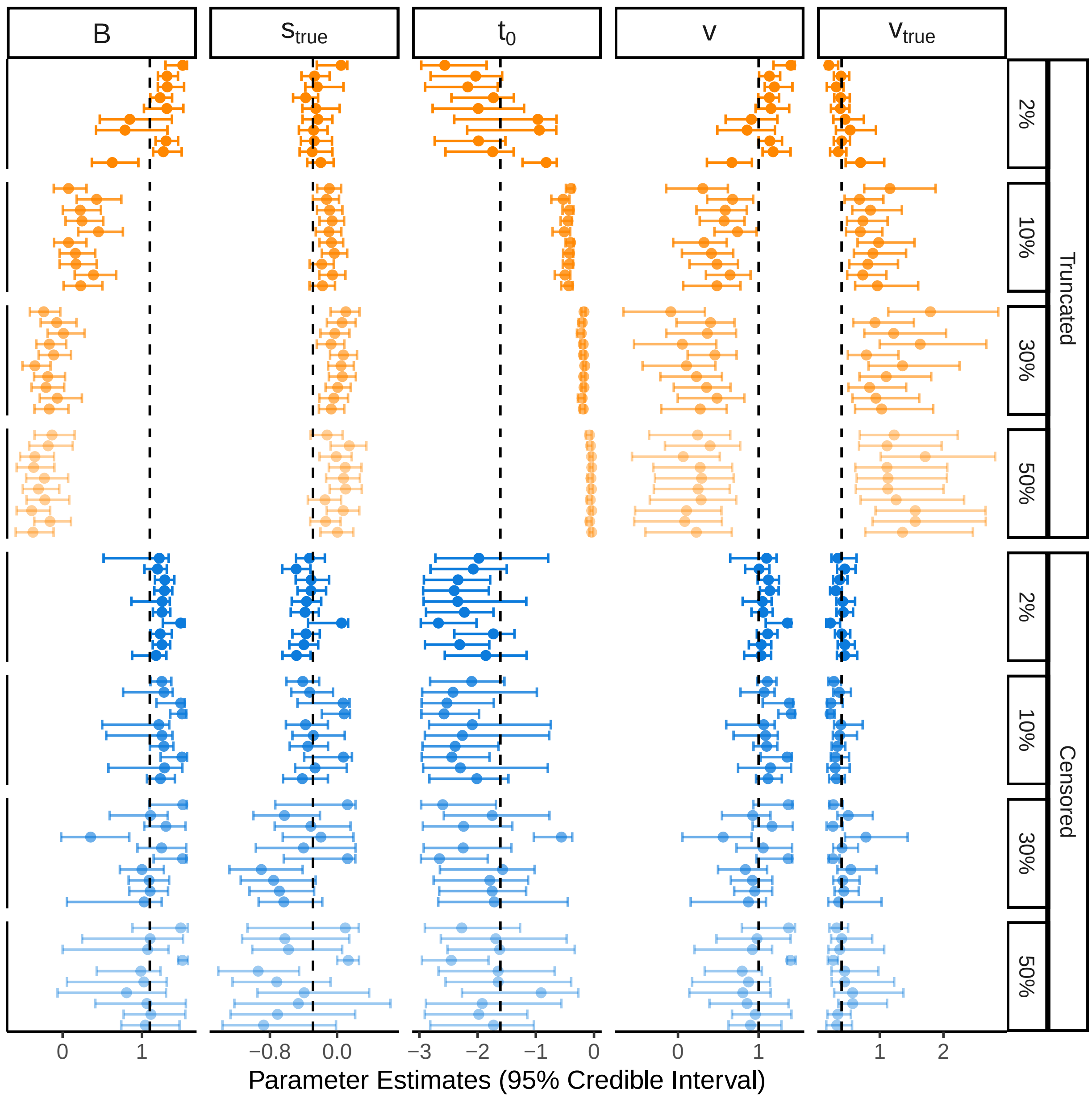


Figure A24

Racing Diffusion Model Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Lower Tail and Responses Unknown



Lower Tail.

