Using machine learning to identify patients at high risk of inappropriate drug dosing in periods with renal dysfunction

Supplement

Benjamin Skov Kaas-Hansen

2 January 2022

Contents

This document contains all supplementary tables and figures for the above-mentioned paper. The figures within each domain all share the same caption, and so to avoid cluttering captions are given only once, under the domain heading.

Supplementary tables

Table S1: Maximum doses in periods when eGFR is as indicated. ATC codes in brackets.

	Drug	eGFR thredshold	Max. daily dose
1	Toxicity issue	<NA>	<NA>
2	Apixaban (B01AF02)	<U+2264> 30	5 mg.
3	Dabigatran (B01AE07)	<U+2264> 30	0 mg.
4	Rivaroxaban (B01AF01)	<U+2264> 30	0 mg.
5	Metformin (A10BA02)	<U+2264> 30	0 mg.
6	Ibuprofene (M01AE01)	<U+2264> 30	0 mg.
7	Celecoxib (M01AH01)	<U+2264> 30	0 mg.
8	Methotrexate (L04AX03)	<U+2264> 30	0 mg.
9	Lithium citrate (N05AN01)	<U+2264> 30	0 mg.

Table S2 (extended version of table 1): Univariate summary statistics of the three data sets. Values are median (inter-quartile range) and N (%).

	variate	develop	test	test_new	in_table1
1	n	42250	10201	5980	yes
2	time_at_risk (median [IQR])	3.5 [1.7, 7.7]	3.5 [1.7, 7.2]	2.9 [1.5, 6.4]	yes
3	daily_rate_not_zero (%)	3786 (9.0)	1080 (10.6)	740 (12.4)	yes
4	daily_rate_geq_1 (%)	2241 (5.3)	588 (5.8)	333 (5.6)	yes
5	daily_rate_geq_2 (%)	1236 (2.9)	288 (2.8)	108 (1.8)	yes
6	daily_rate_geq_3 (%)	783 (1.9)	171 (1.7)	56 (0.9)	yes
7	daily_rate_geq_5 (%)	366 (0.9)	64 (0.6)	9 (0.2)	yes
8	inapp_dose_of_M01AE01 (%)	1713 (4.1)	343 (3.4)	244 (4.1)	yes
9	inapp_dose_of_M01AH01 (%)	67 (0.2)	21 (0.2)	18 (0.3)	yes
10	inapp_dose_of_B01AF01 (%)	341 (0.8)	174 (1.7)	115 (1.9)	yes
11	inapp_dose_of_A10BA02 (%)	1452 (3.4)	340 (3.3)	277 (4.6)	yes
12	inapp_dose_of_B01AE07 (%)	257 (0.6)	76 (0.7)	55 (0.9)	yes
13	inapp_dose_of_N05AN01 (%)	68 (0.2)	17 (0.2)	12 (0.2)	yes
14	inapp_dose_of_L04AX03 (%)	19 (0.0)	5 (0.0)	<5	yes
15	inapp_dose_of_B01AF02 (%)	124 (0.3)	184 (1.8)	73 (1.2)	yes
16	n_previous_admissions (%): 0	4988 (11.8)	1082 (10.6)	1074 (18.0)	yes
17	n_previous_admissions (%): 1-2	10100 (23.9)	2367 (23.2)	1873 (31.3)	yes
18	n_previous_admissions (%): 3-4	7712 (18.3)	1919 (18.8)	1232 (20.6)	yes
19	n_previous_admissions (%): 5-6	5490 (13.0)	1303 (12.8)	685 (11.5)	yes
20	n_previous_admissions (%): >6	13960 (33.0)	3530 (34.6)	1116 (18.7)	yes
21	n_distinct_drugs_before_index (%): 0	6165 (14.6)	1228 (12.0)	762 (12.7)	yes
22	n_distinct_drugs_before_index (%): 1-2	9111 (21.6)	1984 (19.4)	1254 (21.0)	yes
23	n_distinct_drugs_before_index (%): 3-4	8761 (20.7)	2078 (20.4)	1355 (22.7)	yes
24	n_distinct_drugs_before_index (%): 5-6	7197 (17.0)	1852 (18.2)	1095 (18.3)	yes
25	n_distinct_drugs_before_index (%): >6	11016 (26.1)	3059 (30.0)	1514 (25.3)	yes
26	admitted_tue (%)	7092 (16.8)	1755 (17.2)	1016 (17.0)	yes
27	admitted_mon (%)	7667 (18.1)	1803 (17.7)	1066 (17.8)	yes
28	admitted_thu (%)	6665 (15.8)	1626 (15.9)	927 (15.5)	yes
29	admitted_sat (%)	3809 (9.0)	960 (9.4)	606 (10.1)	yes
30	admitted_sun (%)	4140 (9.8)	989 (9.7)	598 (10.0)	yes
31	admitted_wed (%)	6718 (15.9)	1616 (15.8)	936 (15.7)	yes
32	admitted_fri (%)	6159 (14.6)	1452 (14.2)	831 (13.9)	yes
33	n_persons (%)	27253 (64.5)	8412 (82.5)	5341 (89.3)	yes
34	n_women (%)	13759 (32.6)	4049 (39.7)	2629 (44.0)	yes
35	n18_diag_any (%)	13470 (31.9)	3391 (33.2)	732 (12.2)	yes
36	icd10_chapter_4 (%)	19716 (46.7)	5096 (50.0)	2415 (40.4)	yes

## 37	icd10_chapter_9 (%)	25757 (61.0)	6392 (62.7)	3283 (54.9)	yes
## 38	icd10_chapter_14 (%)	23025 (54.5)	5819 (57.0)	2306 (38.6)	yes
## 39	icd10_chapter_18 (%)	18663 (44.2)	5711 (56.0)	2882 (48.2)	yes
## 40	icd10_chapter_19 (%)	20275 (48.0)	4749 (46.6)	2481 (41.5)	yes
## 41	atc_level2_A02 (%)	10635 (25.2)	2776 (27.2)	1407 (23.5)	yes
## 42	atc_level2_B01 (%)	11842 (28.0)	3181 (31.2)	1795 (30.0)	yes
## 43	atc_level2_C03 (%)	13966 (33.1)	3672 (36.0)	1951 (32.6)	yes
## 44	atc_level2_J01 (%)	14719 (34.8)	3257 (31.9)	1938 (32.4)	yes
## 45	atc_level2_N02 (%)	15740 (37.3)	4367 (42.8)	2506 (41.9)	yes
## 46	r (%): 0	38464 (91.0)	9121 (89.4)	5240 (87.6)	no
## 47	r (%): 1-2	1282 (3.0)	446 (4.4)	391 (6.5)	no
## 48	r (%): 3-4	670 (1.6)	200 (2.0)	134 (2.2)	no
## 49	r (%): 5-6	431 (1.0)	116 (1.1)	75 (1.3)	no
## 50	r (%): >6	1403 (3.3)	318 (3.1)	140 (2.3)	no
## 51	age_at_admission (median [IQR])	76.9 [66.6, 85.1]	77.4 [68.2, 85.5]	78.5 [69.3, 86.2]	no
## 52	hour_of_admission_cyclical (%): 0	3560 (8.4)	884 (8.7)	514 (8.6)	no
## 53	hour_of_admission_cyclical (%): 1-3	18460 (43.7)	4521 (44.3)	2727 (45.6)	no
## 54	hour_of_admission_cyclical (%): 4-6	12209 (28.9)	2886 (28.3)	1629 (27.2)	no
## 55	hour_of_admission_cyclical (%): 7-9	5367 (12.7)	1250 (12.3)	712 (11.9)	no
## 56	hour_of_admission_cyclical (%): 10-12	2654 (6.3)	660 (6.5)	398 (6.7)	no
## 57	sex_female (%)	20743 (49.1)	4854 (47.6)	2940 (49.2)	no
## 58	sex_male (%)	21507 (50.9)	5347 (52.4)	3040 (50.8)	no
## 59	n18_diag_N189 (%)	13228 (31.3)	3242 (31.8)	694 (11.6)	no
## 60	n18_diag_N185 (%)	1970 (4.7)	868 (8.5)	62 (1.0)	no
## 61	n18_diag_N184 (%)	385 (0.9)	223 (2.2)	23 (0.4)	no
## 62	n18_diag_N183 (%)	161 (0.4)	138 (1.4)	30 (0.5)	no
## 63	n18_diag_N181 (%)	44 (0.1)	34 (0.3)	<5	no
## 64	n18_diag_N182 (%)	61 (0.1)	36 (0.4)	9 (0.2)	no
## 65	elixhauser_score_ahrq (%): 0	10419 (24.7)	2261 (22.2)	1881 (31.5)	no
## 66	elixhauser_score_ahrq (%): 1-2	1900 (4.5)	432 (4.2)	273 (4.6)	no
## 67	elixhauser_score_ahrq (%): 3-4	2078 (4.9)	496 (4.9)	356 (6.0)	no
## 68	elixhauser_score_ahrq (%): 5-6	4725 (11.2)	1092 (10.7)	609 (10.2)	no
## 69	elixhauser_score_ahrq (%): >6	23128 (54.7)	5920 (58.0)	2861 (47.8)	no
## 70	icd10_chapter_1 (%)	10263 (24.3)	2662 (26.1)	954 (16.0)	no
## 71	icd10_chapter_2 (%)	10386 (24.6)	2748 (26.9)	1410 (23.6)	no
## 72	icd10_chapter_3 (%)	9226 (21.8)	2285 (22.4)	920 (15.4)	no
## 73	icd10_chapter_5 (%)	7673 (18.2)	1979 (19.4)	1059 (17.7)	no
## 74	icd10_chapter_6 (%)	7129 (16.9)	1992 (19.5)	992 (16.6)	no
## 75	icd10_chapter_7 (%)	11452 (27.1)	2901 (28.4)	1495 (25.0)	no
## 76	icd10_chapter_8 (%)	5392 (12.8)	1203 (11.8)	650 (10.9)	no
## 77	icd10_chapter_10 (%)	15905 (37.6)	4131 (40.5)	1930 (32.3)	no
## 78	icd10_chapter_11 (%)	16451 (38.9)	4136 (40.5)	1921 (32.1)	no
## 79	icd10_chapter_12 (%)	6220 (14.7)	1597 (15.7)	680 (11.4)	no
## 80	icd10_chapter_13 (%)	16131 (38.2)	4123 (40.4)	2107 (35.2)	no
## 81	icd10_chapter_15 (%)	159 (0.4)	25 (0.2)	11 (0.2)	no
## 82	icd10_chapter_16 (%)	38 (0.1)	11 (0.1)	5 (0.1)	no
## 83	icd10_chapter_17 (%)	1348 (3.2)	312 (3.1)	93 (1.6)	no
## 84	icd10_chapter_20 (%)	9 (0.0)	<5	<5	no
## 85	icd10_chapter_21 (%)	37872 (89.6)	9254 (90.7)	5221 (87.3)	no
## 86	atc_level2_A03 (%)	1819 (4.3)	395 (3.9)	249 (4.2)	no
## 87	atc_level2_A04 (%)	792 (1.9)	307 (3.0)	200 (3.3)	no
## 88	atc_level2_A06 (%)	2129 (5.0)	523 (5.1)	256 (4.3)	no
## 89	atc_level2_A07 (%)	701 (1.7)	185 (1.8)	95 (1.6)	no
## 90	atc_level2_A09 (%)	22 (0.1)	9 (0.1)	5 (0.1)	no
## 91	atc_level2_A10 (%)	5270 (12.5)	1500 (14.7)	814 (13.6)	no
## 92	atc_level2_A11 (%)	4752 (11.2)	1410 (13.8)	558 (9.3)	no
## 93	atc_level2_A12 (%)	8578 (20.3)	2330 (22.8)	1182 (19.8)	no
## 94	atc_level2_A16 (%)	6 (0.0)	5 (0.0)	0 (0.0)	no
## 95	atc_level2_B02 (%)	1271 (3.0)	295 (2.9)	184 (3.1)	no
## 96	atc_level2_B03 (%)	3513 (8.3)	946 (9.3)	385 (6.4)	no
## 97	atc_level2_B05 (%)	4884 (11.6)	1643 (16.1)	1074 (18.0)	no
## 98	atc_level2_C01 (%)	4318 (10.2)	1068 (10.5)	588 (9.8)	no
## 99	atc_level2_C02 (%)	563 (1.3)	155 (1.5)	45 (0.8)	no
## 100	atc_level2_C07 (%)	8491 (20.1)	2362 (23.2)	1189 (19.9)	no
## 101	atc_level2_C08 (%)	4606 (10.9)	1221 (12.0)	607 (10.2)	no
## 102	atc_level2_C09 (%)	4717 (11.2)	1183 (11.6)	751 (12.6)	no
## 103	atc_level2_C10 (%)	5698 (13.5)	1479 (14.5)	820 (13.7)	no
## 104	atc_level2_D01 (%)	23 (0.1)	9 (0.1)	<5	no

105	atc_level2_D07 (%)	29 (0.1)	8 (0.1)	<5	no
106	atc_level2_G03 (%)	83 (0.2)	29 (0.3)	19 (0.3)	no
107	atc_level2_G04 (%)	836 (2.0)	265 (2.6)	150 (2.5)	no
108	atc_level2_H01 (%)	84 (0.2)	15 (0.1)	12 (0.2)	no
109	atc_level2_H02 (%)	3867 (9.2)	1054 (10.3)	575 (9.6)	no
110	atc_level2_H03 (%)	1600 (3.8)	477 (4.7)	267 (4.5)	no
111	atc_level2_H05 (%)	236 (0.6)	62 (0.6)	5 (0.1)	no
112	atc_level2_J02 (%)	286 (0.7)	84 (0.8)	70 (1.2)	no
113	atc_level2_J04 (%)	22 (0.1)	<5	0 (0.0)	no
114	atc_level2_J05 (%)	244 (0.6)	65 (0.6)	36 (0.6)	no
115	atc_level2_J06 (%)	23 (0.1)	7 (0.1)	<5	no
116	atc_level2_L01 (%)	121 (0.3)	27 (0.3)	15 (0.3)	no
117	atc_level2_L02 (%)	166 (0.4)	61 (0.6)	34 (0.6)	no
118	atc_level2_L03 (%)	12 (0.0)	<5	<5	no
119	atc_level2_L04 (%)	964 (2.3)	207 (2.0)	46 (0.8)	no
120	atc_level2_M01 (%)	1628 (3.9)	401 (3.9)	308 (5.2)	no
121	atc_level2_M03 (%)	247 (0.6)	93 (0.9)	62 (1.0)	no
122	atc_level2_M04 (%)	2341 (5.5)	700 (6.9)	276 (4.6)	no
123	atc_level2_M05 (%)	99 (0.2)	24 (0.2)	17 (0.3)	no
124	atc_level2_N01 (%)	378 (0.9)	99 (1.0)	61 (1.0)	no
125	atc_level2_N03 (%)	1917 (4.5)	557 (5.5)	281 (4.7)	no
126	atc_level2_N04 (%)	174 (0.4)	42 (0.4)	22 (0.4)	no
127	atc_level2_N05 (%)	5284 (12.5)	1454 (14.3)	823 (13.8)	no
128	atc_level2_N06 (%)	3314 (7.8)	843 (8.3)	424 (7.1)	no
129	atc_level2_N07 (%)	260 (0.6)	87 (0.9)	42 (0.7)	no
130	atc_level2_P01 (%)	1456 (3.4)	268 (2.6)	125 (2.1)	no
131	atc_level2_R01 (%)	19 (0.0)	10 (0.1)	5 (0.1)	no
132	atc_level2_R03 (%)	627 (1.5)	191 (1.9)	113 (1.9)	no
133	atc_level2_R05 (%)	691 (1.6)	113 (1.1)	60 (1.0)	no
134	atc_level2_R06 (%)	974 (2.3)	262 (2.6)	96 (1.6)	no
135	atc_level2_S01 (%)	169 (0.4)	40 (0.4)	22 (0.4)	no
136	atc_level2_V03 (%)	2827 (6.7)	748 (7.3)	279 (4.7)	no
137	atc_level2_V08 (%)	37 (0.1)	6 (0.1)	5 (0.1)	no

Supplementary figures

Figure S1: Mapping of admission hour, two alternatives.

Top: Mapping of admission hour (x axis) to unit circle (blue) and distance from midday (red). Bottom: The unit-circle mapping loses the interval nature of the admission hour on the original scale, whereas the distance-to-midday maintains this quality.

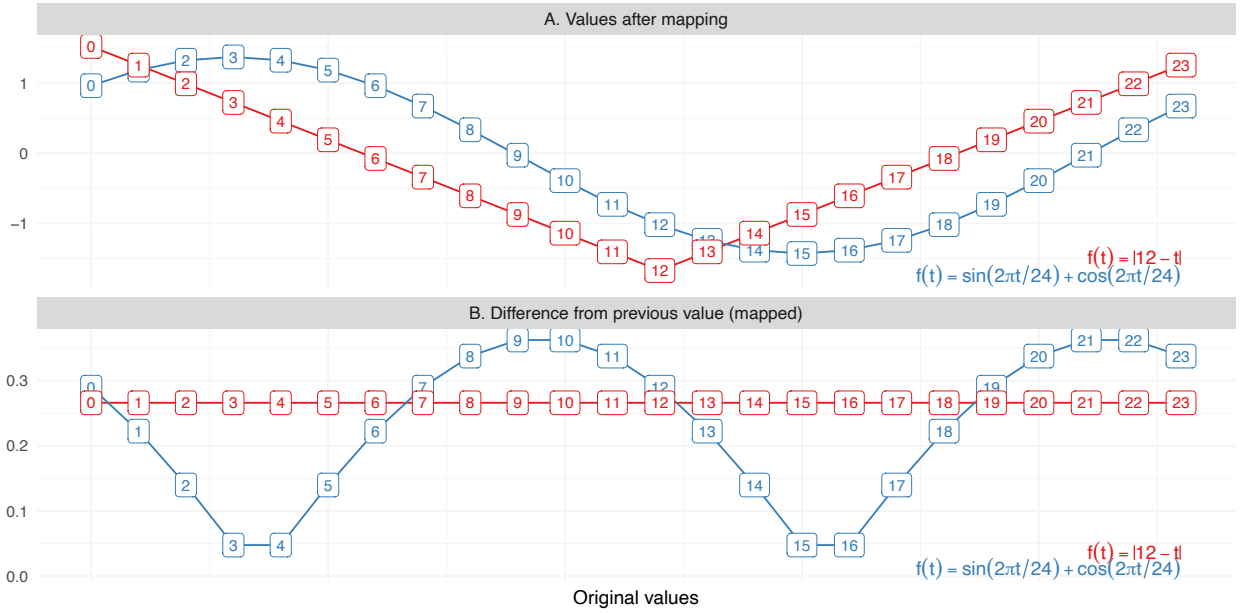

A. Values after mapping
1
0
−1
f(t) = |12 − t|
f(t) = sin(2πt/24) + cos(2πt/24)
B. Difference from previous value (mapped)
0.3
0.2
0.1
0.0
f(t) = |12 − t|
f(t) = sin(2πt/24) + cos(2πt/24)
Original values

Figure S2: The so-called rulegraph of the our Snakemake pipeline illustrating the end-to-end workflow with dependencies between processing, training and visualisation steps.

Optuna hyperparameter optimisation

Sampled hyperparameter configurations. Point colours represent the loss value: blue = low (preferable), red = high, grey = above the third quartile. Dashed lines indicate the best configuration.

