**Appendices**

**Appendix A. Equation for Relative Entropy**

The equation and syntax for calculating relative entropy were provided by Brilleman (2019 p.42; personal communications). Relative entropy is calculated as

,

where is the estimated posterior probability of individual *i* (*i* = 1…, *N*) being in latent class *g* (*g* = 1…, *G*).

**Appendix B. Estimation of Equivalent Joint Models without Latent Class Structure**

Shared random-effects modelling (SREM; see Rizopoulos, 2012) is an alternative approach to joint modelling that assumes a homogenous distribution of hazard. In this approach, random effects from a longitudinal model are included as covariates of a survival model.

In order to account for error in measurement and incomplete knowledge of the full trajectory of the longitudinal variable (dynamic risk in our project), SREM joint models use observed values to estimate the true values of the longitudinal variable. These are used to reconstruct a complete longitudinal history for each individual to inform the hazard model.

We tested predictive ability of SREM models with identical covariates, link functions, and baseline hazard distributions as our JLCM models.

**Appendix C. Assumption Testing: Variability of Slopes and Intercepts**

A fundamental criticism of group-based trajectory analysis is that statistical programs are likely to identify multiple trajectory groups, even when the data reflects a single underlying mean trajectory (Erosheva et al., 2014; Bauer & Curran, 2003). According to the framework suggested by Qureshi and Fang (2011) one must demonstrate that individuals differ both in terms of starting points (intercepts) and growth over time (slopes) before determining that a group-based trajectory method is appropriate. In the context of this project, we sought to establish that different individuals entered the community with different levels of risk and that there were between-individual differences in how risk changed over time. We sought to demonstrate this using multilevel modelling, as it is readily able to accommodate the large number of reassessments and unbalanced measurement schedules.

Using the lme4 package (Bates et al. 2015) for R (R Core Team 2015), we first compared model fit for linear and higher order models of the relationship between DRAOR subscale scores and time. Weeks were selected as our meaningful units of time. For each DRAOR subscale, mixed-models were run using linear, quadratic, and cubic specifications of time. For each subscale, higher order specifications of time resulted in optimal fit (see **C.1**).

After selecting an appropriate representation of time, we then examined the variance component of intercept-only models to establish whether significant variation around intercepts was present. To establish whether individual differences existed in change in risk over time, we compared models with fixed slopes to models where slopes were allowed to vary across individuals. For each of the DRAOR subscales, allowing slopes to vary between individuals resulted in significant improvements of model fit (see **C.2**). These results indicate that there is significant variation in both initial risk levels and change in risk over time. Therefore it was appropriate to move forward with testing whether distinct classes of risk trajectories were identifiable based on this unobserved heterogeneity.

**Appendix D. Selection of Initial Values for Joint Latent Class Analyses**

Latent class joint modelling in lcmm uses a maximum likelihood framework. Specifying appropriate initial values is critical, as it reduces computation time (see appendix of Brilleman et al., 2019) and increases the likelihood that convergence reflects a *global* maxima rather than a local maxima (Proust-Lima et al., 2017). Our approach for specifying initial values was to first estimate a model with the same structure as our intended joint model but with a single class. The parameter estimates from this model were used along with lcmm’s gridsearch function to randomly generate three sets of initial values (argument ‘rep = 3’), each ran for 50 maximum iterations (argument ‘maxiter=50’). In cases where this was not sufficient to facilitate model convergence, the number of reps was increased to 5. If models still failed to converge, we increased reps to 7 and maximum iterations to 150.

**Appendix E. Model Selection Outcomes for DRAOR Stable and DRAOR Protect**

Model selection criteria for DRAOR Stable and Protect trajectories are presented in **E.1**. For DRAOR Stable subscales, BIC values continued to drop as more groups were added. We eliminated the six-group solution as group membership fell below 1% for one of the groups. To select between four- and five- group solutions we considered trajectory plots and chose to use the more parsimonious four-group solution. Trajectory estimates from the test sample did not match the calibration sample (increasing group not replicated). We therefore tested the five-group solution across the split sample. Trajectories from the calibration sample were consistent with the test sample (see Appendix F) and the five-group solution was upheld.