Figure A25

Racing Diffusion Model Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Upper and Lower Tail and Responses Known

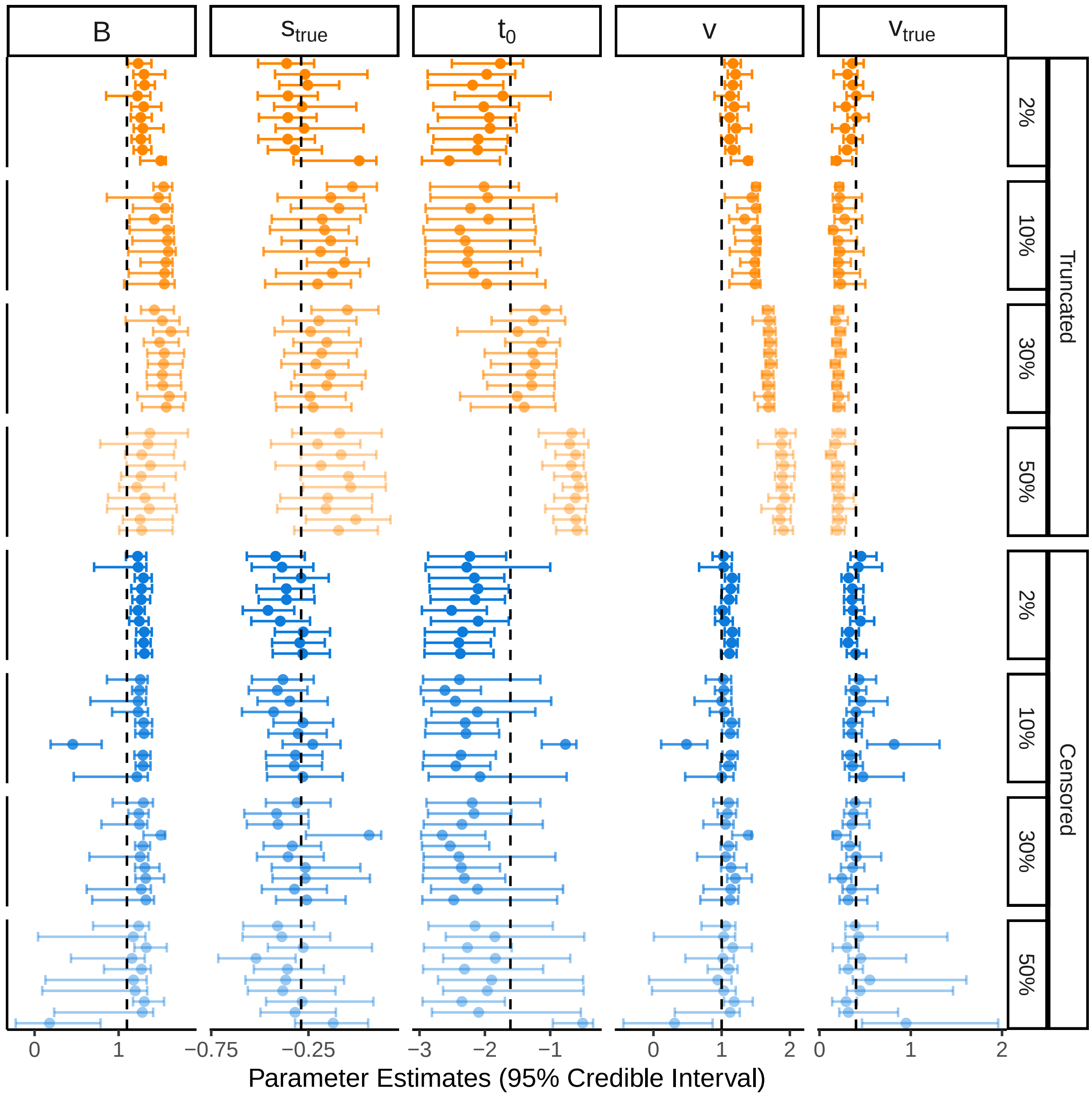