Figure S3: MLP, daily rate > 0.

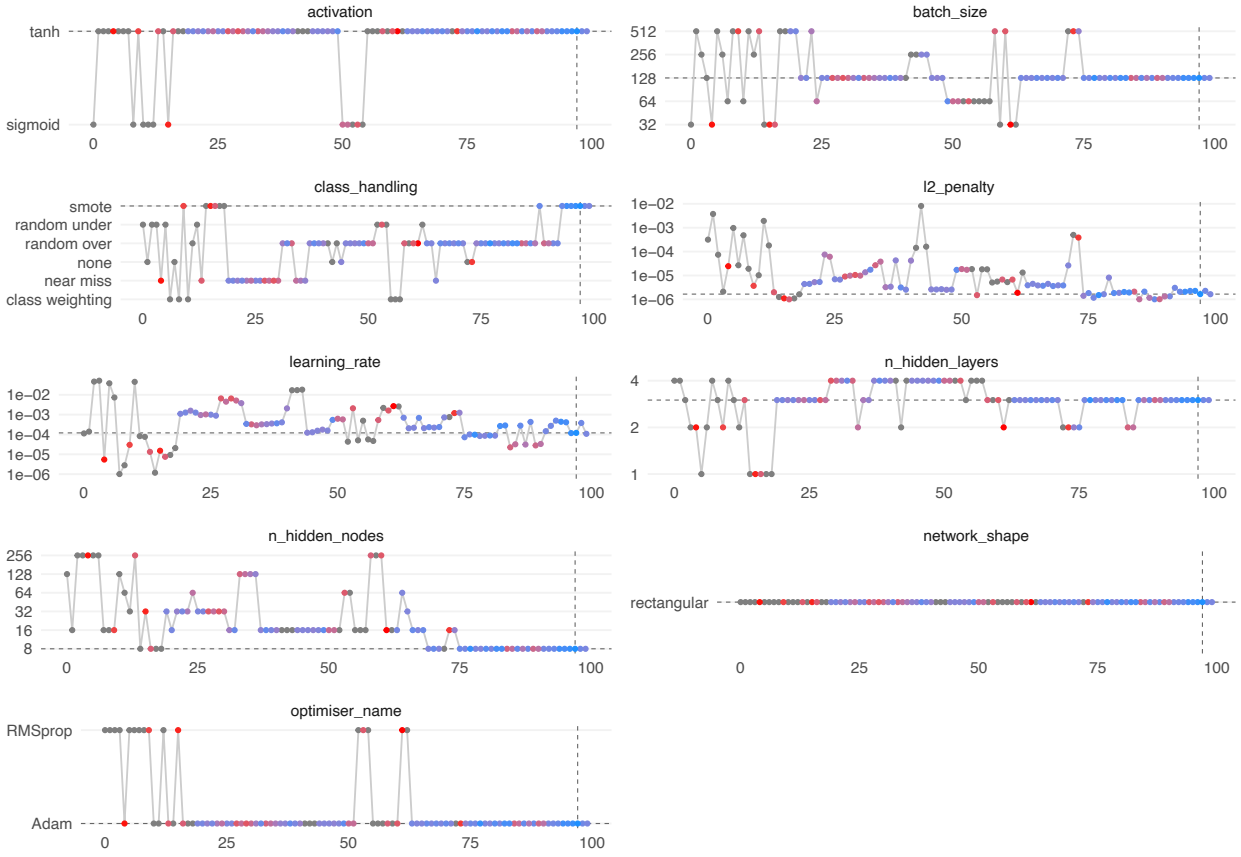


Figure S4: Linear, daily rate > 0.

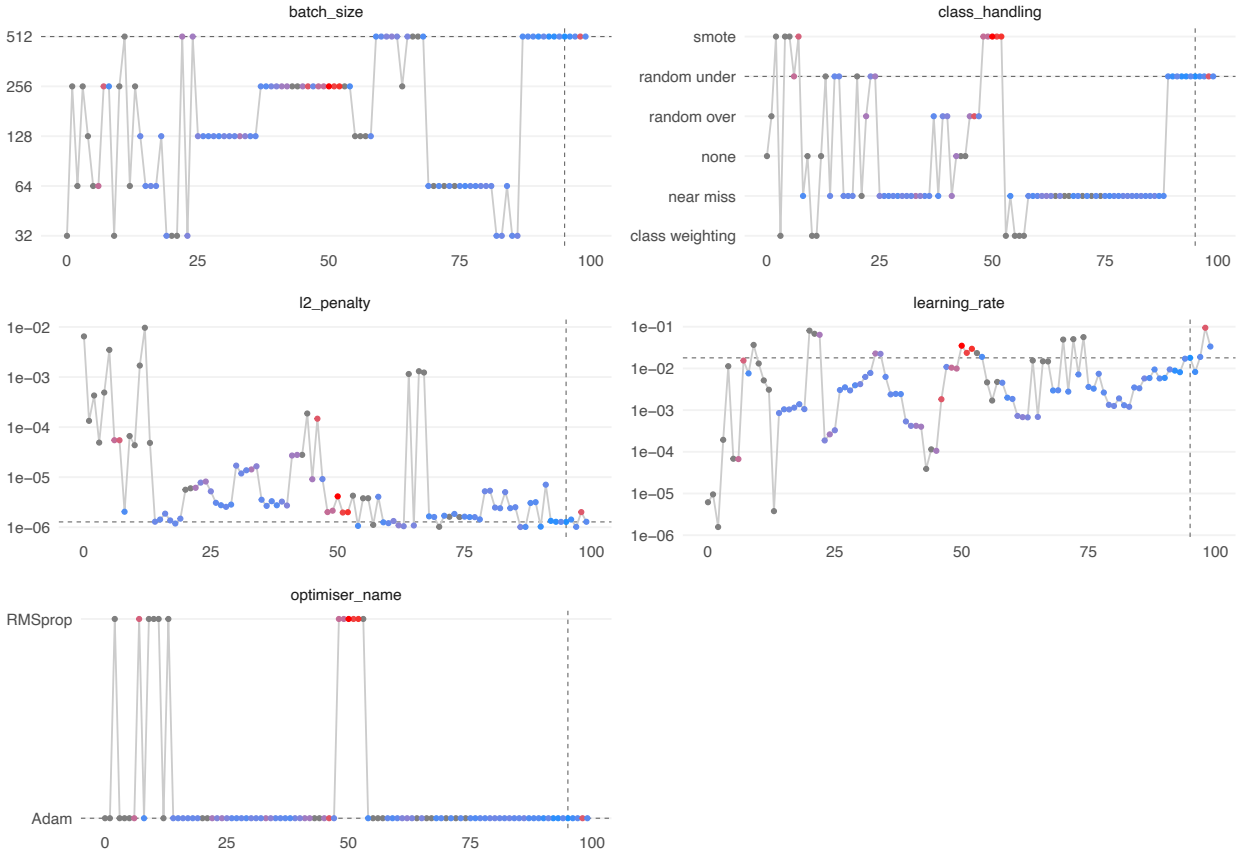


Figure S5: MLP, daily rate >= 1.

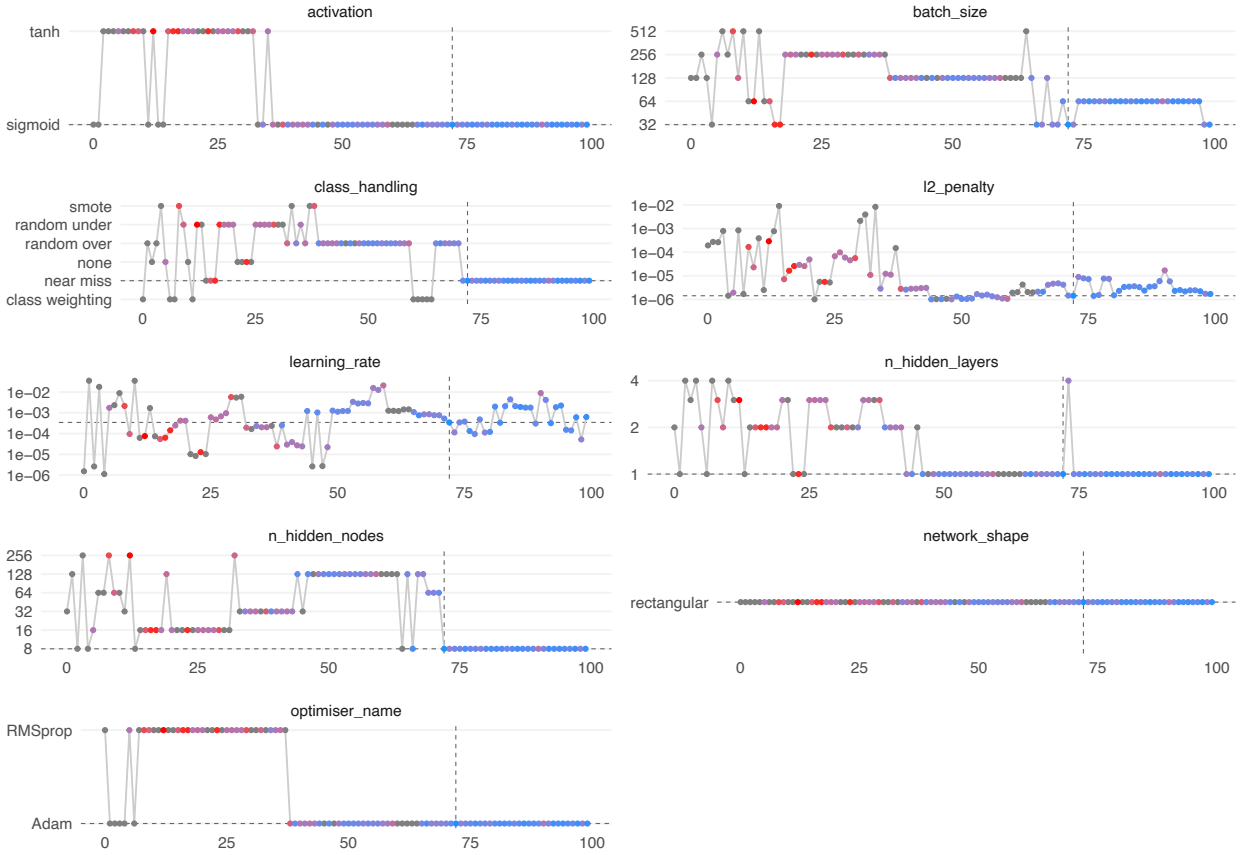


Figure S6: Linear, daily rate >= 1.

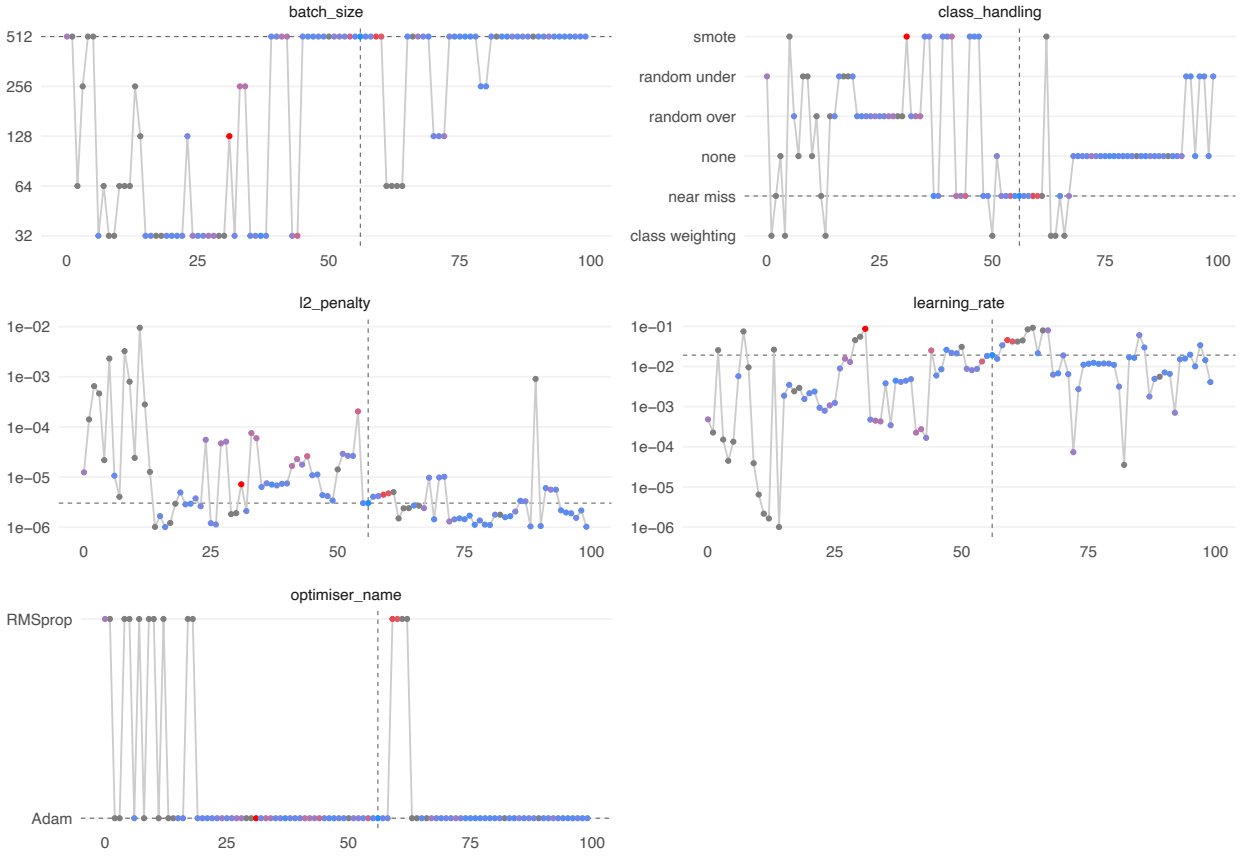


Figure S7: MLP, daily rate >= 2.

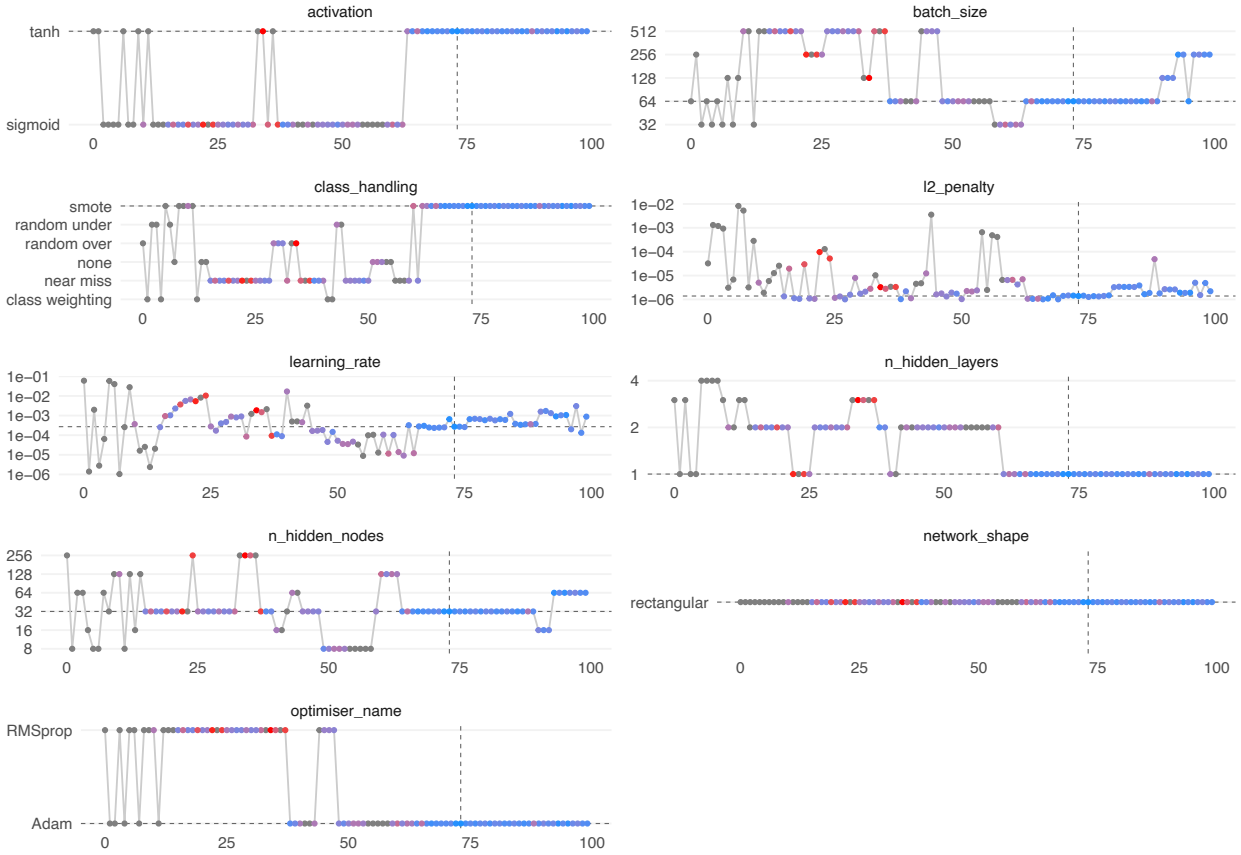


Figure S8: Linear, daily rate >= 2.

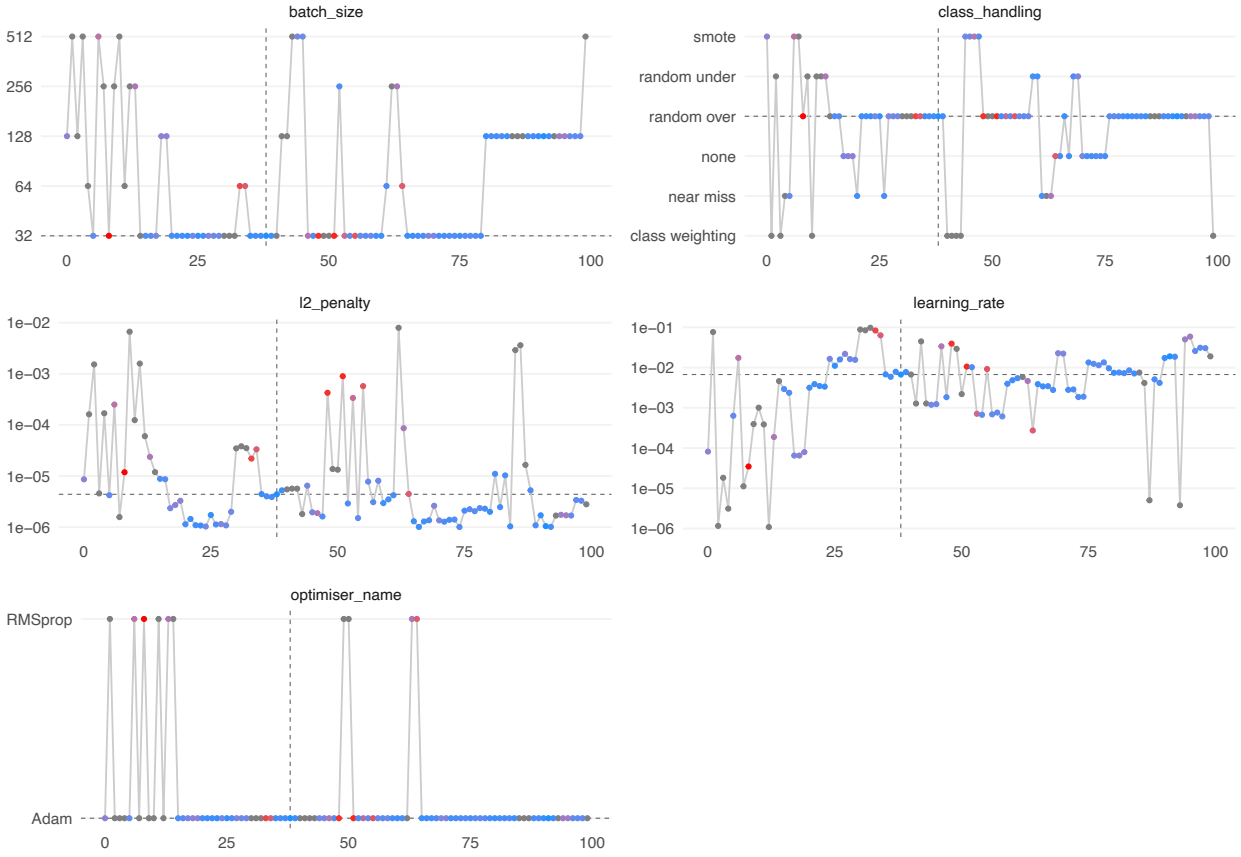


Figure S9: MLP, daily rate >= 3.

