

Supplementary Material F: Aalen-Johansen, Nelson-Aalen, and product-integral applications

The contents include:

- (I) Violin plots for additional simulation studies to illustrate the contexts when the Aalen-Johansen and Nelson-Aalen estimators result in unbiased and biased estimation of transition rate parameters.
- (II) Computational time differences between using approach (A) with `deSolve` versus with `prodint`.

Part (I)

Setup

The motivation for these simulation studies is to verify that the biased estimation from approach (E) in Sections 3.3 and 3.4 is due precisely to violating the assumptions that all observation times are transition times and all transitions are observed. The conclusion being, if the modeling assumptions for the Aalen-Johansen and Nelson-Aalen estimators are met, applying approach (E) detailed in Section 5.2 would lead to unbiased estimation of the transition rate parameters. The data generating mechanism for these simulation studies is the same as that defined in Section 3.3 of the manuscript, where the true parameter values defining the model are:

$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{pmatrix} = \begin{pmatrix} \beta_{0,1} & \beta_{1,1} & \beta_{2,1} \\ \beta_{0,2} & \beta_{1,2} & \beta_{2,2} \\ \beta_{0,3} & \beta_{1,3} & \beta_{2,3} \\ \beta_{0,4} & \beta_{1,4} & \beta_{2,4} \\ \beta_{0,5} & \beta_{1,5} & \beta_{2,5} \end{pmatrix} = \begin{pmatrix} -2.316173 & 0.2705000 & -0.39079609 \\ -1.287563 & -0.4930641 & -0.05894252 \\ -1.101164 & 0.2705000 & -0.32509646 \\ -2.523675 & 0.2886209 & 0.48631653 \\ -2.103848 & 0.2273128 & 0.99565810 \end{pmatrix}, \quad \pi = \begin{pmatrix} 0.996593594 \\ 0.001464950 \\ 0.001941456 \\ 0 \end{pmatrix}, \quad \begin{pmatrix} p_1 \\ p_2 \\ p_3 \\ p_4 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}.$$

Note that because $p_1 = p_2 = p_3 = p_4 = 0$, there is no misclassification of states (i.e., state observations are without error). This eliminates the possibility that biased parameter estimation is a result of state misclassifications. For each simulated multistate Markov process, we will save three different versions, each representing different assumptions about the observed process:

- (i) observation times do not necessarily correspond to transitions times; not all state transitions are necessarily observed
- (ii) it is known which observation times correspond to transition times; not all state transitions are necessarily observed
- (iii) all observation times correspond to transition times; all state transitions are observed

Case (i) corresponds to the assumptions made in the data generating mechanism for the simulation study in Section 3.3 modulo no state misclassification error, and case (iii) corresponds to the modeling assumptions for the Aalen-Johansen and Nelson-Aalen estimators. One hundred data sets are simulated for each case (each with 3000 sampled subjects), and for each data set, approaches (A) and (E) are fit. Violin plots of the estimated transition rate parameters are presented in Figures 1, 2, and 3 for cases (i), (ii), and (iii), respectively.

Likelihood construction

Because the three cases represent different modeling assumptions, the likelihood function will vary depending on each case. To start, the likelihood expression (7) from the manuscript is used for case (i) with $\mathbf{D}_{(i,k)} = \text{diag}(\mathbf{1}\{s_{i,k} = 1\}, \dots, \mathbf{1}\{s_{i,k} = m\})$ because no state misclassifications are allowed in this simulation. However, for cases (ii) and (iii), a further transformation of the likelihood in (7) is needed.

More broadly, using the ideas from Section 2.4.1 of Williams et al. (2020), we can account for observing exact transition times in cases (ii) and (iii) by the following. For a multistate process, $X(t)$, suppose $X(t_{i,k-1}) = s_{i,k-1}$ and $X(t_{i,k}) = s_{i,k}$ such that $s_{i,k-1} \neq s_{i,k}$ for $i \in \{1, 2, \dots, n\}$, $k \in \{2, 3, \dots, n_i\}$, and $s_{i,k-1}, s_{i,k} \in \{1, 2, \dots, m\}$. For case (ii), it is not guaranteed that all state transitions are observed, and so we have to account for all possible transitions into $s_{i,k}$ at precisely $t_{i,k}$. Accordingly, the likelihood contribution is

$$\sum_{\{\ell \in \{1, \dots, m\} \mid \ell \neq s_{i,k-1}\}} P\{X(t_{i,k}) = \ell \mid X(t_{i,k-1}) = s_{i,k-1}\} \cdot [\mathbf{Q}(t_{i,k})]_{\ell, s_{i,k}}.$$