Models for DRAOR Protect had progressively better fit as indicated by BIC as more groups were added. Likewise, relative entropy increased as more classes were added. The six-group solution failed to converge. We therefore selected a five-group solution.

**Appendix F. Split Sample Model Comparisons**

For each of the DRAOR subscales, the models selected from the 60% calibration sample (calibration sample) were applied to the remaining 40% of the sample (test sample) for visual comparison. We accepted the model from the calibration sample if the respective intercepts, trajectory shapes, and proportion of participants identified were replicated in the test sample. Comparisons for the selected models are presented in **F.1.**

For the four-group solution of DRAOR Stable, the intercepts, shapes, and proportion of participants allotted for three of the groups identified were similar, but the fourth, increasing group from the calibration sample was not replicated. We therefore rejected the four-group solution and revisited the five-group solution. For this solution, trajectories identified in the test sample were consistent with those identified in the calibration sample. The proportion of participants designated to each group was also similar across samples. We therefore upheld the five-group solution.

For DRAOR Acute, the proportion of participants designated into the first two groups was consistent across calibration and test samples, and the shapes and intercepts of the predicted trajectories were very similar. For this reason, we upheld the four-group solution.

For DRAOR Protect, the five-group solution yielded similar trajectories across calibration and test samples. The intercepts and trajectory shapes estimated were similar, and similar proportions of participants were identified as belonging in each trajectory group. We upheld the five-group solution accordingly.

**Appendix G. Descriptive Statistics of Selected DRAOR Stable Model**

Among the five trajectory groups identified for DRAOR Stable scores some differed in both intercepts and trajectory shapes, while others differed primarily in intercepts (see **G.1**). Each group was associated with different survival and hazard curves. Two groups had roughly parallel trajectories: one characterized by high initial DRAOR Stable scores that stayed high (*high stable* group, about 17% of participants; see **G.2** for sample trajectories within groups and **G.3** for descriptive statistics) and one characterized by low initial DRAOR Stable scores that decreased slightly (*low decreasing* group, about 8% of participants). A third group was characterized by moderate initial DRAOR Stable scores that decreased consistently across follow up (*moderate decreasing* group, about 70% of participants). A fourth group had initially high DRAOR Stable scores which decreased rapidly (*rapid decreasing* group, about 4% of participants). A fifth group had low initial DRAOR scores that increased over time (*increasing* group, about 1% of participants).

The *high stable*, *rapid decreasing*, and *increasing* groups had relatively high rates of recidivism and were assessed in the community for similar amounts of time. Participants designated to the *increasing* group had the highest rate of any recidivism and higher rates of each type of recidivism relative to other groups, despite having lower initial DRAOR scores. This group was the only to show substantial deterioration, and also had the highest mean net change, indicating greater fluctuation in scores from week to week. *High stable* and *rapid decreasing* groups also had elevated mean net change in DRAOR Acute scores relative to *moderate decreasing* and *low decreasing* groups.

The *moderate decreasing* group and *low decreasing* group each had lower rates of recidivism than the other three groups. The *low decreasing* group had substantially less recidivism and was assessed in the community for longer than the *moderate decreasing* group despite the *moderate decreasing* group being older, with lower static risk. Conversely, the *low decreasing* group had lower initial risk as indicated by DRAOR scores, and had less fluctuation in DRAOR scores as indicated by mean net change.

**Appendix H. Descriptive Statistics of selected DRAOR Protect model**

Of the five trajectory groups of DRAOR Protect scores identified, three reflected scores that increased over time (see **H.1**). The first trajectory group was characterized by high initial Protect scores that increased somewhat over time (*high increasing*; about 57% of participants; see **H.2** for sample trajectories within groups). The participants in this group were older, spent more time in the community, and had low rates of any recidivism relative to other groups and sample averages. They had low risk at the time of re-entry and improved moderately across subscales, with low mean net change (descriptive statistics in **H.3**).