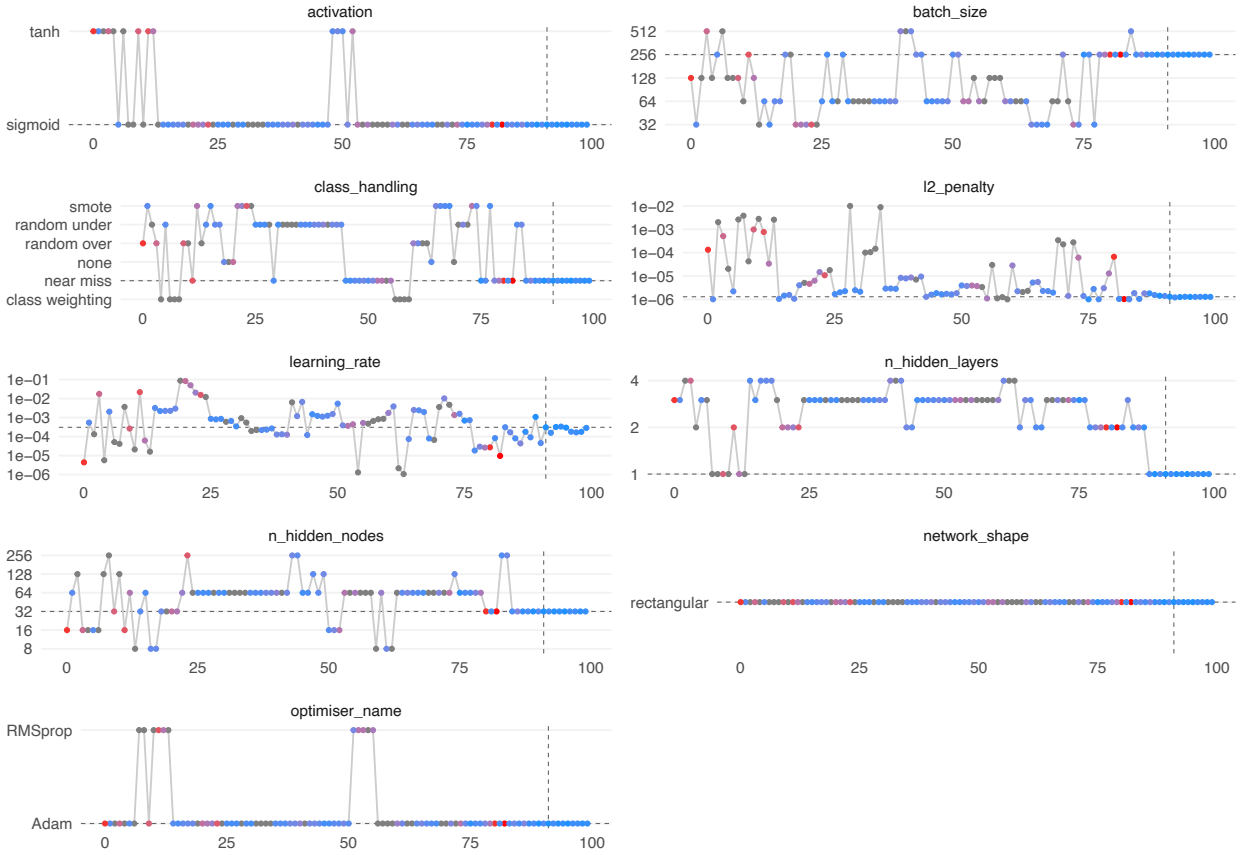


Figure S10: Linear, daily rate >= 3.

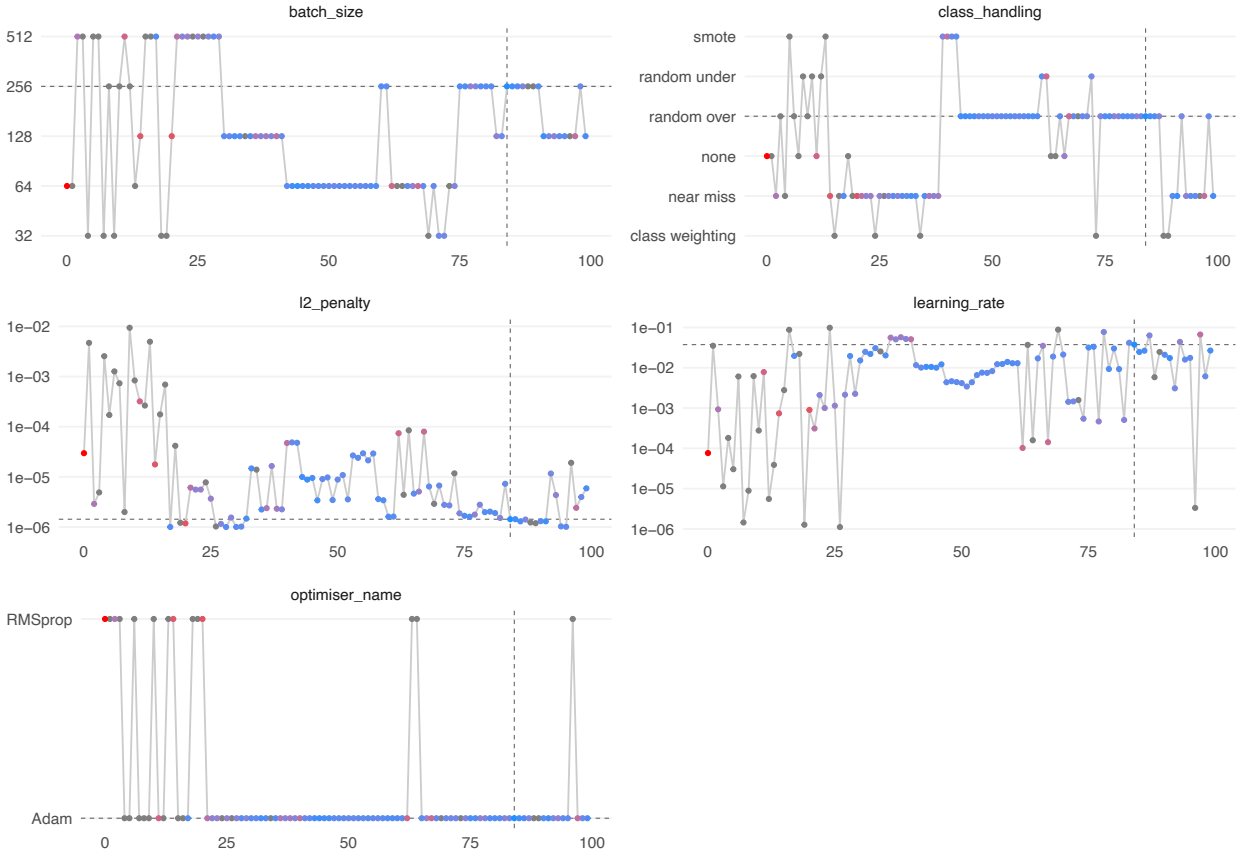


Figure S11: MLP, daily rate >= 5.

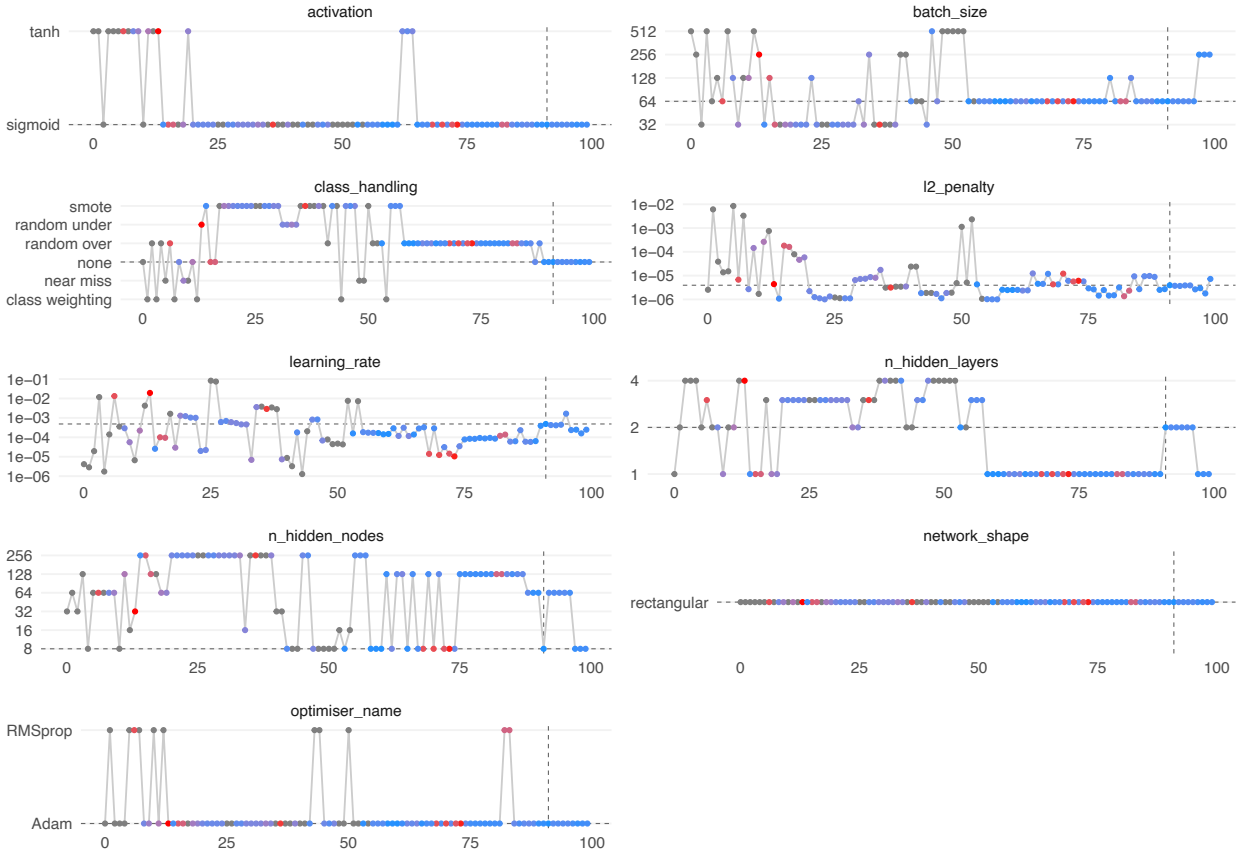


Figure S12: Linear, daily rate >= 5.

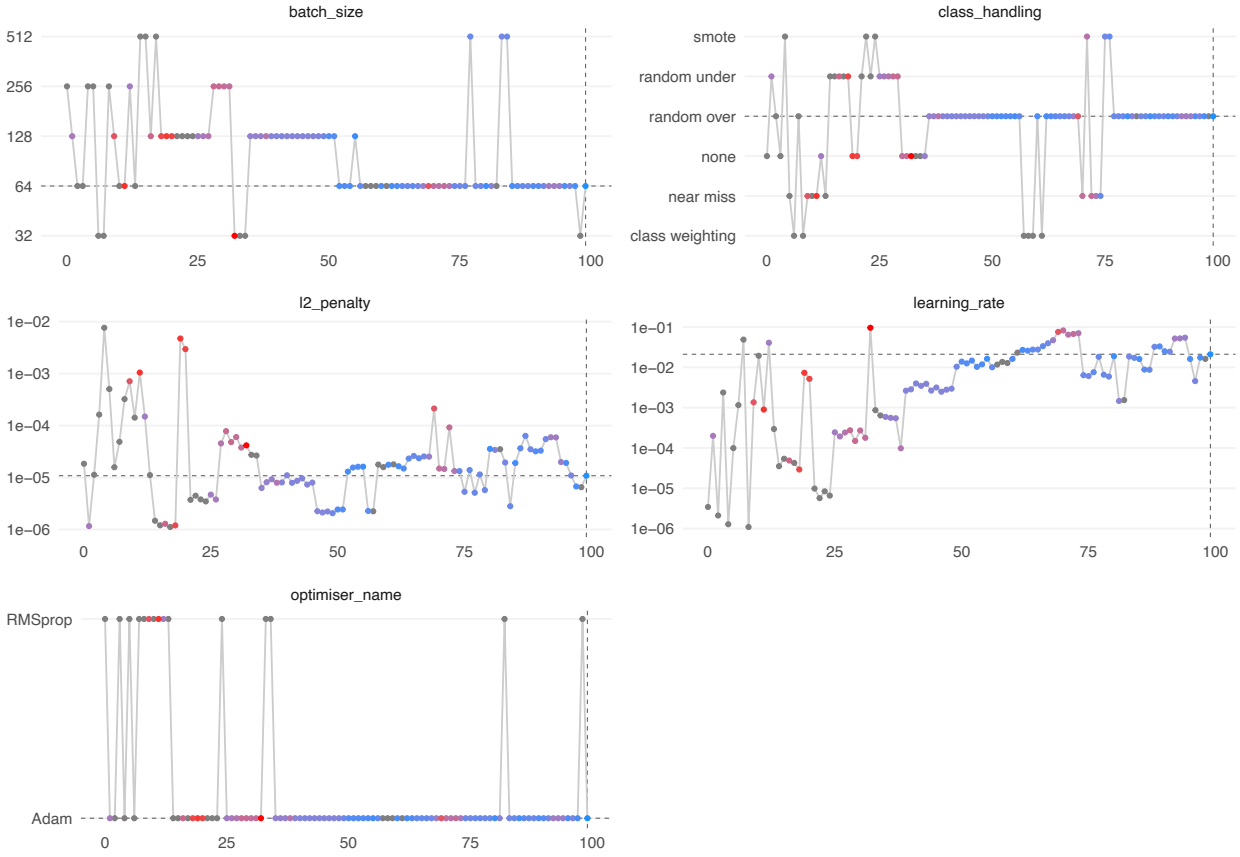


Figure S13: Ref. model, MLP, daily rate > 0.

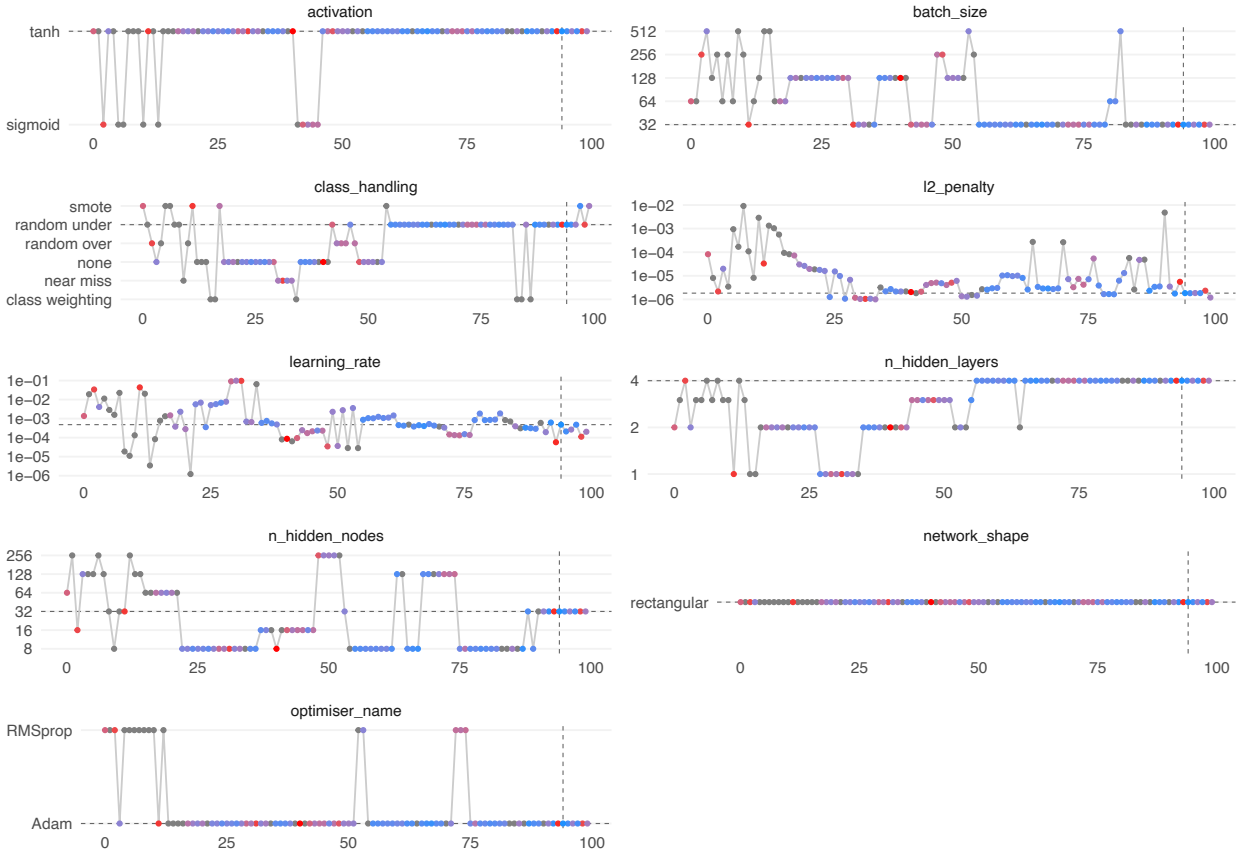


Figure S14: Ref. model, Linear, daily rate > 0.