Thus, for case (ii), the likelihood in (7) is written as

$$f_{s_1, \dots, s_n}(\mathbf{s}_1, \dots, \mathbf{s}_n) = \prod_{i=1}^n \boldsymbol{\pi}^T \mathbf{D}_{(i,1)} \cdot \mathbf{W}^*(t_{i,1}, t_{i,2}) \mathbf{D}_{(i,2)} \cdots \mathbf{W}^*(t_{i,n_i-1}, t_{i,n_i}) \mathbf{D}_{(i,n_i)} \cdot \mathbf{1}_m, \quad (*)$$

with $\mathbf{D}_{(i,k)} = \text{diag}(\mathbf{1}\{s_{i,k} = 1\}, \dots, \mathbf{1}\{s_{i,k} = m\})$ and

$$\mathbf{W}^*(t_{i,k-1}, t_{i,k}) := \begin{cases} \mathbf{P}(t_{i,k-1}, t_{i,k}) \mathbf{Q}^*(t_{i,k}), & s_{i,k-1} \neq s_{i,k}, \\ \mathbf{P}(t_{i,k-1}, t_{i,k}), & s_{i,k-1} = s_{i,k}, \end{cases}$$

where $\mathbf{Q}^*(t_{i,k})$ is the transition rate matrix with zeros along the diagonal. In the case of this simulation study,

$$\mathbf{Q}^*(t_{i,k}) = \begin{pmatrix} 0 & q_1 & 0 & q_2 \\ 0 & 0 & q_3 & q_4 \\ 0 & 0 & 0 & q_5 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$

with $q_j = \exp(\beta_{0,j} + \beta_{1,j} \cdot t + \beta_{2,j} \cdot \text{sex})$ for $j \in \{1, \dots, 5\}$, where t is time and “sex” $\in \{0, 1\}$ is the biological sex of the subject.

For case (iii), it is given that *all* state transitions are observed. As a result, we no longer need to account for all possible transitions into state $s_{i,k}$ at $t_{i,k}$ as in (*). Instead, we need to model the multistate process *remaining* in state $s_{i,k-1}$ between $t_{i,k-1}$ and $t_{i,k}$, and immediately transitioning to state $s_{i,k}$ at $t_{i,k}$. This is accomplished using a similar approach to how the product integral is defined, where

$$\begin{aligned} p_{s_{i,k-1}}(t_{i,k-1}, t_{i,k}) &:= P\{X(t) = s_{i,k-1}, \forall t \in [t_{i,k-1}, t_{i,k}) \mid X(t_{i,k-1}) = s_{i,k-1}\} \\ &= \lim_{N \rightarrow \infty} \prod_{j=0}^{N-1} P\{X(t_{i,k-1} + [j+1] \cdot \Delta t) = s_{i,k-1} \mid X(t_{i,k-1} + j \cdot \Delta t) = s_{i,k-1}\}, \end{aligned}$$

and $\Delta t := (t_{i,k} - t_{i,k-1})/N$. Hence, the likelihood for case (iii) is given by

$$f_{s_1, \dots, s_n}(\mathbf{s}_1, \dots, \mathbf{s}_n) = \prod_{i=1}^n [\boldsymbol{\pi}]_{s_{i,1}} \cdot \mathbf{w}^*(t_{i,1}, t_{i,2}) \cdots \mathbf{w}^*(t_{i,n_i-1}, t_{i,n_i}), \quad (**)$$

where

$$\mathbf{w}^*(t_{i,k-1}, t_{i,k}) := \begin{cases} p_{s_{i,k-1}}(t_{i,k-1}, t_{i,k}) \cdot [\mathbf{Q}^*(t_{i,k})]_{s_{i,k-1}, s_{i,k}}, & s_{i,k-1} \neq s_{i,k} \\ p_{s_{i,k-1}}(t_{i,k-1}, t_{i,k}), & s_{i,k-1} = s_{i,k} \end{cases}.$$

For a sufficiently large N , it is feasible to estimate $p_{s_{i,k-1}}(t_{i,k-1}, t_{i,k})$; however, for a forward-only Markov model (as is the case here and in Section 3.3), it turns out that

$$p_{s_{i,k-1}}(t_{i,k-1}, t_{i,k}) = [\mathbf{P}(t_{i,k-1}, t_{i,k})]_{s_{i,k-1}, s_{i,k-1}},$$

which can be estimated using either approach (A) or (E).