The remaining two increasing groups were characterised by rapid increases in DRAOR Protect scores. One started with slightly higher initial Protect scores (*moderate-rapid increasing* group, about 7% of participants) and one started with low initial Protect scores (*low-rapid increasing* group, about 1.5% of participants. Participants in the *low-rapid increasing* group fared worse than those in the *moderate-rapid increasing* group, with less time spent in the community and higher rates of recidivism, although they demonstrated greater improvement in DRAOR Acute and DRAOR Protect scores and had lower RoC\*RoI scores. Participants in the *low-rapid increasing* group had high mean net change across subscales, indicating more fluctuation in scores across weeks.

Another trajectory group was characterized by moderate initial Protect scores that changed little (*moderate stable* group, about 34% of participants). These participants had slightly higher rates of recidivism than the *moderate-rapid increasing* group, with similar levels of mean net change. Of all groups identified, participants in this group had the highest static risk scores.

Finally, one group was characterised by decreasing Protect scores (*decreasing* group, about 1% of participants). This group had similar outcomes to the *low-rapid* *increasing* group, with relatively high recidivism rates and little time spent in the community. This was the only group to demonstrate deterioration over time, and they had a high degree of mean net change.

**Appendix I Dynamic Predictions and Comparison of Predictive Ability for DRAOR Stable and DRAOR Protect**

As with DRAOR Acute models, there was little difference between DRAOR Stable models with and without latent classes during the early prediction window (**I.1**), whereas models without latent class structures outperformed those with latent classes during the second prediction window (12 through 20 weeks), and, to a lesser extent, during the third prediction window (24 through 30 weeks). For DRAOR Protect, the models without latent class structures outperformed the latent class models across all prediction windows.

**C.1**

*Model Fit Comparisons for Linear and Higher Order Specifications of Time and DRAOR Scores*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Model 1: Random intercept, random linear slope | Model 2: Random intercept, random linear slope, fixed quadratic term | Model 3: Random intercept, random linear slope, fixed quadratic term, fixed cubic term |
| AIC/BIC | 261569/261625 | 259263/259329 | 258941/259017 |
| χ2 Δ from Model 1 | - | 2307.55\*\*\* | 2631.43\*\*\* |
| χ2 Δ from Model 2 | - | - | 323.87\*\*\* |
| χ2 Δ from Model 3 | - | - | - |
| AIC/BIC | 290998/291055 | 288977/289043 | 288544/288620 |
| χ2 Δ from Model 1 | - | 2023.42\*\*\* | 2458.13\*\*\* |
| χ2 Δ from Model 2 | - | - | 434.71\*\*\* |
| χ2 Δ from Model 3 | - | - | - |
| AIC/BIC | 259819/259876 | 257118/257184 | 256682/256757 |
| χ2 Δ from Model 1 | - | 2703.38\*\*\* | 3141.14\*\*\* |
| χ2 Δ from Model 2 | - | - | 437.77\*\*\* |
| χ2 Δ from Model 3 | - | - | - |

Note. N =92104 observations of 3421 individuals. DRAOR = Dynamic Risk Assessment for Offender Re-Entry (Serin, 2007). AIC = Akaike Information Criteria. BIC = Bayesian Information Criteria.