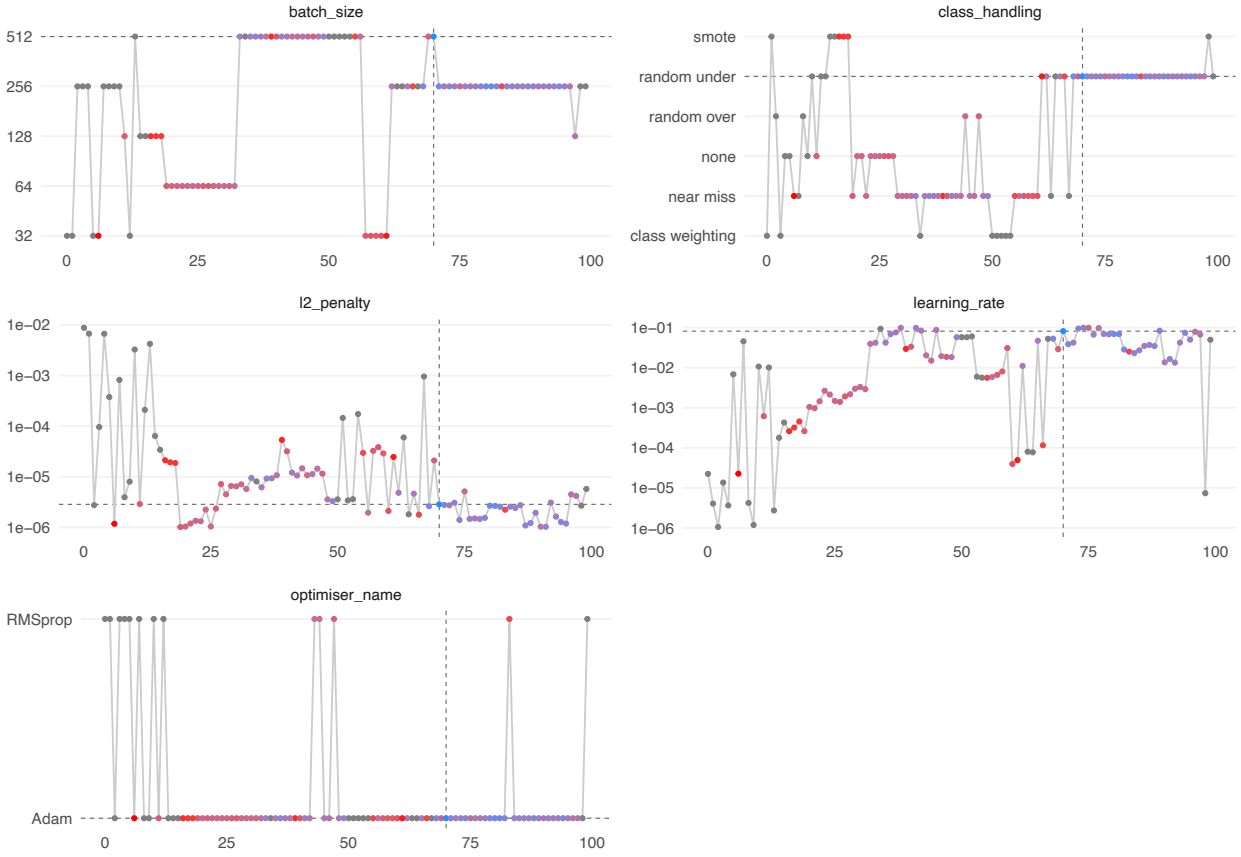


Figure S15: Ref. model, MLP, daily rate >= 1.

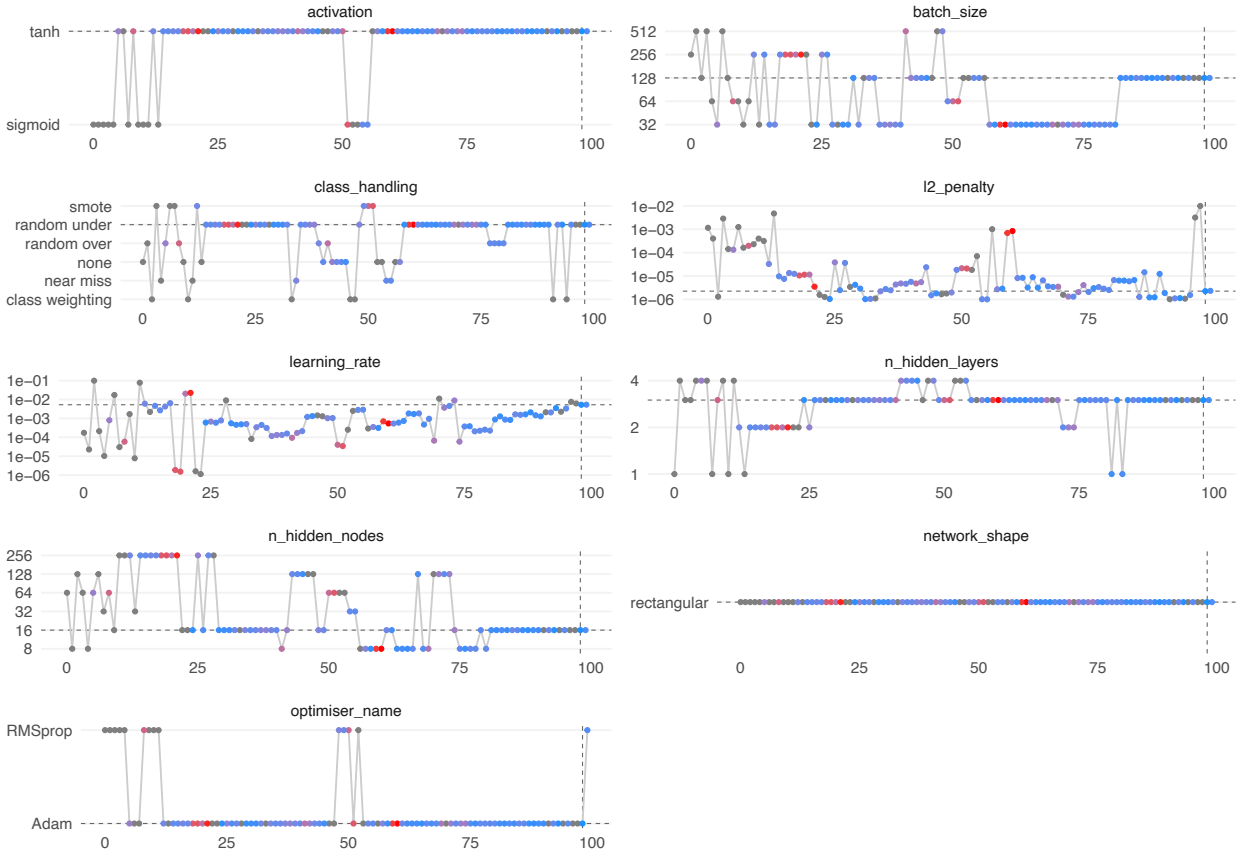
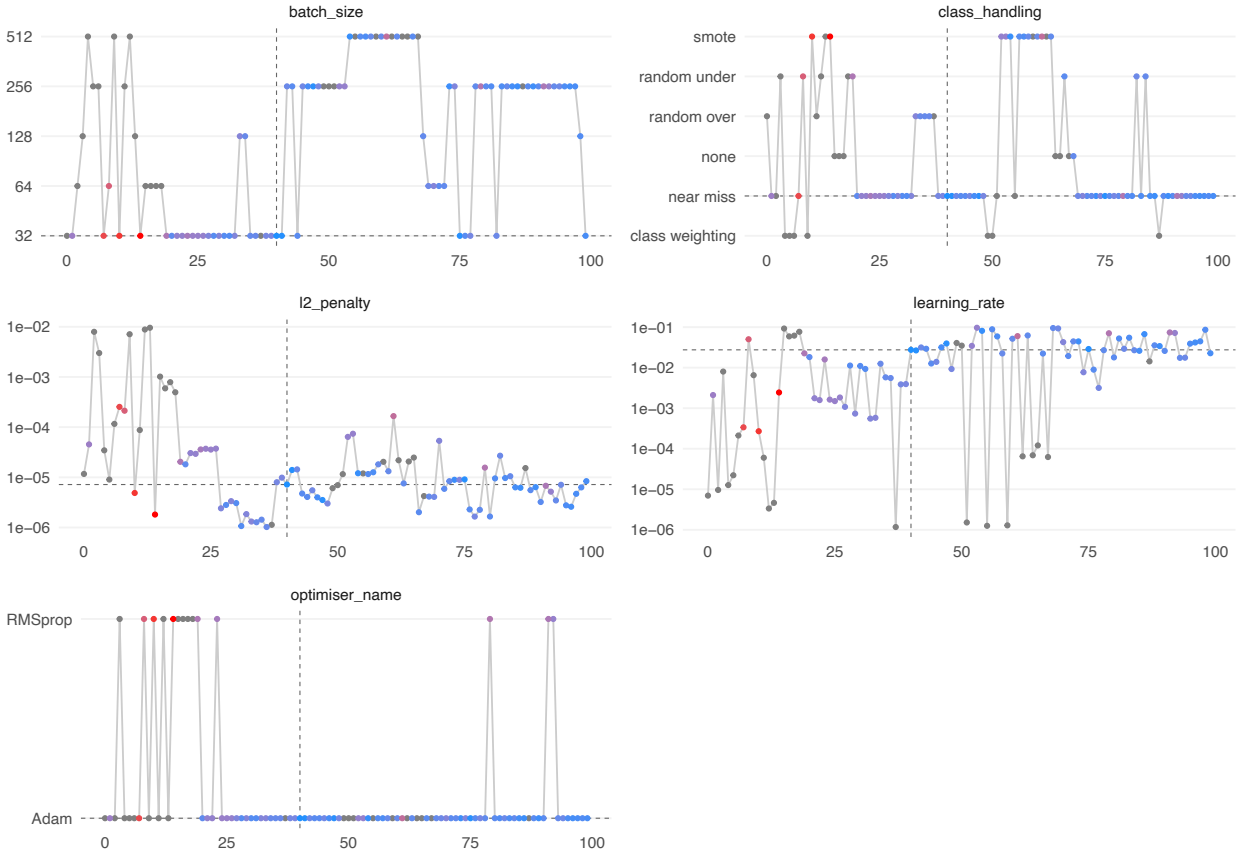


Figure S16: Ref. model, Linear, daily rate >= 1.



ROC curves

Receiver-operating characteristic (ROC) curves in the development and test sets.

Figure S17: MLP, daily rate > 0.

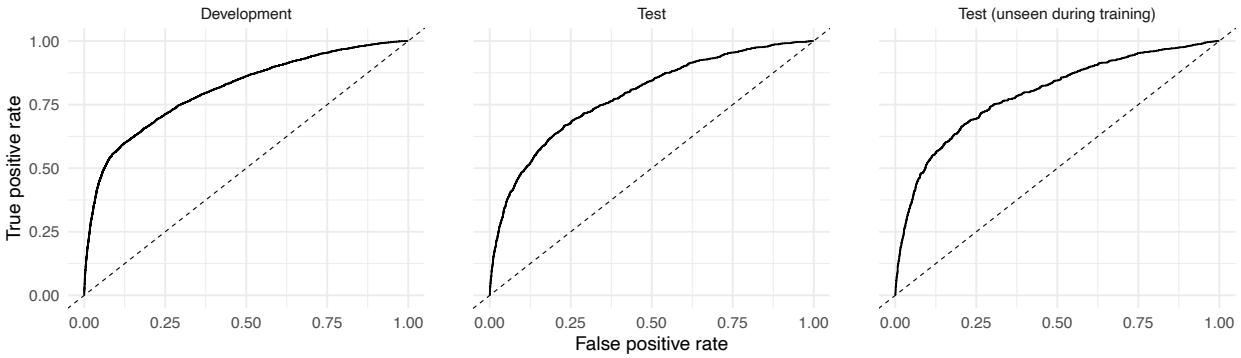


Figure S18: Linear, daily rate > 0.

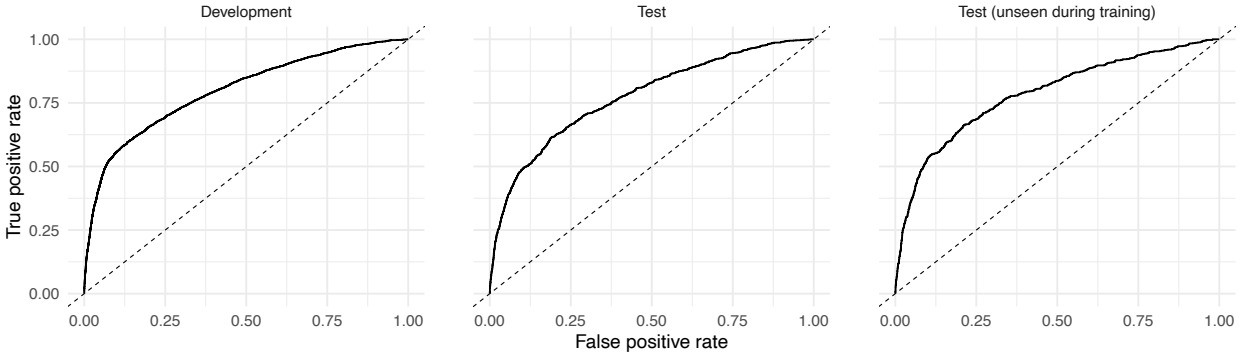


Figure S19: MLP, daily rate >= 1.

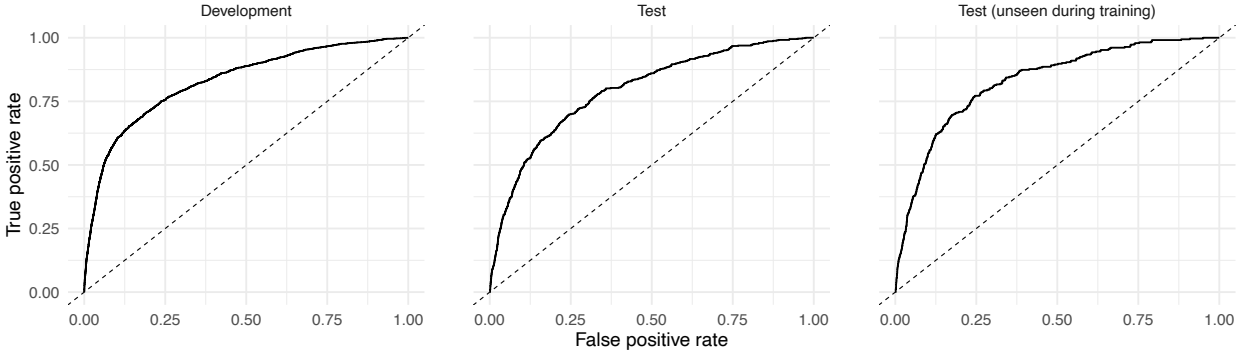


Figure S20: Linear, daily rate >= 1.

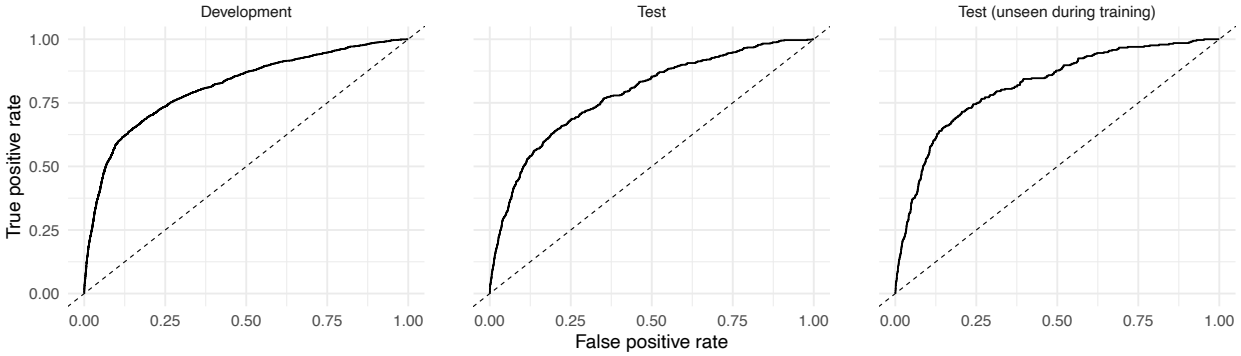


Figure S21: MLP, daily rate >= 2.

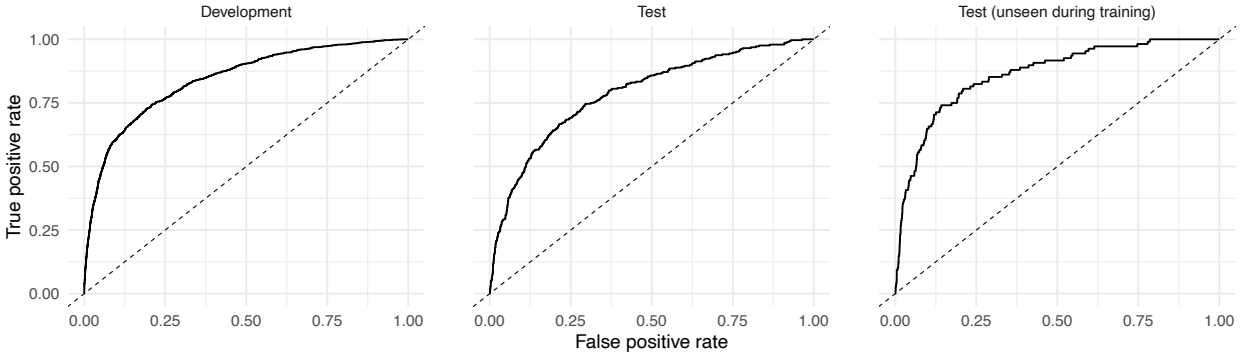


Figure S22: Linear, daily rate >= 2.

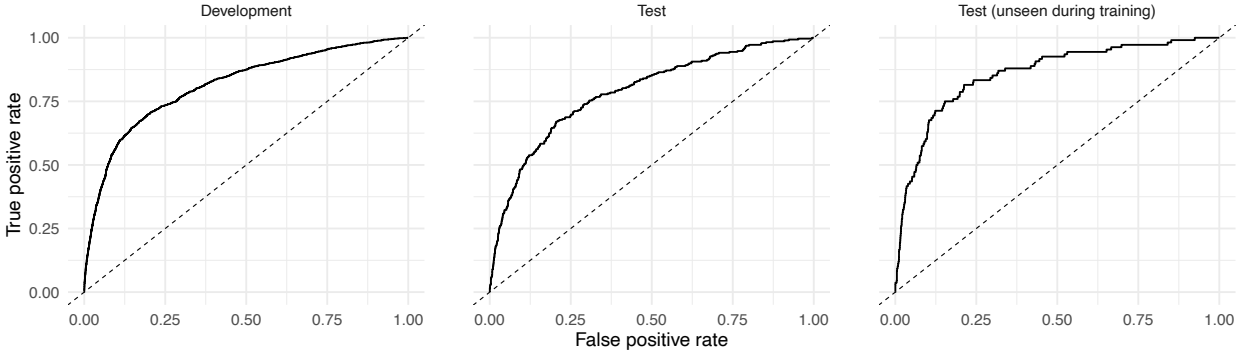


Figure S23: MLP, daily rate >= 3.