Results

Figures 1 and 2, corresponding to cases (i) and (ii), respectively, demonstrate the bias in the Aalen-Johansen and Nelson-Aalen estimators (i.e., approach (E)) due to violations of the assumptions that all observation times are transition times and all transitions are observed. Alternatively, Figure 3 illustrates the consistency of parameter estimation from approach (E) when the model is correctly specified, as in case (iii). Moreover, this justifies and accredits our proposed mechanism for associating parameter estimates with the Nelson-Aalen estimator. In Figures 1, 2, and 3, the unbiased parameter estimates for approach (A) indicate its robustness to assumptions about the observed transitions, so long as the likelihood function is properly specified.

Part (II)

Using the exact same simulated data and modeling structure as found in Section 3.3 of the manuscript, here, we compare the computational time differences between using approach (A) with `deSolve` as opposed to with `prodint`. In order to numerically compute the transition probability matrix between two time points, say t_k and t_{k+1} ($t_{k+1} > t_k$), with `prodint`, it requires defining a partition of the interval $[t_k, t_{k+1}]$ over which to numerically integrate. Define

$$\Delta t_k := \frac{t_{k+1} - t_k}{s},$$

where s is a fixed, positive integer defining the resolution of the partition. Recall that applying approach (A) with `deSolve` does *not* require defining a partition. Thus, for a fixed data set, the likelihood (equation (7) from the manuscript) is computed using approach (A) with `deSolve`, as well as computed using approach (A) with `prodint` for various partitions. See Figure 4. For a coarser partition, `prodint` computes a likelihood with a larger deviation away from the likelihood computed using `deSolve`. Additionally, as the partition becomes finer, the computation time using `prodint` increases almost linearly. Hence, because `prodint` requires defining a partition, it quickly becomes computationally infeasible compared to using `deSolve`; however, aside from computation time, we see that the two numerical-integration techniques lead to nearly the exact same likelihood computation, for a sufficient resolution of the partition used for `prodint`.

References

Williams, J. P., Storlie, C. B., Therneau, T. M., Jack Jr, C. R. & Hannig, J. (2020), ‘A Bayesian approach to multi-state hidden Markov models: application to dementia progression’, *Journal of the American Statistical Association* **115**(529), 16–31.