p\* < .05, p\*\* < .01, p\*\*\* < .001

**C.2**

*Multilevel Models Demonstrating Variability around Slopes and Intercepts*

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Variable | Model 1: Random Intercept | | Model 2: Random Intercept and Fixed Slope | | Model 3: Random Intercept and Random Slope | |
| B (SE) | *t*- value | B (SE) | *t*- value | B (SE) | *t*- value |
| |  | | --- | | **DRAOR Stable (*N* = 75917 observations from 3416 individuals)** | | | | | | | |
| Intercept | 5.68 (0.05) | 126 | 5.55 (0.04) | 127.03 | 5.56 (0.04) | 123.80 |
| Linear slope (Time in weeks) | - | - | -108.20 (1.30) | -82.97 | -105.99 (4.97) | -21.32 |
| Quadratic slope (Time in weeks, squared) | - | - | 26.89 (1.22) | 22.11 | 29.86 (3.15) | 9.47 |
| **Model Fit** |  | |  | |  | |
| AIC/BIC | 261465/261492 | | 254519/254566 | | 213267/213360 | |
| ICC/ χ2 Δ from prior model | ICC = 0.82 | | χ2 Δ = 6949.2\*\*\* | | χ2 Δ = 41262.3\*\*\* | |
| |  | | --- | | **DRAOR Acute (*N* = 75917 observations from 3416 individuals)** | | | | | | | |
| Intercept | 4.97 (0.04) | 136.5 | 4.82 (0.03) | 138.37 | 4.81 (0.04) | 136.67 |
| Linear slope (Time in weeks) | - | - | -130.81 (1.42) | -92.42 | -127.24 (1.42) | -89.68 |
| Quadratic slope (Time in weeks, squared) | - | - | 37.53 (1.32) | 28.41 | 44.98 (3.46) | 12.98 |
| **Model Fit** |  | |  | |  | |
| AIC/BIC | 273815/273843 | | 265119/265165 | | 256115 256207 | |
| ICC/ χ2 Δ from prior model | ICC = 0.71 | | χ2 Δ = 8700.7\*\*\* | | χ2 Δ = 9013.9\*\*\* | |
| **DRAOR Protect (*N* = 75917 observations from 3416 individuals)** | | | | | | |
| Intercept | 6.81 (0.04) | 165.7 | 6.95 (0.04) | 176.37 | 6.95 (0.04) | 171.09 |
| Linear slope (Time in weeks) | - | - | 117.35 (1.26) | 92.84 | 120.36 (4.67) | 25.78 |
| Quadratic slope (Time in weeks, squared) | - | - | -32.32 (1.18) | -27.42 | -34.06 (3.04) | -11.22 |
| **Model Fit** |  | |  | |  | |
| AIC/BIC | 258044/258072 | | 249329/249375 | | 213681/213774 | |
| ICC/ χ2 Δ from prior model | ICC = 0.80 | | χ2 Δ = 8719.3\*\*\* | | χ2 Δ = 35657.8\*\*\* | |

Note. N = 75917. DRAOR = Dynamic Risk Assessment for Offender Re-Entry (Serin, 2007). AIC = Akaike Information Criteria. BIC = Bayesian Information Criteria. Quadratic specification was used instead of cubic, as singularity became an issue with the cubic specification of the DRAOR Stable subscale.

p\* < .05, p\*\* < .01, p\*\*\* < .001

**E.1**

*DRAOR Stable and DRAOR Protect Model Selection Criteria Derived From 60% Calibration Sample*

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Model | BIC | Max LL | CI Test Statistic | Relative Entropy | Mean Posterior Probability within Class (Percentage) | | | | | |
|  |  |  | 1 | 2 | 3 | 4 | 5 | 6 |
| DRAOR Stable |  |  |  |  |  |  |  |  |  |  |
| 2 classes | 154675.32 | -77257.59 | 0.196 | 0.48 | 0.86 (51.17) | 0.82 (48.83) |  |  |  |  |
| 3 Classes | 154324.20 | -77059.15 | 0.79 | 0.66 | 0.86 (54.34) | 0.82 (43.22) | 0.89 (2.44) |  |  |  |
| 4 Classes | 154236.00 | -76992.18 | 0.997 | 0.72 | 0.85 (50.73) | 0.82 (45.41) | 0.88 (2.83) | 0.86 (1.02) |  |  |
| 5 Classes | 154082.54 | -76892.57 | 2.11 | 0.71 | 0.85 (52.2) | 0.78 (37.76) | 0.78 (7.37) | 0.86 (1.51) | 0.84 (1.17) |  |
| 6 Classes | 154008.22 | -76832.53 | 2.07 | 0.73 | 0.84 (50.20) | 0.76 (37.17) | 0.76 (9.22) | 0.87 (1.61) | 0.84 (1.32) | 0.93 (0.49) |
| DRAOR Protect |  |  |  |  |  |  |  |  |  |  |
| 2 classes | 152242.05 | -76040.96 | 1.92 | 0.51 | 0.86 (51.32) | 0.84 (48.68) |  |  |  |  |
| 3 classes | 151992.62 | -75893.37 | 2.20 | 0.66 | 0.86 (51.07) | 0.82 (45.76) | 0.83 (3.17) |  |  |  |
| 4 classes | 151910.65 | -75829.5 | 0.23 | 0.69 | 0.87 (56.05) | 0.78 (36.44) | 0.77 (6.20) | 0.84 (1.32) |  |  |
| 5 classes | 151779.84 | -75741.22 | 0.12 | 0.71 | 0.85 (52.49) | 0.77 (37.51) | 0.80 (7.27) | 0.86 (1.46) | 0.86 (1.27) |  |
| *Note.* Calibration sample comprised of 51633 observations of N = 2050 randomly selected participants. There were 787 recidivism events recorded in this sample.  BIC indicates Bayesian information criterion  \*indicates significant conditional independence test statistic | | | | | | | | | | |