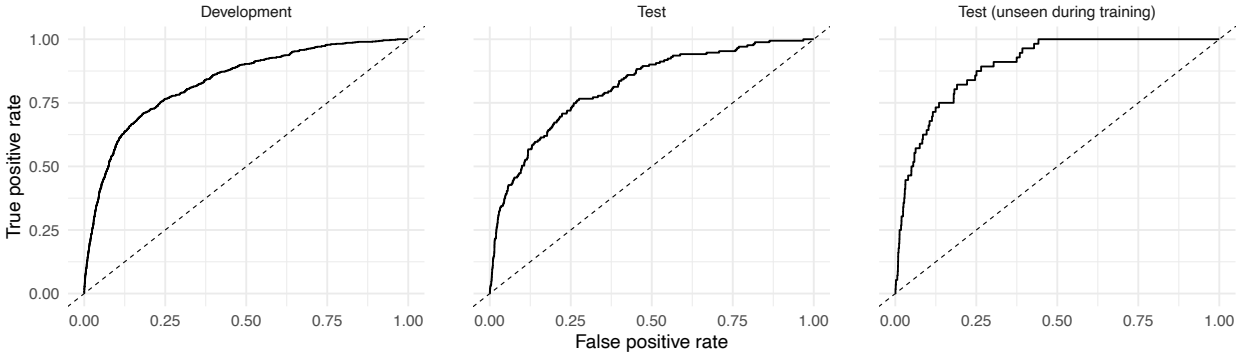


Figure S24: Linear, daily rate >= 3.

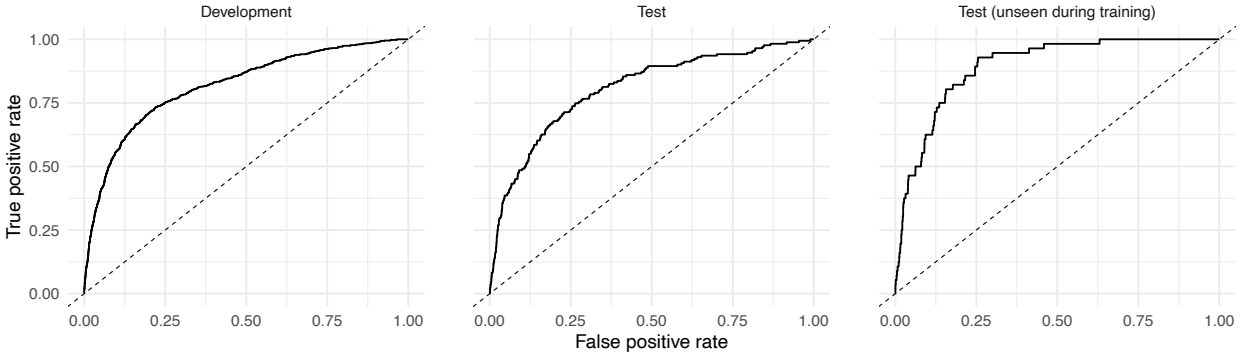


Figure S25: MLP, daily rate >= 5.

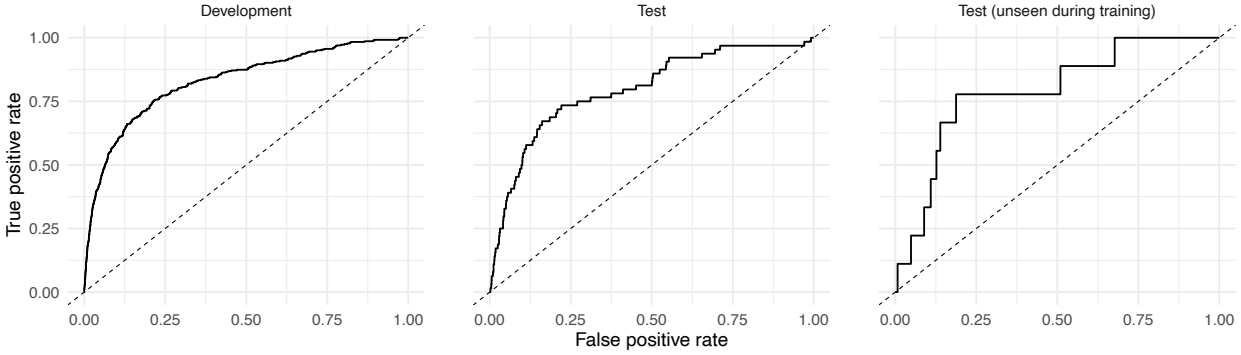


Figure S26: Linear, daily rate >= 5.

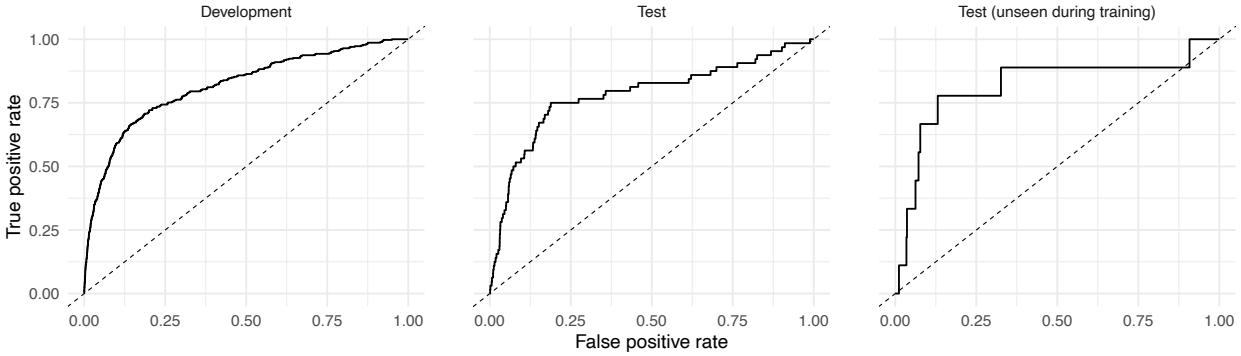


Figure S27: Ref. model, MLP, daily rate > 0.

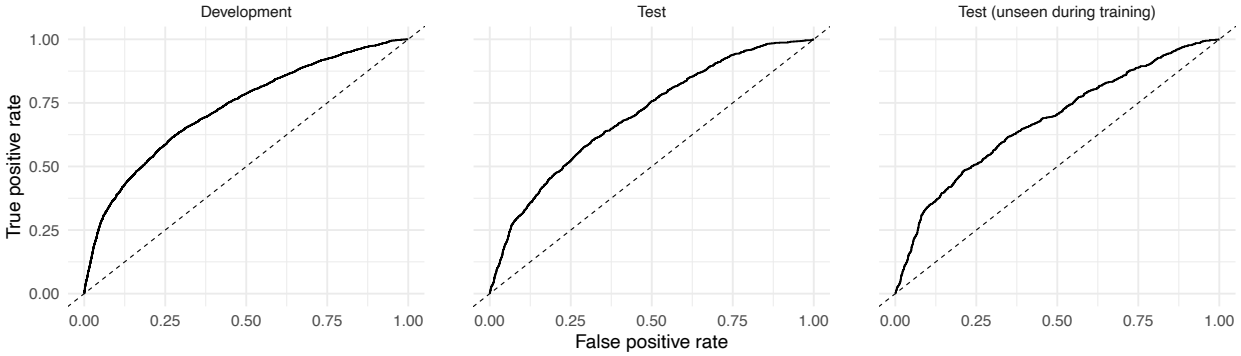


Figure S28: Ref. model, Linear, daily rate > 0.

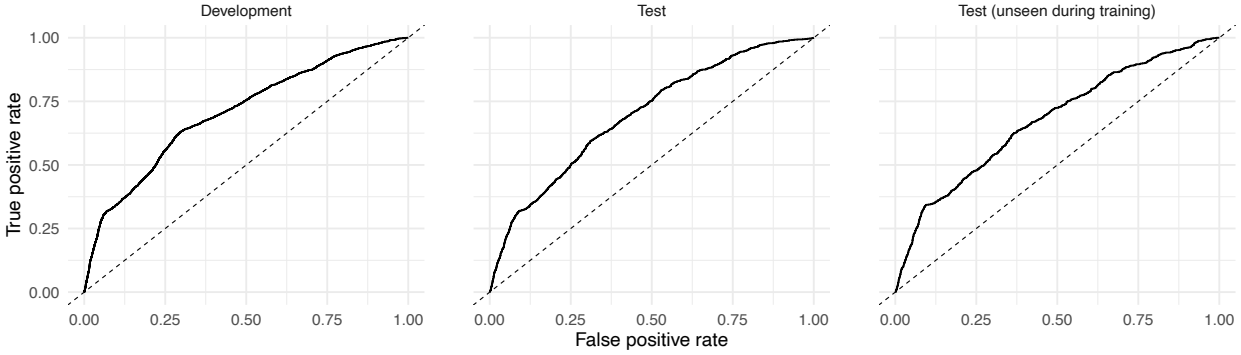


Figure S29: Ref. model, MLP, daily rate >= 1.

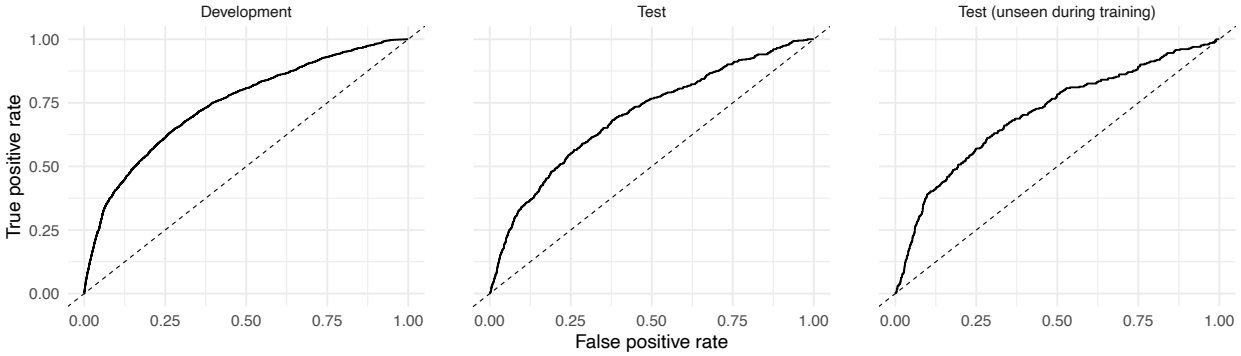
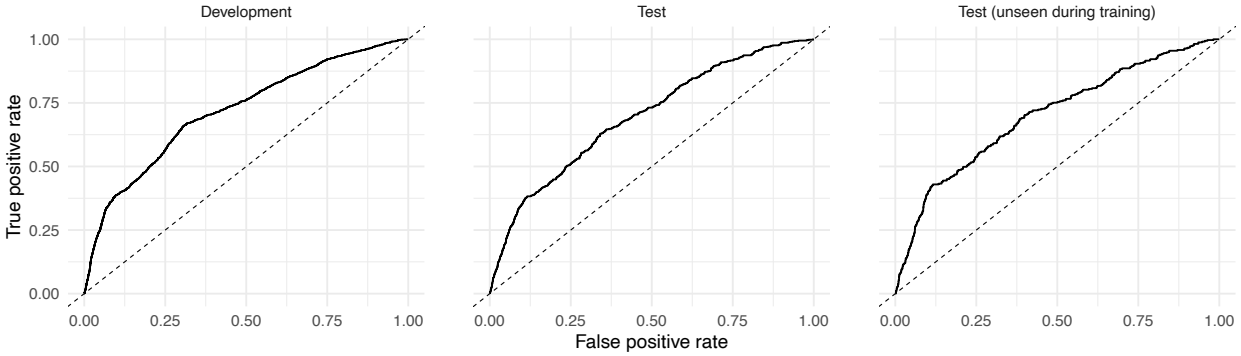


Figure S30: Ref. model, Linear, daily rate >= 1.



Calibration plots

Calibration curves in the development and test sets.

Figure S31: MLP, daily rate > 0.

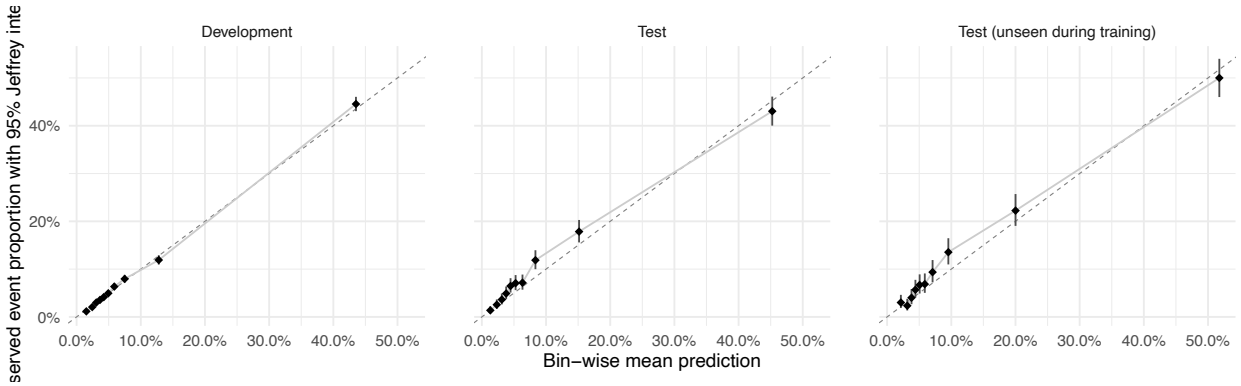


Figure S32: Linear, daily rate > 0.

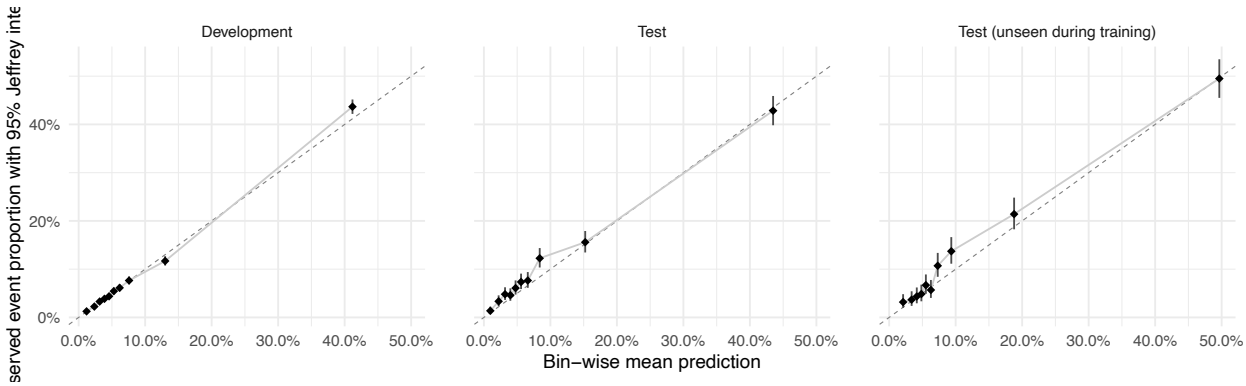


Figure S33: MLP, daily rate >= 1.

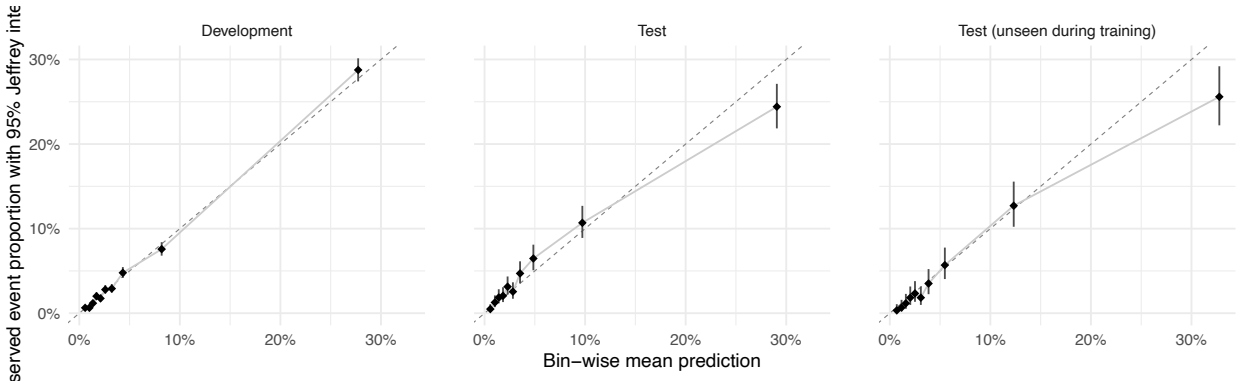


Figure S34: Linear, daily rate >= 1.

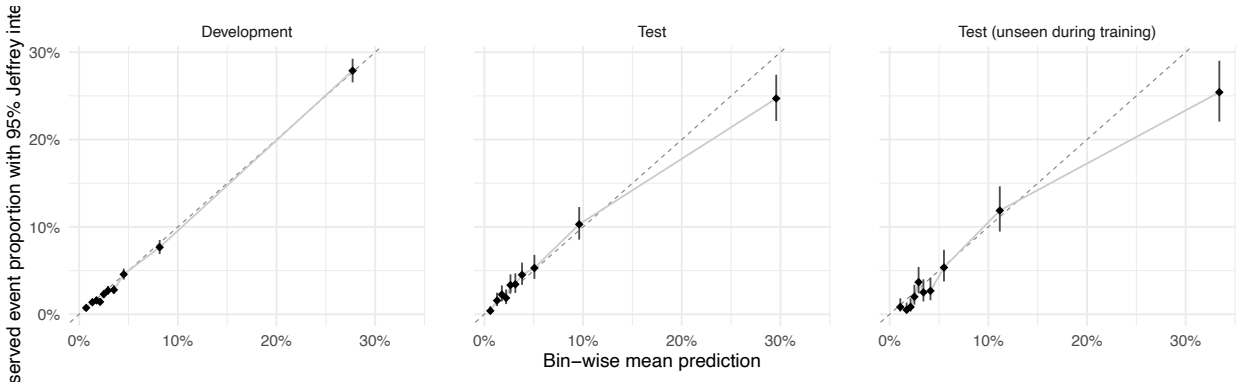


Figure S35: MLP, daily rate >= 2.

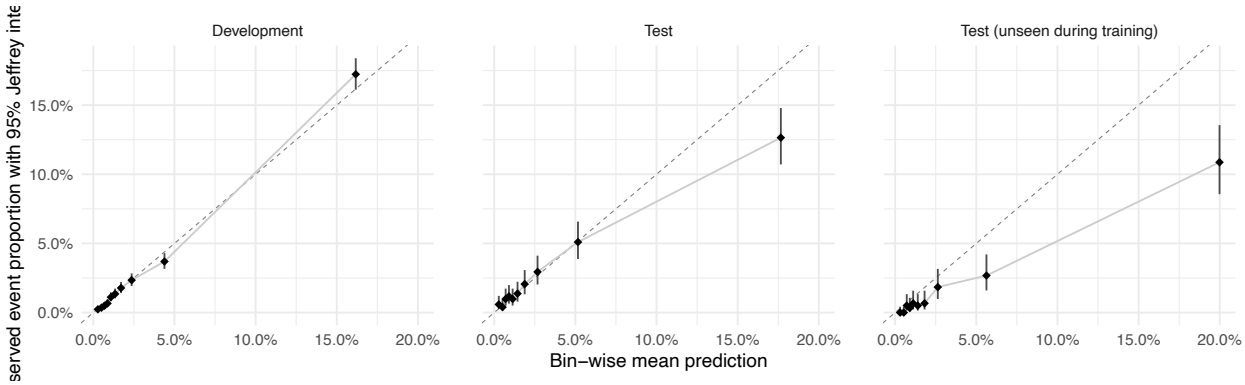


Figure S36: Linear, daily rate >= 2.