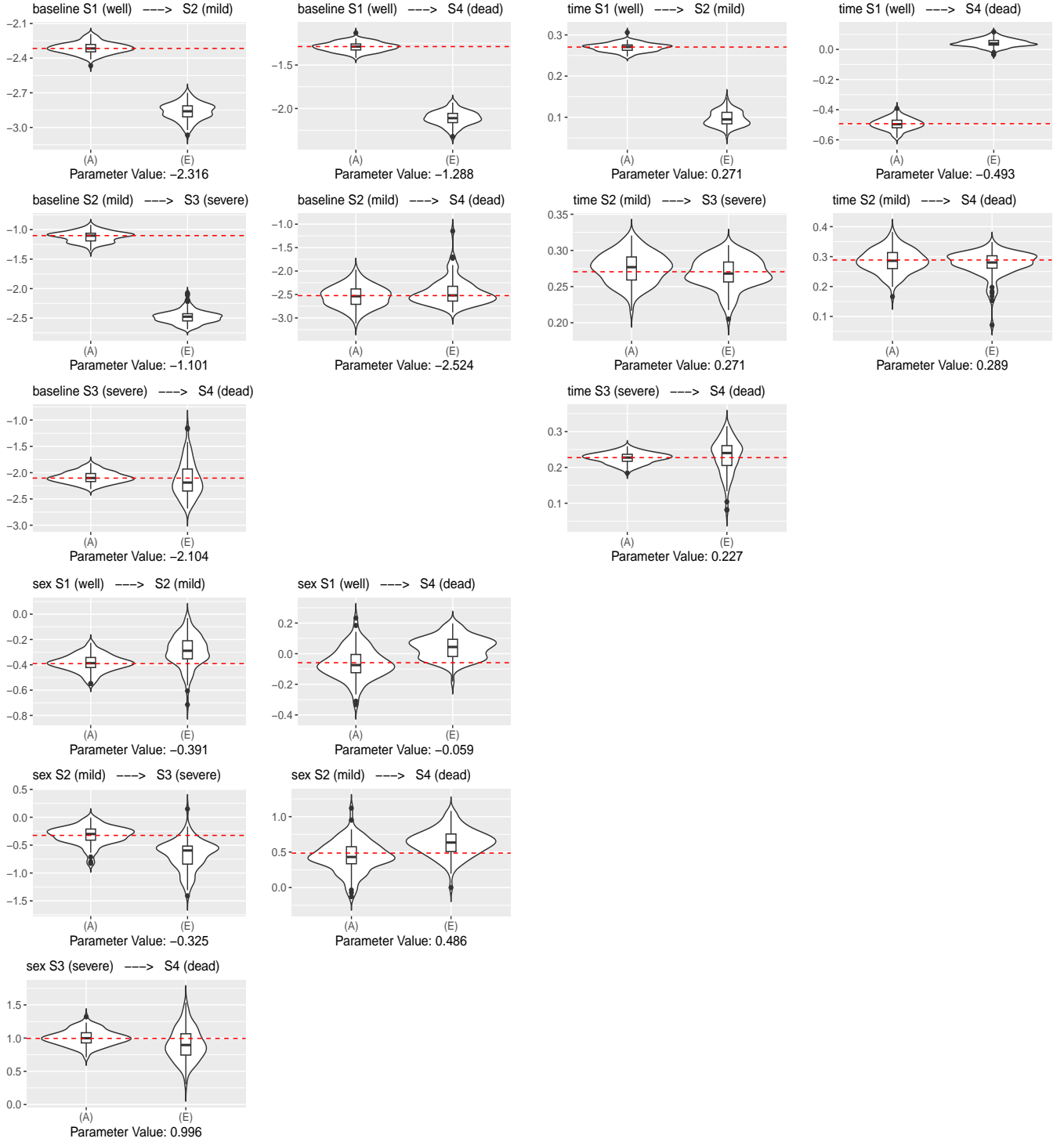


Figure 1: Violin plots of the estimated transition rate parameters from applying approaches (A) and (E) to the data for case (i). The posterior means of the 100 simulated data sets are plotted for approach (A), and the point estimates for each data set are plotted for approach (E).