|  |  |
| --- | --- |
| **F.1** |  |
| A. Predicted DRAOR Stable Trajectories for 60% Calibration Sample | B. Predicted DRAOR Stable Trajectories for 40% Test Sample |
|  |  |
| C. Predicted DRAOR Acute Trajectories for 60% Calibration Sample | D. Predicted DRAOR Acute Trajectories for 40% Test Sample |
|  |  |
| E. Predicted DRAOR Protect Trajectories for 60% Calibration Sample | F. Predicted DRAOR Protect Trajectories for 40% Test Sample |

Cross sample comparisons of identified trajectories. Calibration sample comprised of 51633 observations of *N* = 2050 randomly selected participants, with 787 recidivism events. Calibration sample comprised of 40312 observations of *N* = 1598 remaining participants, with 646 recidivism events.

**G.1**

|  |  |
| --- | --- |
| **A** | **B** |
| **C** | **D** |
| Mean group predicted trajectories of DRAOR Stable plotted with (**A**) and without (**B**) Monte Carlo confidence intervals, mean group survival curves (**C**) and mean group hazard curves (**D**). | |

**G.2**

|  |  |  |  |
| --- | --- | --- | --- |
| **A** | | **B** | |
| **C** | **D** | | **E** |
| Sample DRAOR Stable Trajectories with Mean Predicted Group Trajectories Superimposed. Heavily weighted line represents predicted mean trajectories of DRAOR Stable for (**A)** *Moderate Decreasing*, (**B**) *High Stable*, (**C**) *Low Decreasing*, (**D**) *Rapid Decreasing,* and (**E**) *Increasing*. Unweighted lines represent 50 randomly selected individual sample trajectories within each group (except *increasing* which only had 47 participants designated to it), jittered to reduce overlap. Jittering increases readability, facilitating illustration of within-group noise, but creates appearance of oscillation, where scores may be constant across measurement occasions. | | | |