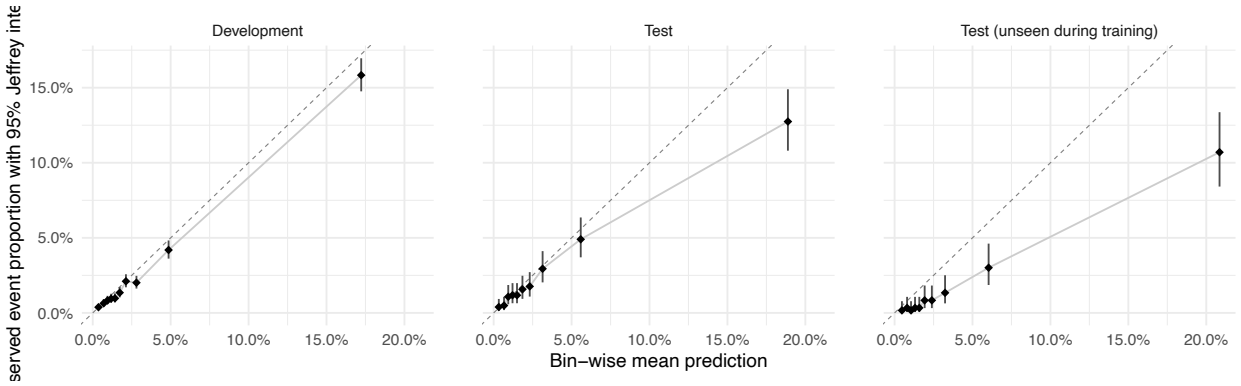


Figure S37: MLP, daily rate >= 3.

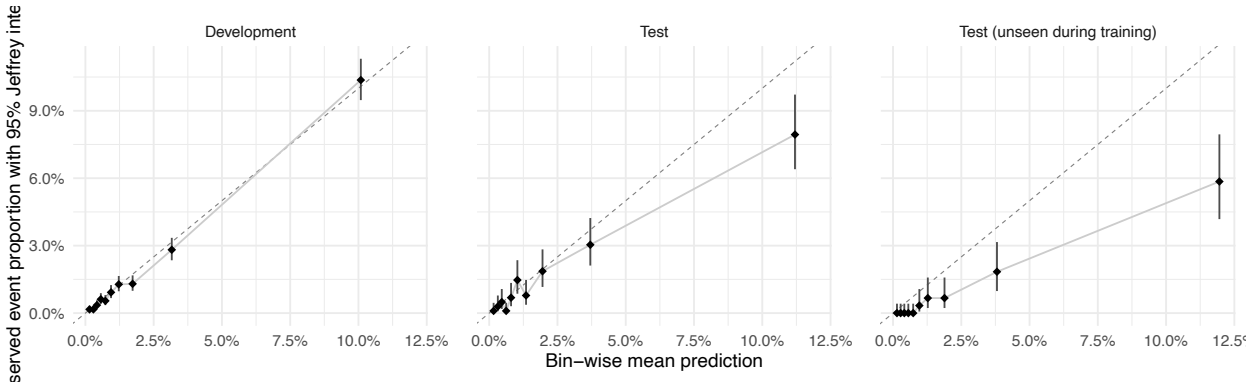


Figure S38: Linear, daily rate >= 3.

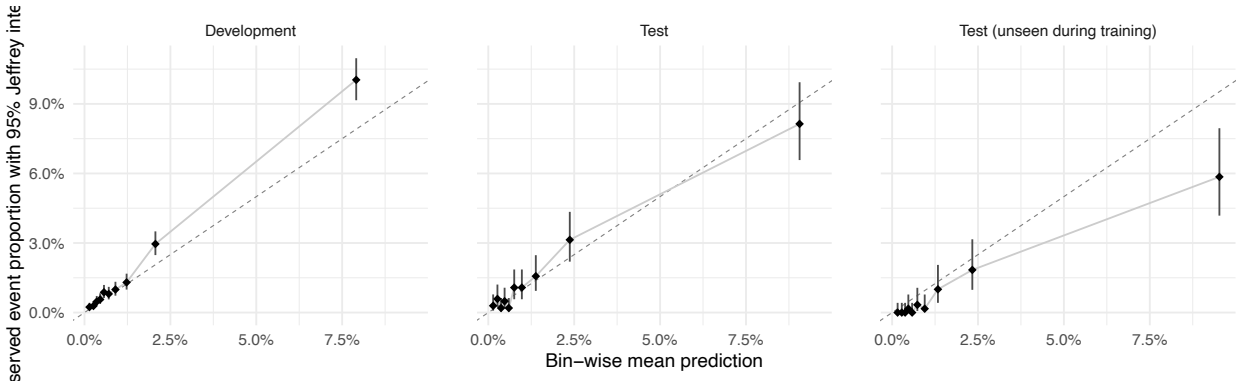


Figure S39: MLP, daily rate >= 5.

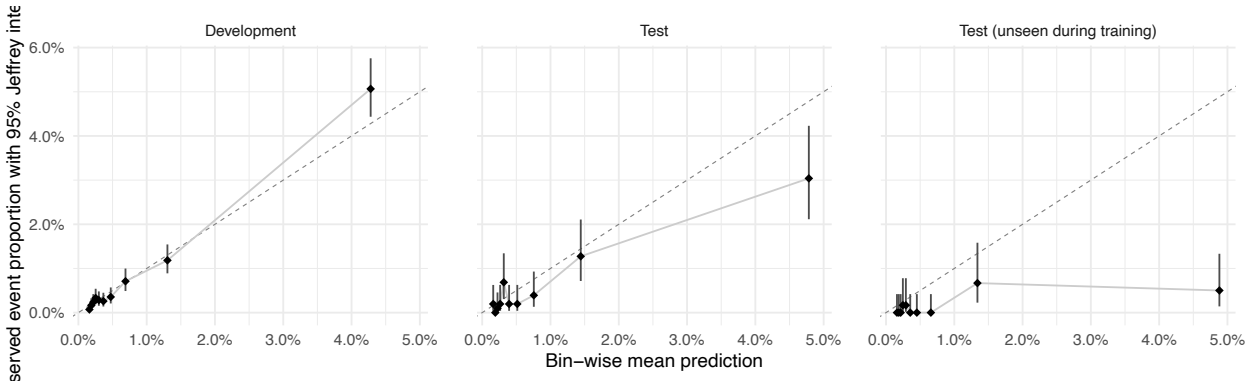


Figure S40: Linear, daily rate >= 5.

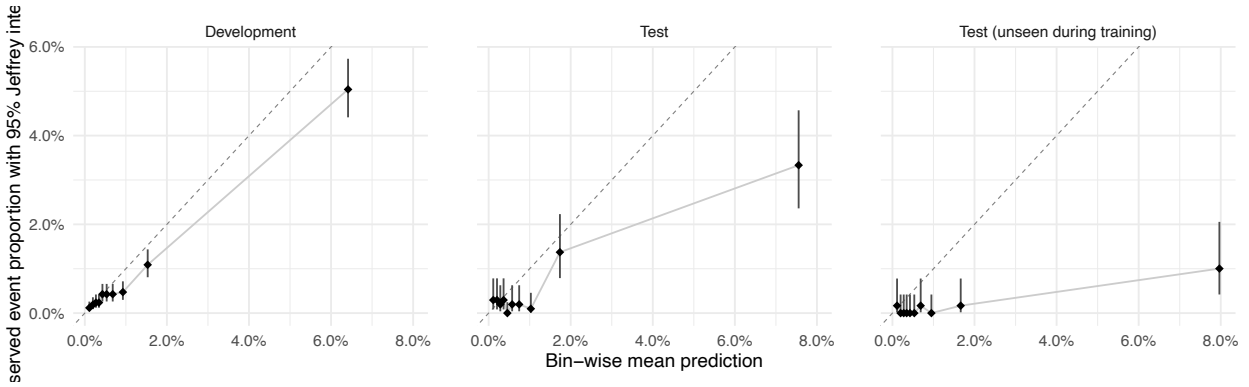


Figure S41: Ref. model, MLP, daily rate > 0.

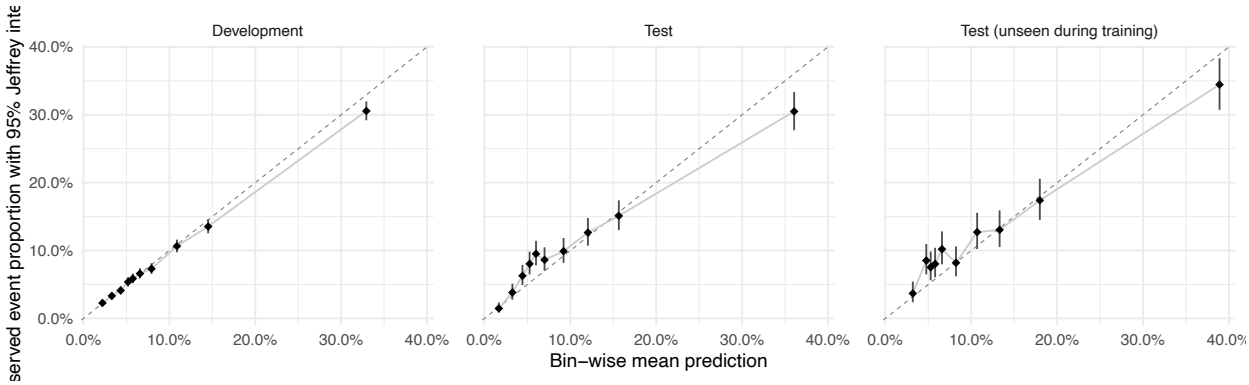


Figure S42: Ref. model, Linear, daily rate > 0.

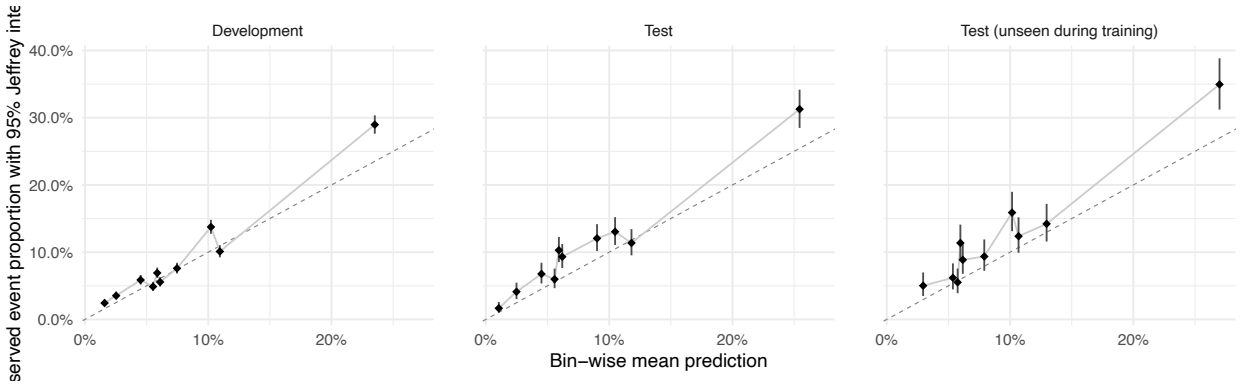


Figure S43: Ref. model, MLP, daily rate >= 1.

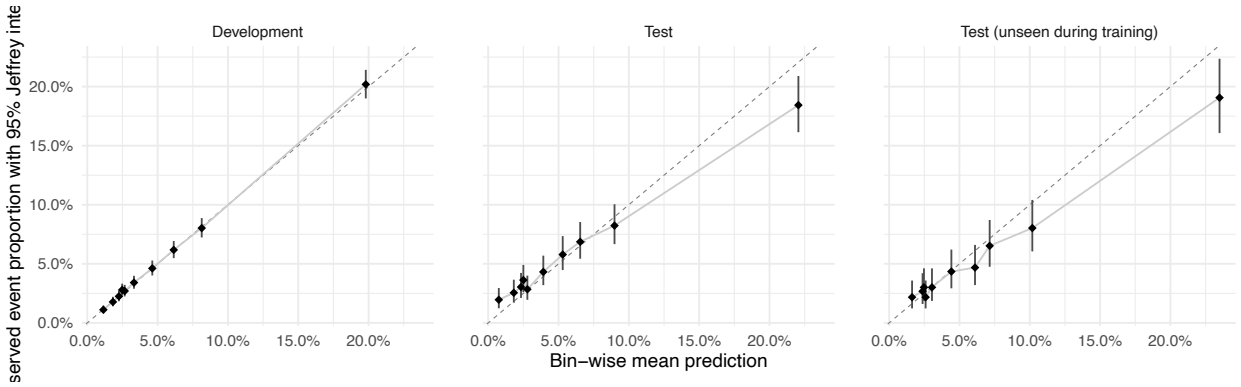
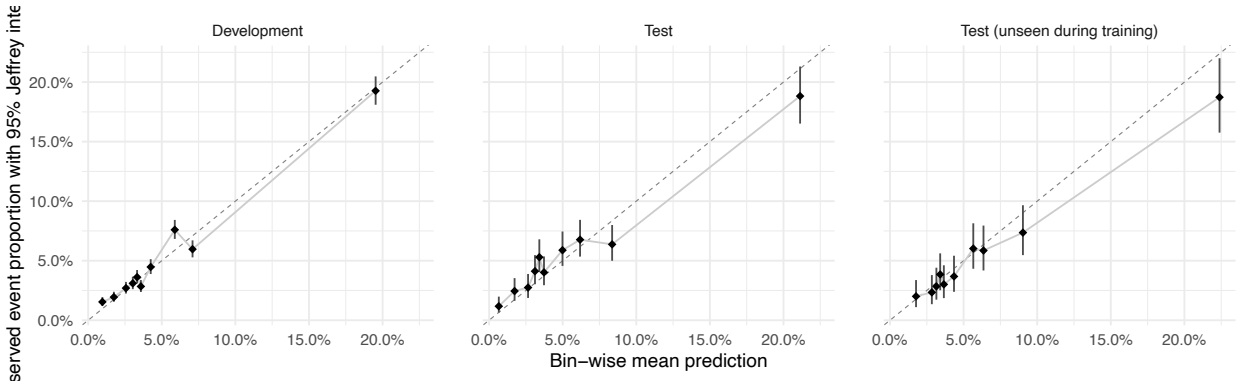


Figure S44: Ref. model, Linear, daily rate >= 1.



Decision curves

Decision curves based on the test set. The curves show the clinical utility (in the unit of standardised net benefit) of intervening in all patients (magenta), no patients (dark grey, follows the x axis) and patients identified by the model (blue).

Figure S45: MLP, daily rate > 0.

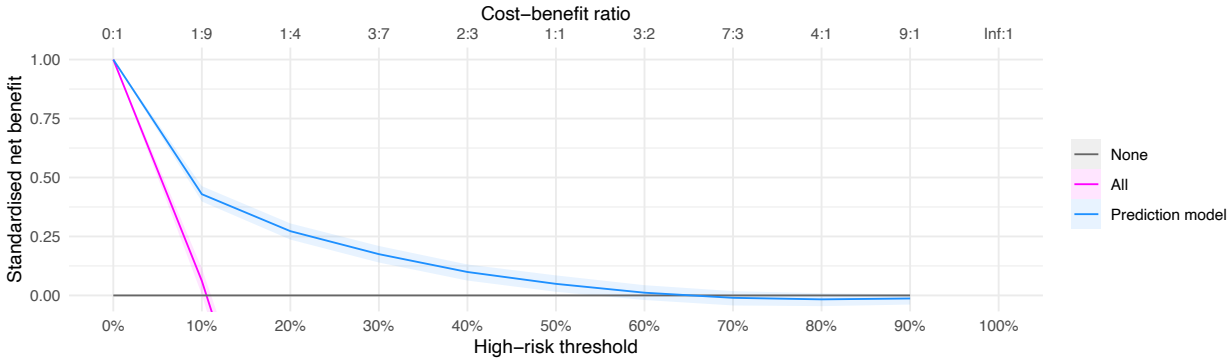


Figure S46: Linear, daily rate > 0.

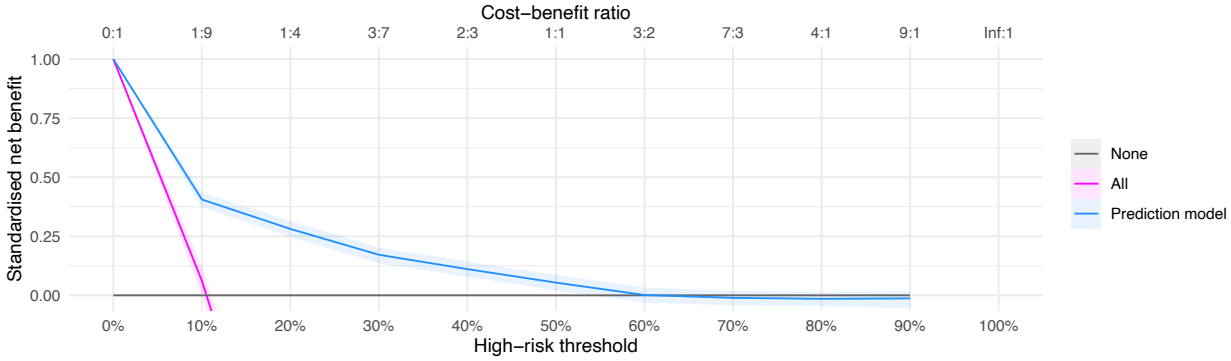


Figure S47: MLP, daily rate >= 1.

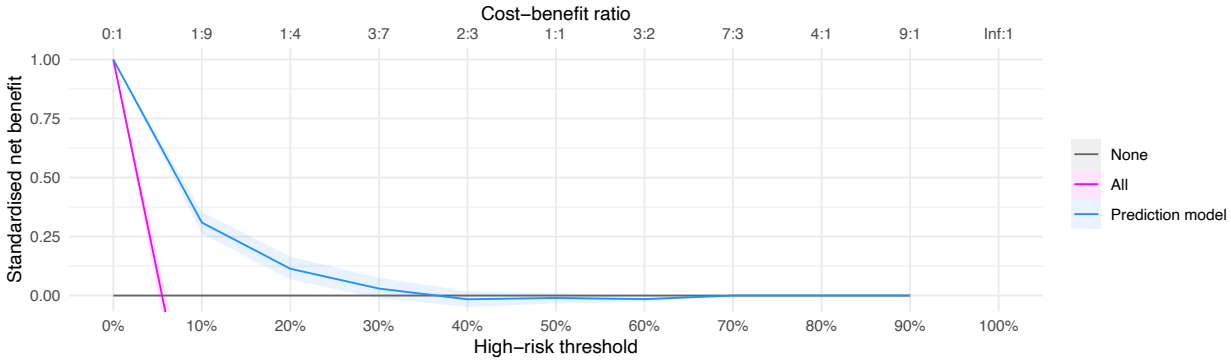


Figure S48: Linear, daily rate >= 1.