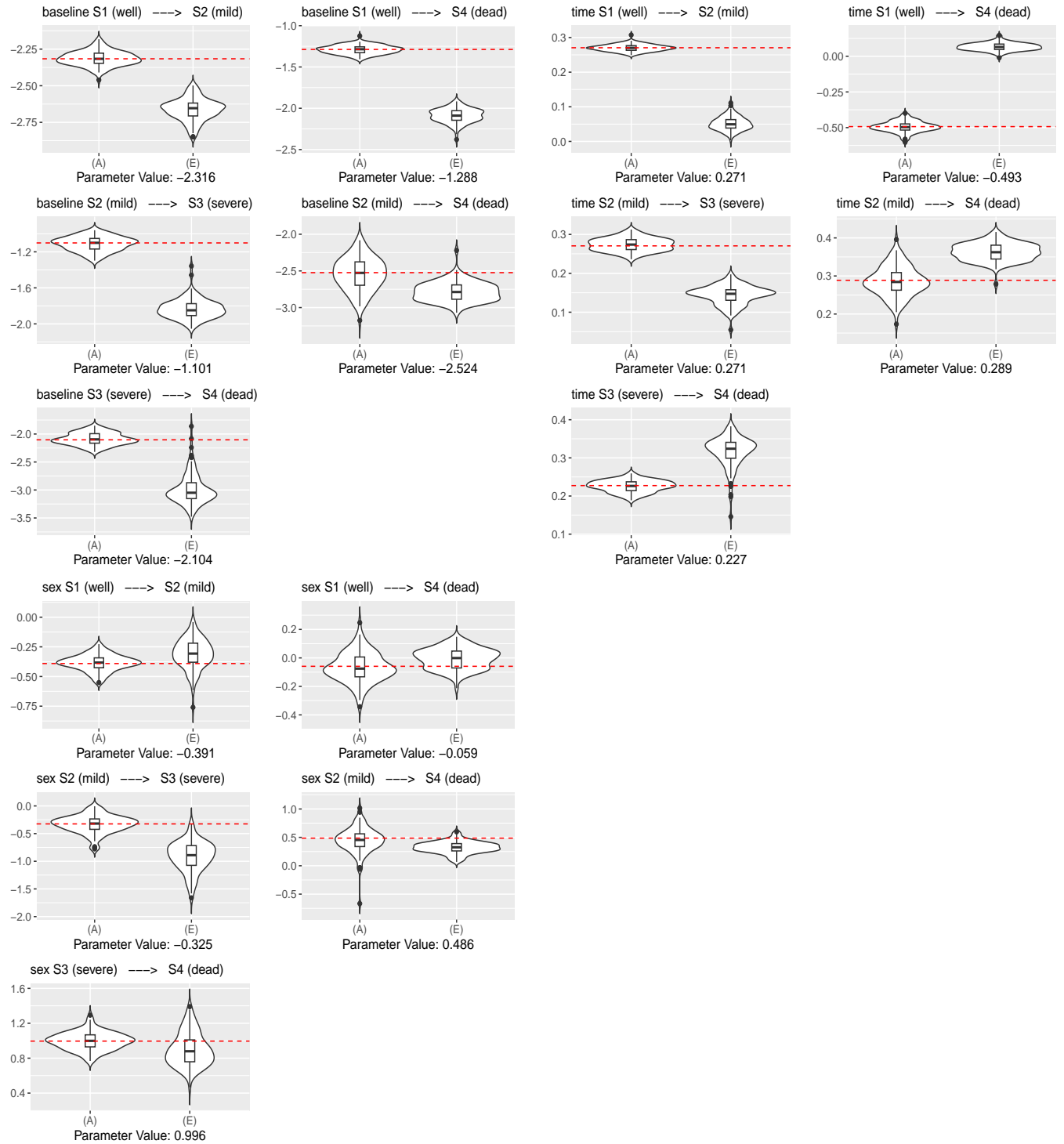


Figure 2: Violin plots of the estimated transition rate parameters from applying approaches (A) and (E) to the data for case (ii). The posterior means of the 100 simulated data sets are plotted for approach (A), and the point estimates for each data set are plotted for approach (E).