**G.3**

Descriptive Qualities of Five DRAOR Stable Trajectories

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Moderate Decreasing | High Stable | Low Decreasing | Rapid Decreasing | Increasing |  |
|  | N (%) | | | | |  |
| Number of Participants | 2544 (69.74) | 617 (16.91) | 297 (8.14) | 143 (3.92) | 47(1.29) |  |
|  | Mean (SD) | | | | | Kruskal-Wallis χ2 |
| Age | 35.58 (11.81) | 33.79 (11.05) | 32.80 (10.51) | 31.55 (10.11) | 28.00 (10.03) | 42.91\*\*\* |
| Weeks follow up | 26.76 (16.46) | 13.06 (11.78) | 43.96 (11.82) | 14.76 (8.92) | 13.94 (7.98) | 761.38\*\*\* |
| RoC\*RoI | 0.48 (0.25) | 0.57 (0.23) | 0.61 (0.20) | 0.56 (0.22) | 0.60 (0.19) | 128.69\*\*\* |
| Baseline Stable | 5.94 (2.15) | 9.05 (1.82) | 4.14 (2.36) | 8.19 (2.42) | 4.57 (2.04) | 1046.2\*\*\* |
| Baseline Acute | 5.63 (2.23) | 7.58 (2.39) | 4.74 (2.19) | 6.83 (2.80) | 5.70 (2.11) | 377.83\*\*\* |
| Baseline Protect | 6.30 (2.20) | 4.32 (2.28) | 7.55 (2.32) | 5.16 (2.41) | 6.68 (2.49) | 452.56\*\*\* |
| Change Stable | -1.04 (2.25) | 0.20 (1.35) | -0.88 (2.40) | -4.54 (2.49) | 4.23 (1.59) | 585.94\*\*\* |
| Change Acute | -1.40 (2.46) | -0.70 (2.34) | -1.64 (2.45) | -2.81 (2.67) | 1.19 (2.50) | 142.36\*\*\* |
| Change Protect | 1.18 (2.26) | 0.13 (1.81) | 1.37 (2.41) | 2.71 (2.67) | -1.91 (3.02) | 275.89\*\*\* |
| Mean Net Change Stable | 0.21 (0.46) | 0.29 (0.72) | 0.12 (0.13) | 0.47 (0.45) | 0.69 (0.65) | 338.15\*\*\* |
| Mean Net Change Acute | 0.36 (0.48) | 0.50 (0.64) | 0.25 (0.18) | 0.46 (0.41) | 0.86 (0.65) | 95.05\*\*\* |
| Mean Net Change Protect | 0.21 (0.48) | 0.32 (0.69) | 0.13 (0.16) | 0.37 (0.42) | 0.58 (0.76) | 122.15\*\*\* |
|  | N (%) | | | | | Pearson’s χ2 |
| Any Recidivism | 891 (35.02) | 468 (75.85) | 41 (13.80) | 97 (67.83) | 38 (80.85) | 506.01\*\*\* |
| Technical Violations | 626 (24.61) | 348 (56.40) | 29 (9.76) | 74 (51.75) | 31 (65.96) | 356.35\*\*\* |
| Nonviolent Criminal Recidivism | 237 (9.32) | 136 (22.04) | 7 (2.36) | 21 (14.69) | 6 (12.77) | 107.58\*\*\* |
| Violent Recidivism | 153 (6.01) | 83 (13.45) | 5 (1.68) | 18 (12.59) | 9 (19.14) | 69.81\*\*\* |

Note. *N* = 3648 participants, assigned to groups based on posterior probabilities.

**H.1**

|  |  |
| --- | --- |
| **A** | **B** |
| **C** | **D** |
| Mean group predicted trajectories of DRAOR Protect plotted with (**A**) and without (**B**) Monte Carlo confidence intervals, mean group survival curves (**C**) and mean group hazard curves (**D**). | |

**H.2**

|  |  |  |  |
| --- | --- | --- | --- |
| **A** | | **B** | |
| **C** | **D** | | **E** |
| Sample DRAOR Protect Trajectories with Mean Predicted Group Trajectories Superimposed. Heavily weighted line represents predicted mean trajectories of DRAOR Protect for (**A)** *High Increasing*, (**B**) *Moderate Stable*, (**C**) *Moderate-Rapid Increasing*, (**D**) *Low-Rapid Increasing,* and (**E**) *Decreasing*. Unweighted lines represent 50 randomly selected individual sample trajectories within each group (except *decreasing* which only had 33 participants designated to it), jittered to reduce overlap. Jittering increases readability, facilitating illustration of within-group noise, but creates appearance of oscillation, where scores may be constant across measurement occasions. | | | |