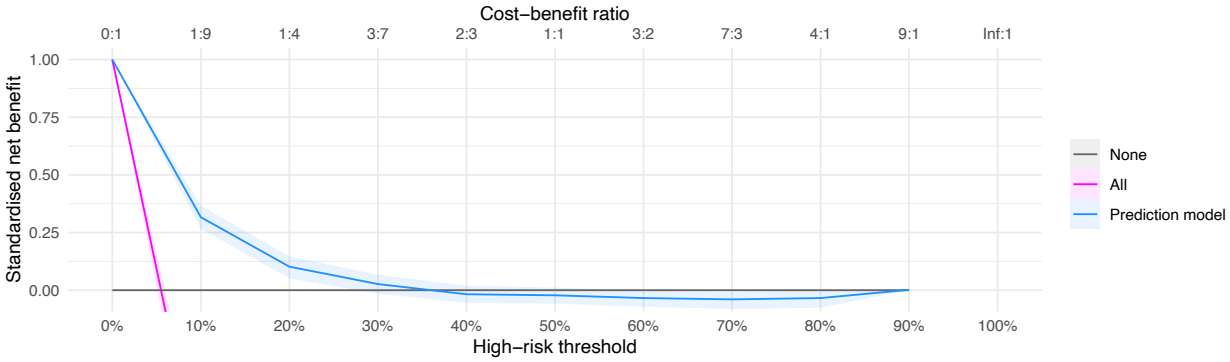


Figure S49: MLP, daily rate >= 2.

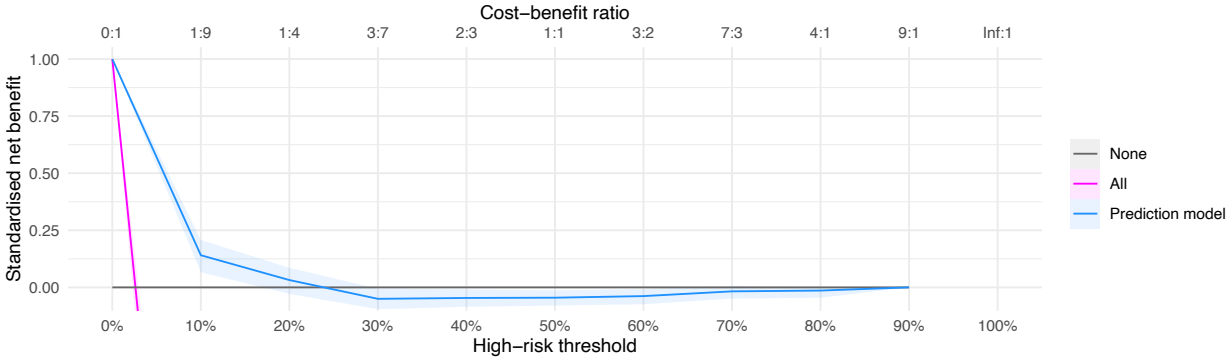


Figure S50: Linear, daily rate >= 2.

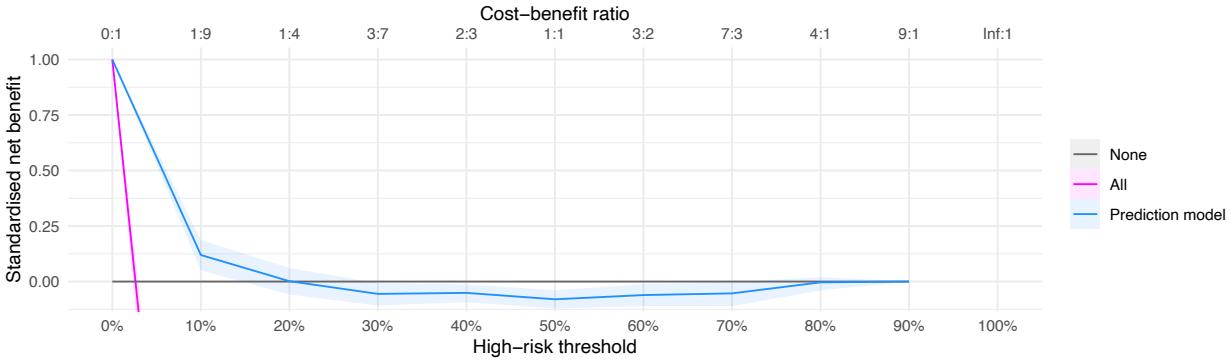


Figure S51: MLP, daily rate >= 3.

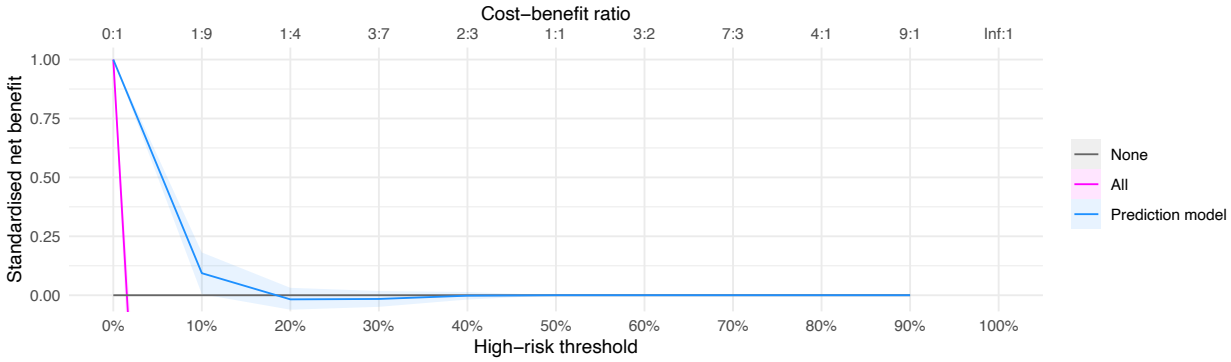


Figure S52: Linear, daily rate >= 3.

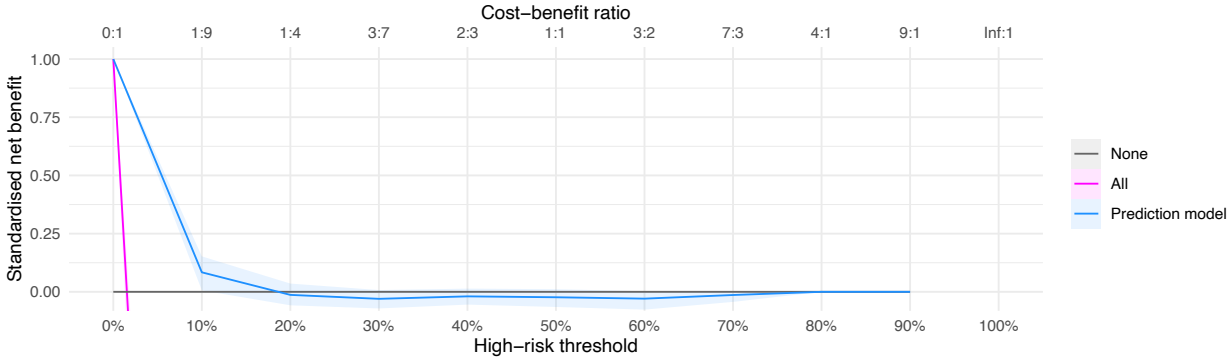


Figure S53: MLP, daily rate >= 5.

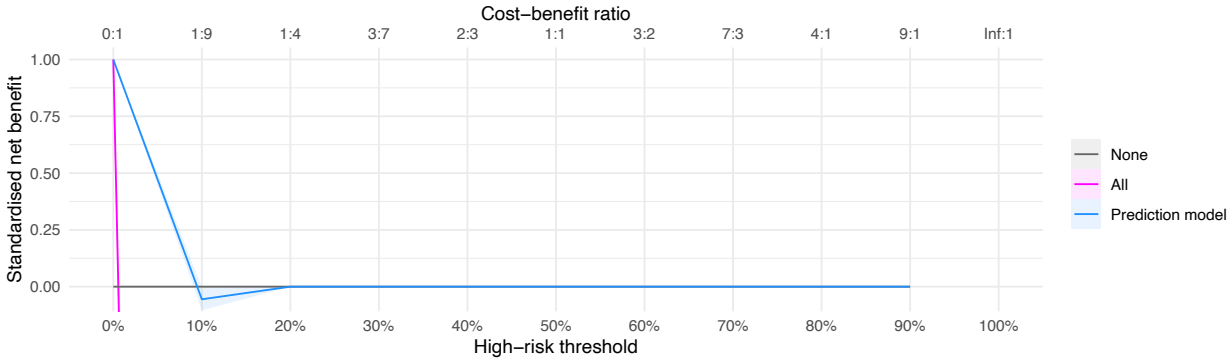


Figure S54: Linear, daily rate >= 5.

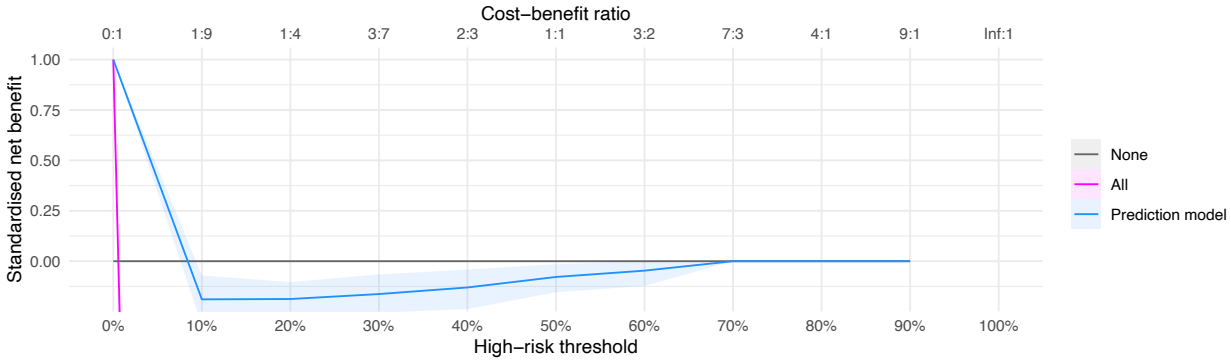


Figure S55: Ref. model, MLP, daily rate > 0.

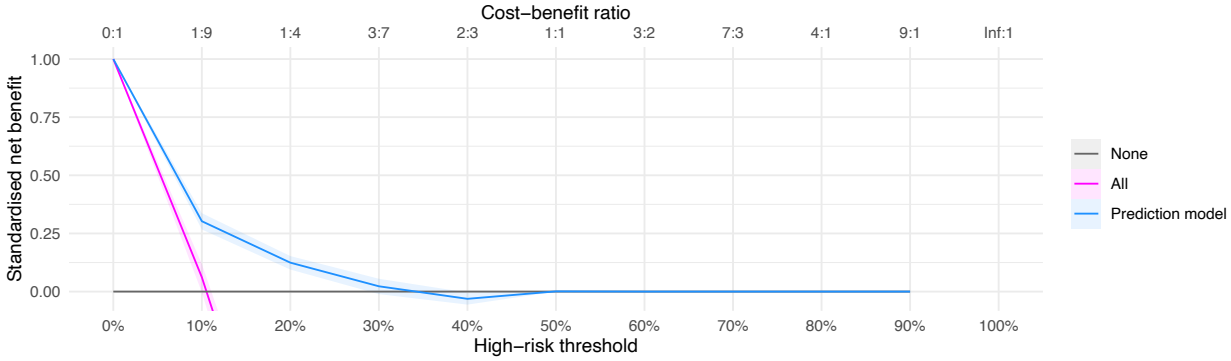


Figure S56: Ref. model, Linear, daily rate > 0.

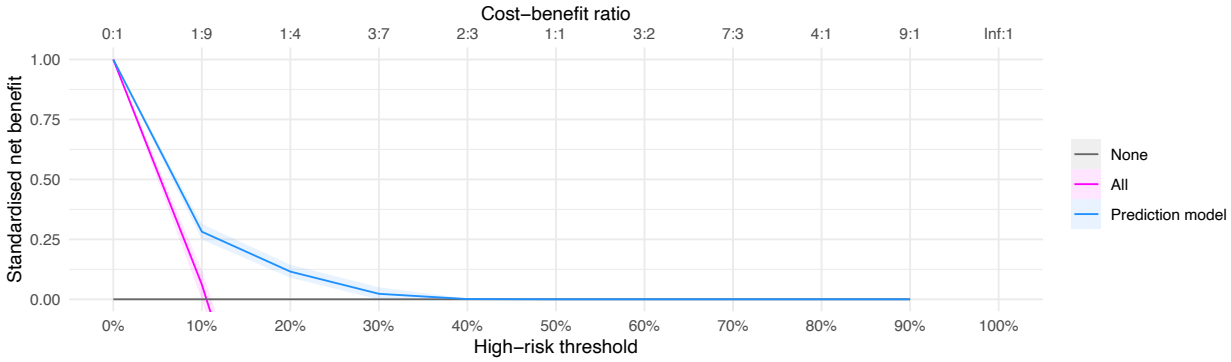


Figure S57: Ref. model, MLP, daily rate >= 1.

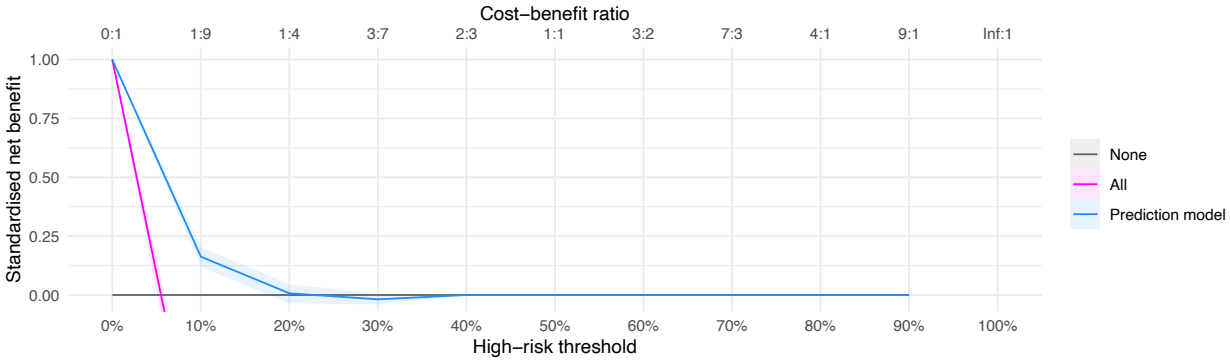
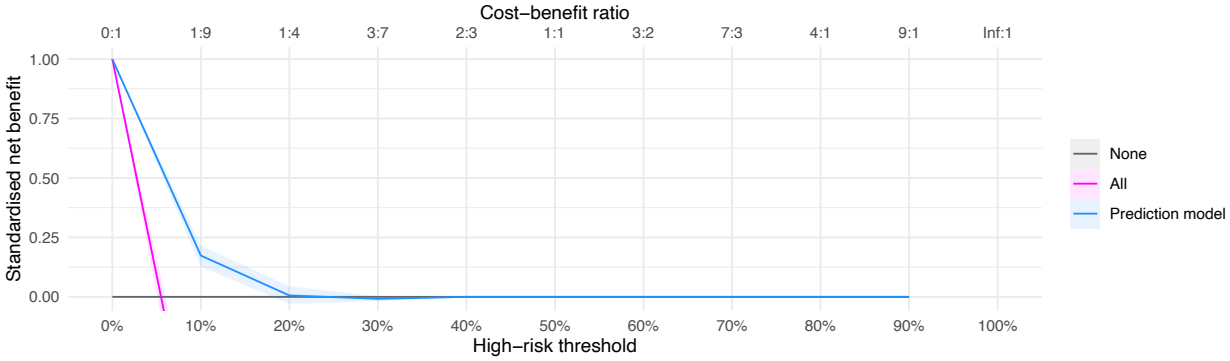


Figure S58: Ref. model, Linear, daily rate >= 1.



SHAP plots

The individual shap plots below all visualise the the distributions as density bands of shap values by feature. Blue represents low feature values (0 for binary features) and red high feature values (1 for binary features). Continuous features were binned into deciles.