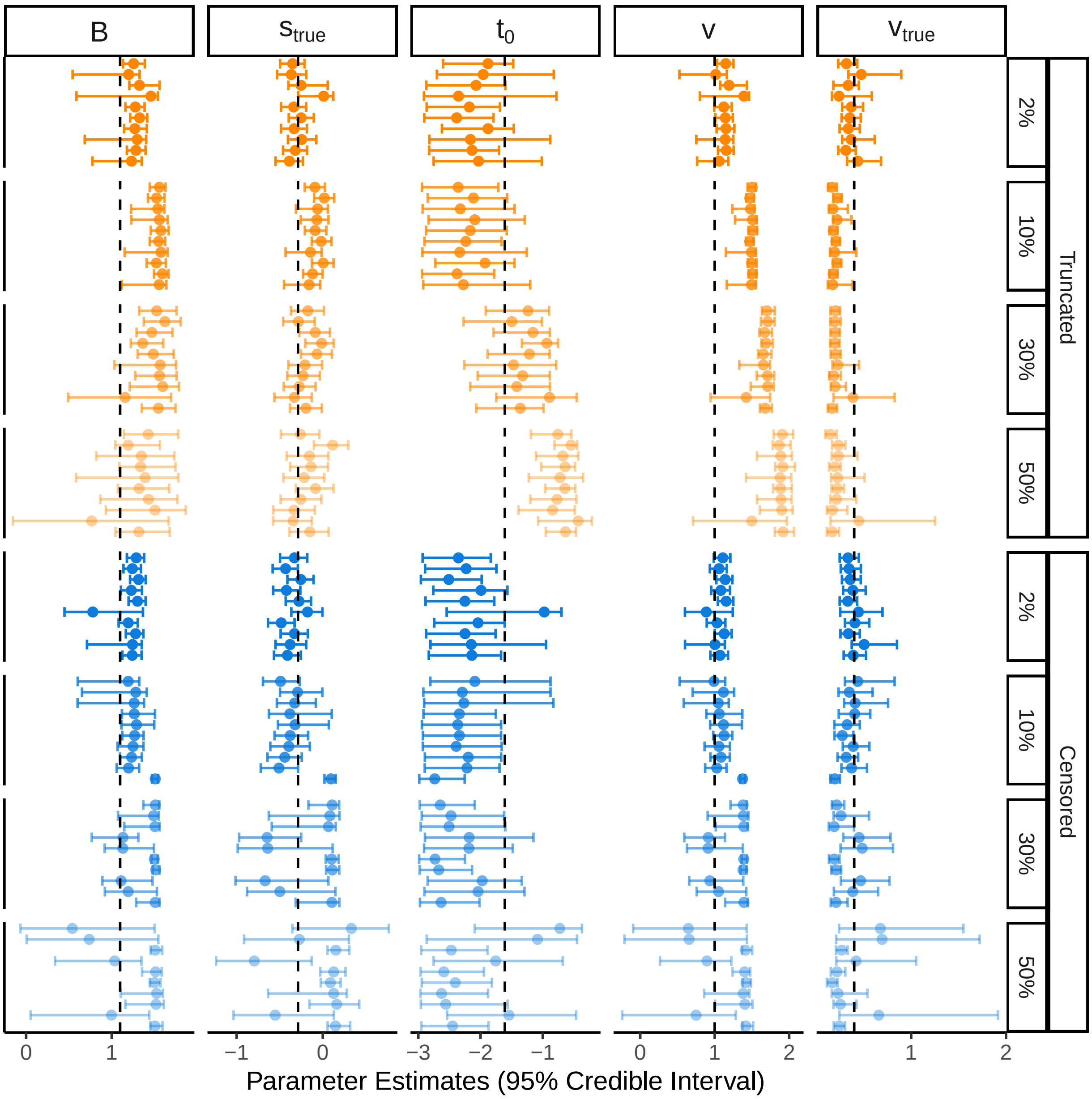


Figure A26

Racing Diffusion Model Posterior Medians and 95% Equal Tailed Credible Intervals with Missing Upper and Lower Tail and Responses Unknown



Both Tails.