**H.3**

Descriptive Qualities of Five DRAOR Protect Trajectories

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | High Increasing | Moderate Stable | Moderate-Rapid Increasing | Low-Rapid Increasing | Decreasing |  |
|  | N (%) | | | | |  |
| Number of Participants | 2069 (56.72) | 1240 (33.99) | 253 (6.94) | 53 (1.45) | 33 (0.90) |  |
|  | Mean (SD) | | | | | Kruskal-Wallis χ2 |
| Age | 37.43 (12.31) | 31.06 (9.25) | 32.20 (9.67) | 34.13 (12.78) | 33.70 (8.88) | 233.48\*\*\* |
| Weeks follow up | 34.06 (16.00) | 13.05 (9.95) | 18.17 (8.00) | 6.98 (3.20) | 9.79 (5.01) | 1330.10\*\*\* |
| RoC\*RoI | 0.40 (0.23) | 0.67 (0.16) | 0.60 (0.19) | 0.55 (0.22) | 0.59 (0.19) | 986.09\*\*\* |
| Baseline Stable | 5.61 (2.40) | 7.51 (2.27) | 7.13 (2.48) | 6.94 (2.50) | 6.24 (2.97) | 476.17\*\*\* |
| Baseline Acute | 5.31 (2.20) | 6.82 (2.39) | 6.52 (2.67) | 7.13 (2.53) | 6.33 (2.15) | 317.16\*\*\* |
| Baseline Protect | 6.88 (2.13) | 4.89 (2.23) | 4.83 (2.26) | 4.08 (2.49) | 7.73 (2.35) | 648.80\*\*\* |
| Change Stable | -1.27 (2.34) | 0.05 (1.83) | -2.54 (2.97) | -1.62 (2.86) | 1.79 (2.56) | 430.84\*\*\* |
| Change Acute | -1.69 (2.39) | -0.50 (2.36) | -2.35 (2.80) | -2.38 (2.44) | 0.30 (2.39) | 262.53\*\*\* |
| Change Protect | 1.40 (2.15) | -0.09 (1.58) | 3.68 (2.25) | 4.04 (2.72) | -4.73 (2.17) | 916.25\*\*\* |
| Mean Net Change Stable | 0.14 (0.25) | 0.36 (0.73) | 0.35 (0.54) | 0.60 (0.79) | 0.49 (0.53) | 159.84\*\*\* |
| Mean Net Change Acute | 0.27 (0.30) | 0.53 (0.67) | 0.53 (0.61) | 0.65 (0.74) | 0.64 (0.35) | 211.25\*\*\* |
| Mean Net Change Protect | 0.14 (0.28) | 0.31 (0.72) | 0.42 (0.41) | 0.94 (0.79) | 0.85 (0.51) | 456.01\*\*\* |
|  | N (%) | | | | | Pearson’s χ2 |
| Any Recidivism | 320 (15.47) | 977 (78.79) | 167 (66.01) | 44 (83.02) | 27 (81.81) | 1404.2\*\*\* |
| Technical Violations | 213 (10.29) | 716 (57.74) | 127 (50.20) | 32 (60.38) | 20 (60.61) | 917.46\*\*\* |
| Nonviolent Criminal Recidivism | 68 (3.29) | 271 (21.85) | 46 (18.18) | 14 (26.42) | 8 (24.24) | 303.21\*\*\* |
| Violent Recidivism | 54 (2.61) | 173 (13.95) | 24 (9.49) | 10 (18.87) | 7 (21.21) | 169.03\*\*\* |

*Note. N* = 3648 participants, assigned to groups based on posterior probabilities.

**I.1**

Predictive discrimination and calibration of selected joint latent class model and equivalent model without latent class structure for DRAOR Stable and Protect.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Prediction Window | | | | | |
|  | 4 weeks through 12 weeks | | 12 weeks through 20 weeks | | 24 weeks through 32 weeks | |
| Model | AUC\*100 | Brier Score\*100 | AUC \*100 | Brier Score\*100 | AUC \*100 | Brier Score\*100 |
| DRAOR Stable | | | | | | |
| Five class JLCM | 73.46 | 11.24 | 70.16 | 12.49 | 74.50 | 9.79 |
| Equivalent SREM | 73.44 | 10.87 | 73.63 | 11.28 | 75.28 | 9.94 |
|  | | | | | | |
| Five class JLCM | 70.85 | 13.29 | 68.88 | 13.76 | 71.60 | 10.22 |
| Equivalent SREM | 73.56 | 10.83 | 75.02 | 11.18 | 75.37 | 9.84 |

*Note.* Survival predictions from JLCM used Weibull baseline hazard stratified on probable group membership with RoC\*RoI as covariate. Survival predictions from SREM used Weibull baseline hazard with RoC\*RoI as covariate.