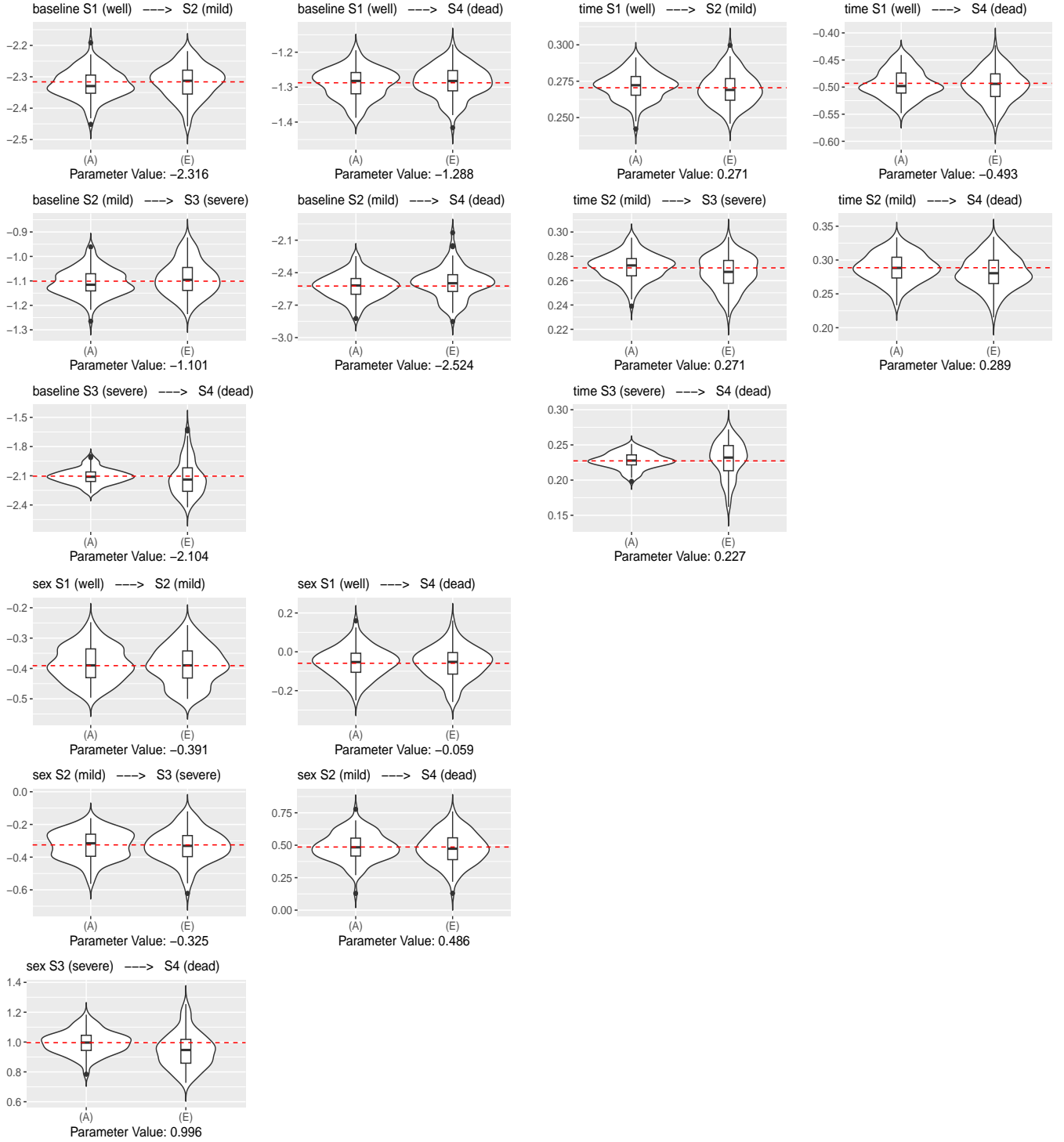


Figure 3: Violin plots of the estimated transition rate parameters from applying approaches (A) and (E) to the data for case (iii). The posterior means of the 100 simulated data sets are plotted for approach (A), and the point estimates for each data set are plotted for approach (E).

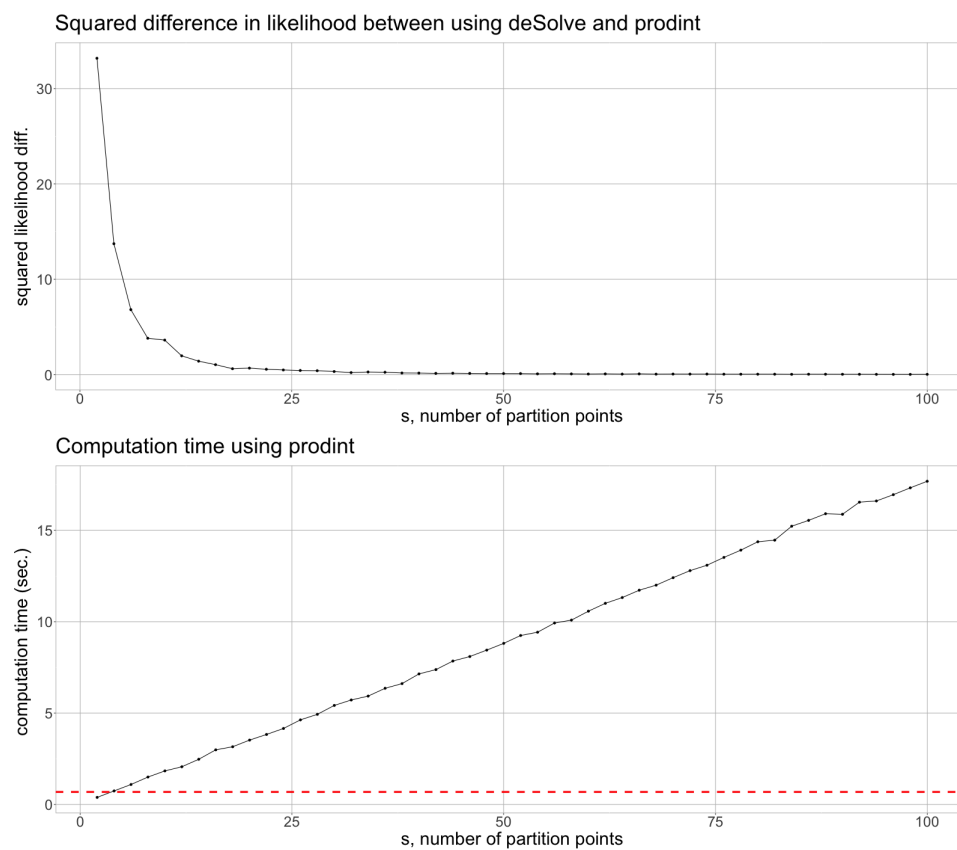


Figure 4: Differences between using approach (A) with `deSolve` compared to with `prodint`. Note that the red dotted line in the bottom figure represents the computation time for evaluating the likelihood using `deSolve`.