Figure S59: MLP, daily rate > 0

Figure S60: Linear, daily rate > 0

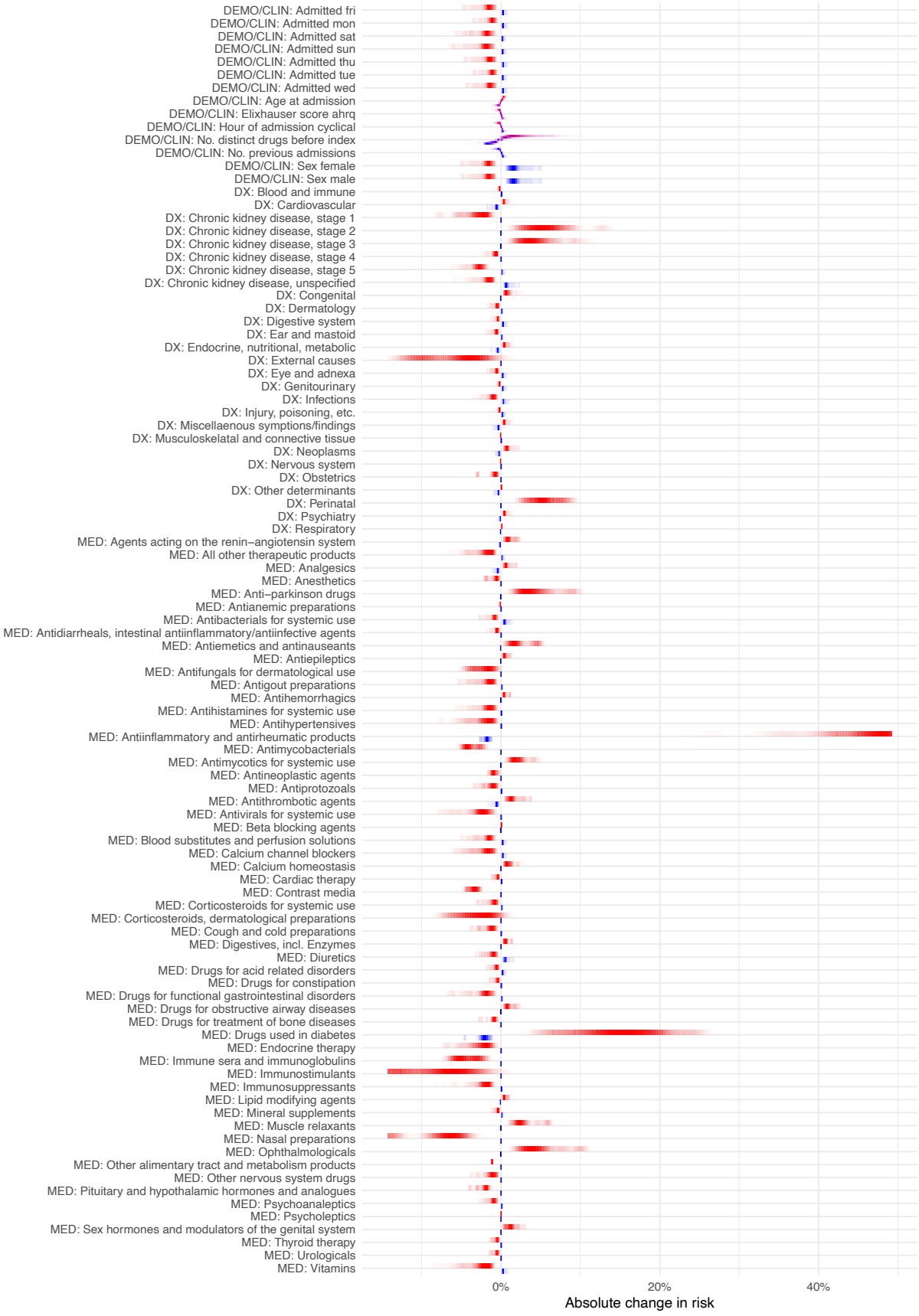


Figure S61: MLP, daily rate >= 1

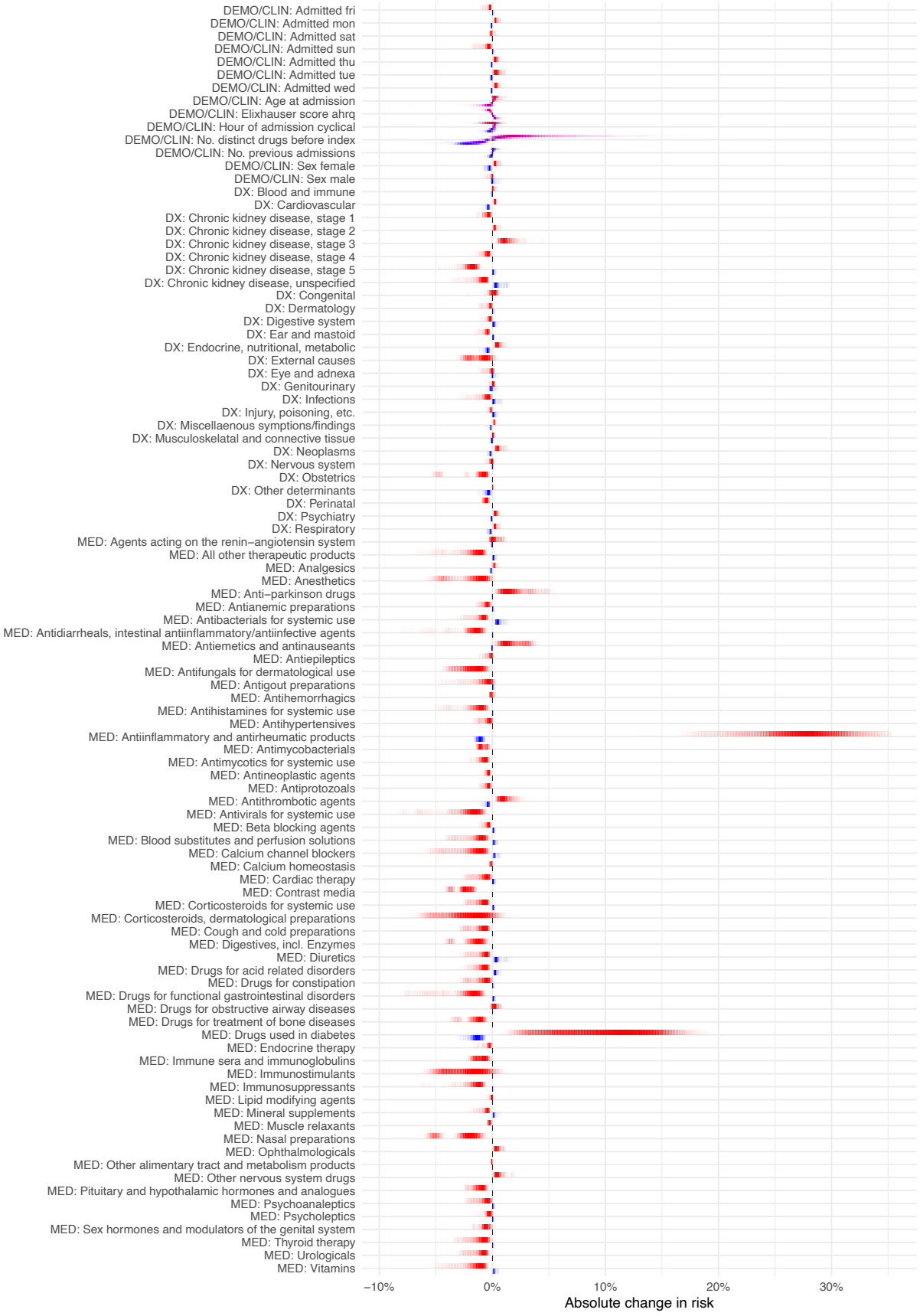


Figure S62: Linear, daily rate >= 1

Figure S63: MLP, daily rate >= 2

Figure S64: Linear, daily rate >= 2

Figure S65: MLP, daily rate >= 3

Figure S66: Linear, daily rate >= 3

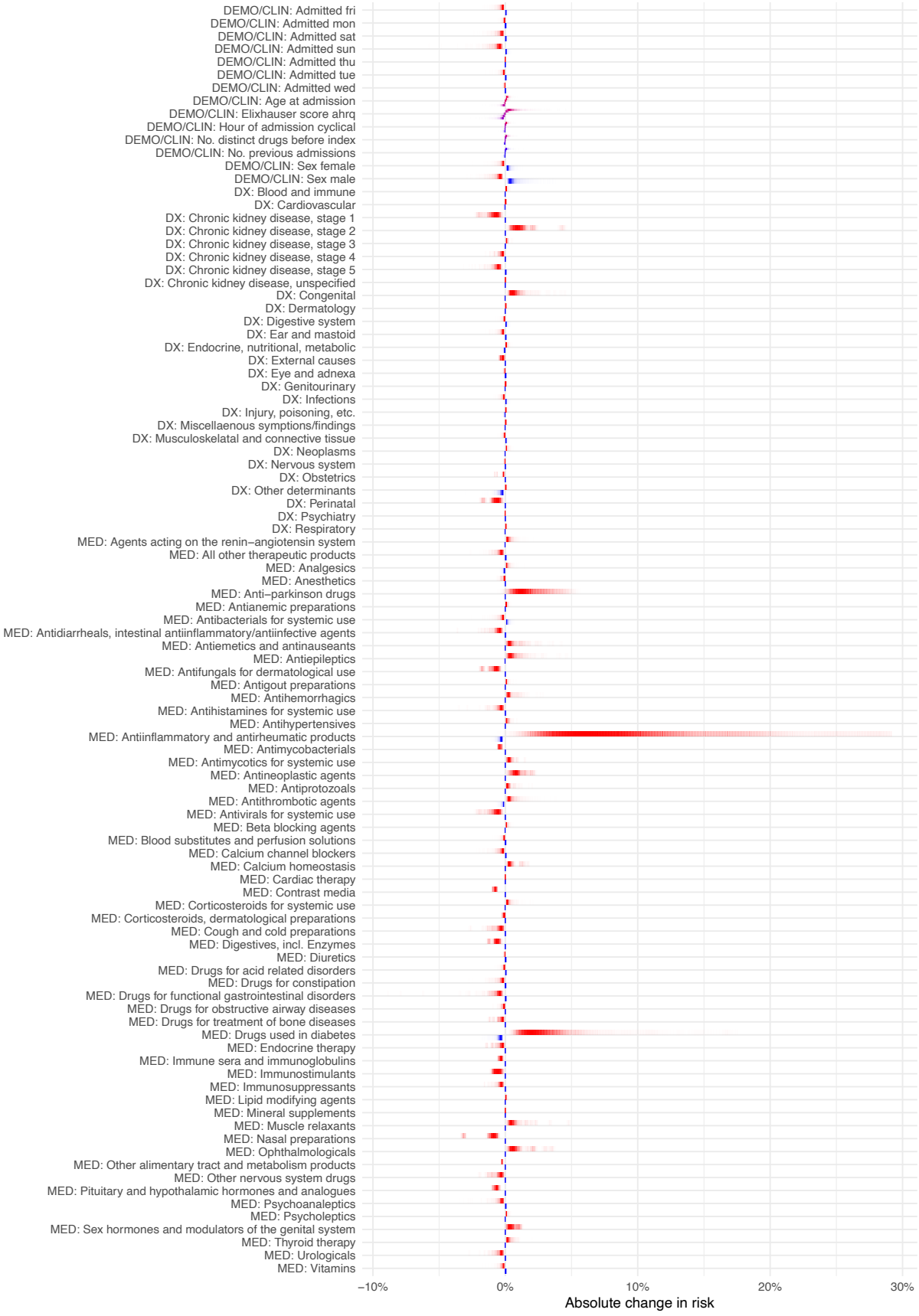


Figure S67: MLP, daily rate >= 5

Figure S68: Linear, daily rate >= 5

Figure S69: Ref. model, MLP, daily rate > 0

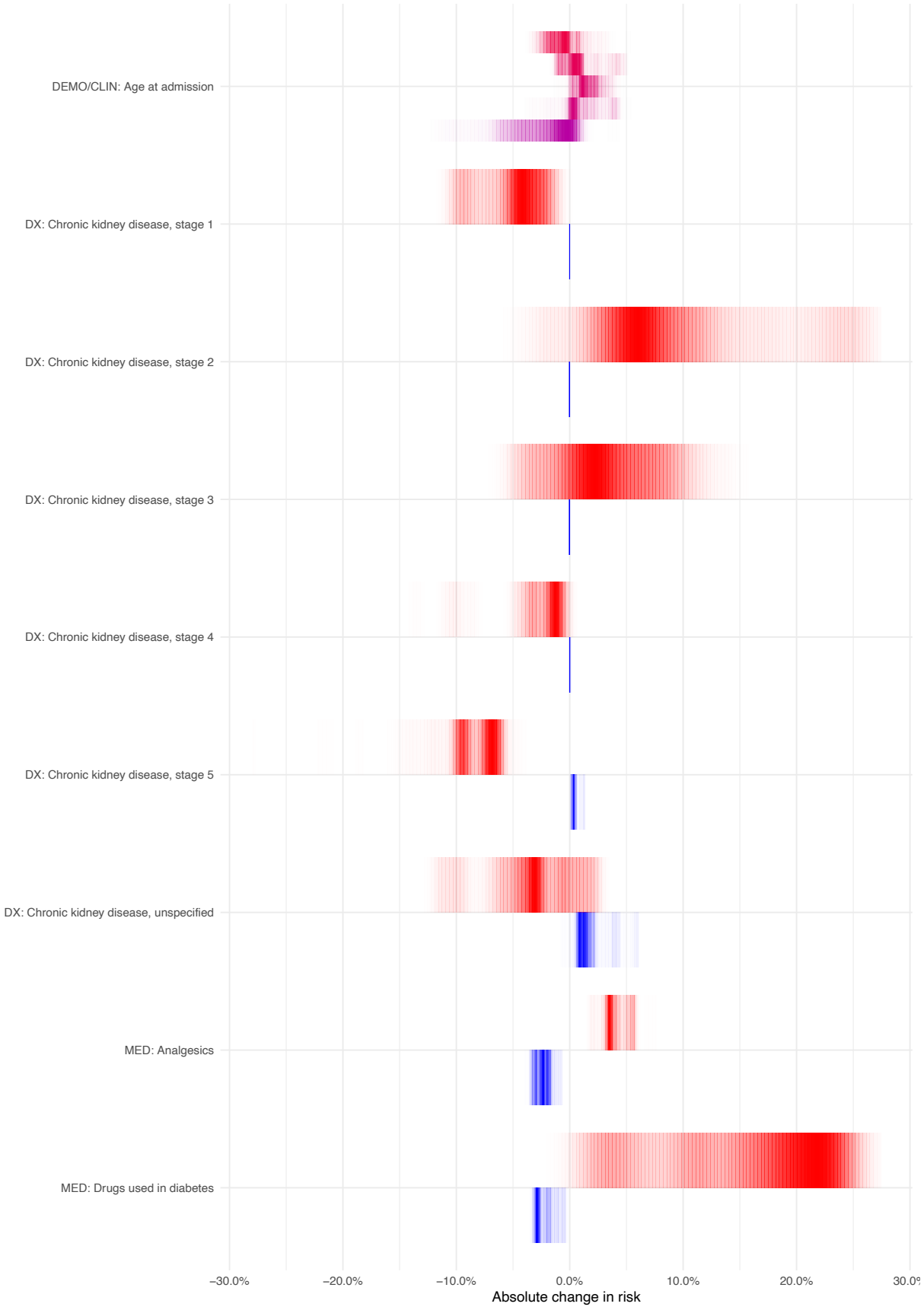


Figure S70: Ref. model, Linear, daily rate > 0

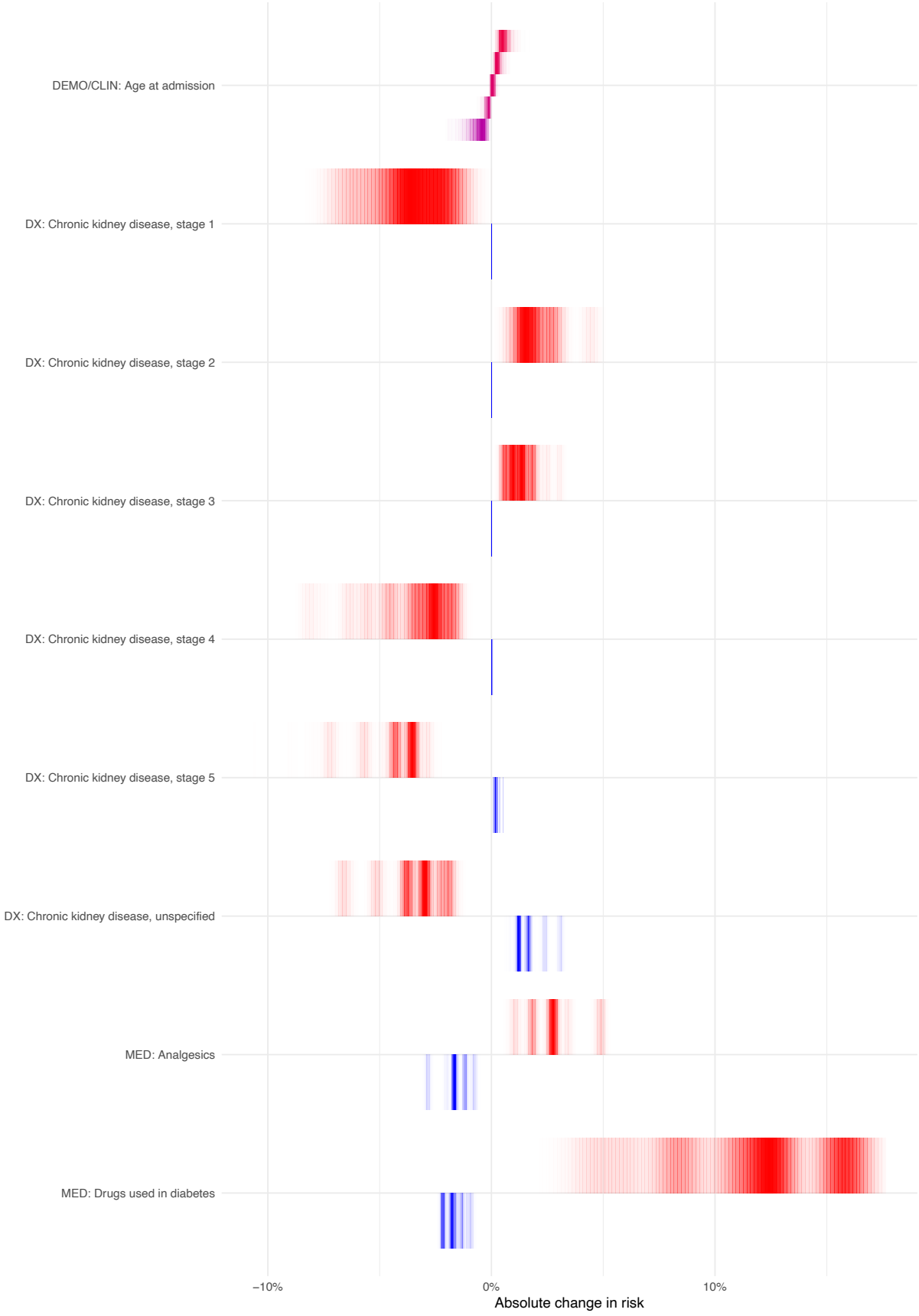


Figure S71: Ref. model, MLP, daily rate >= 1

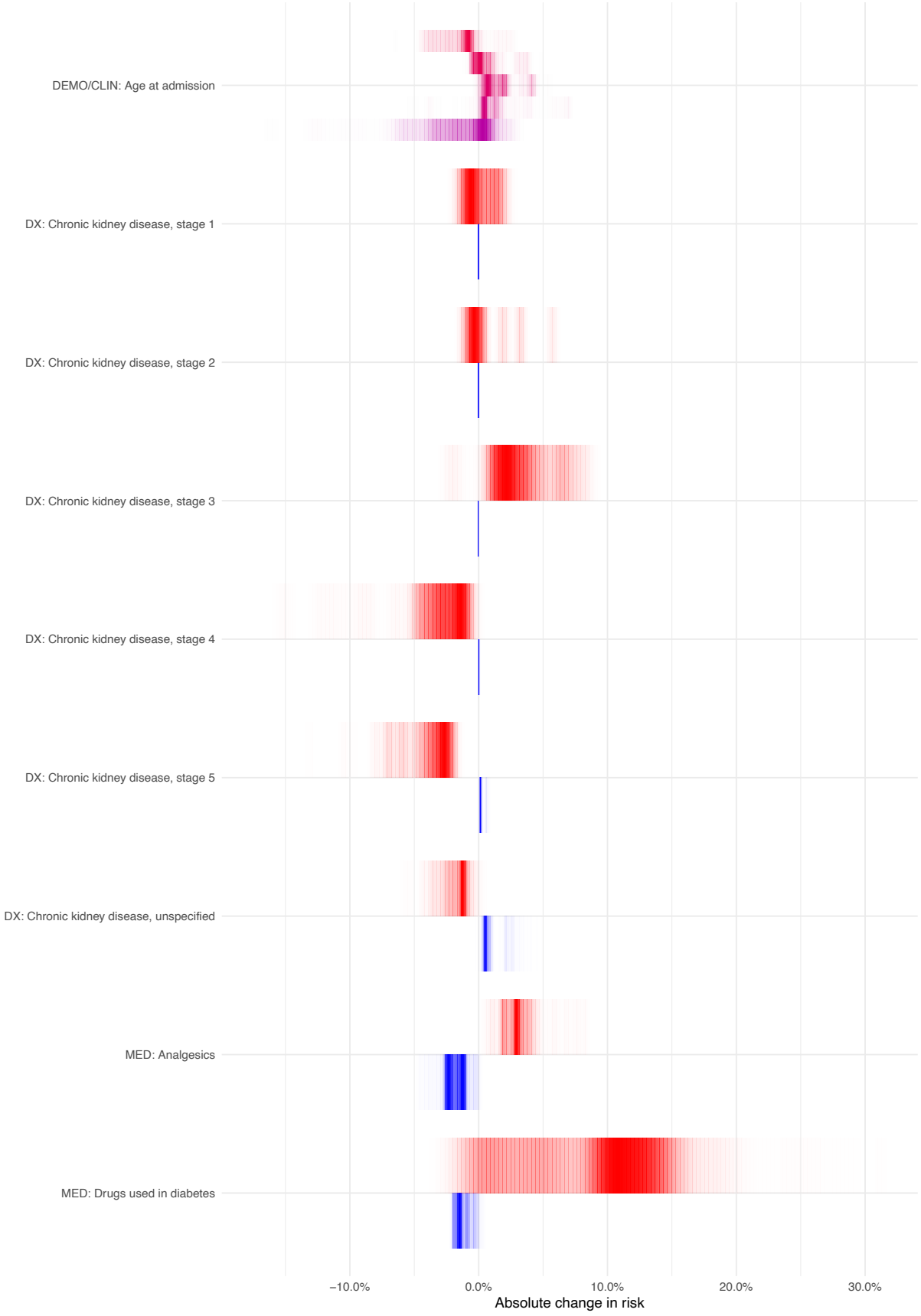
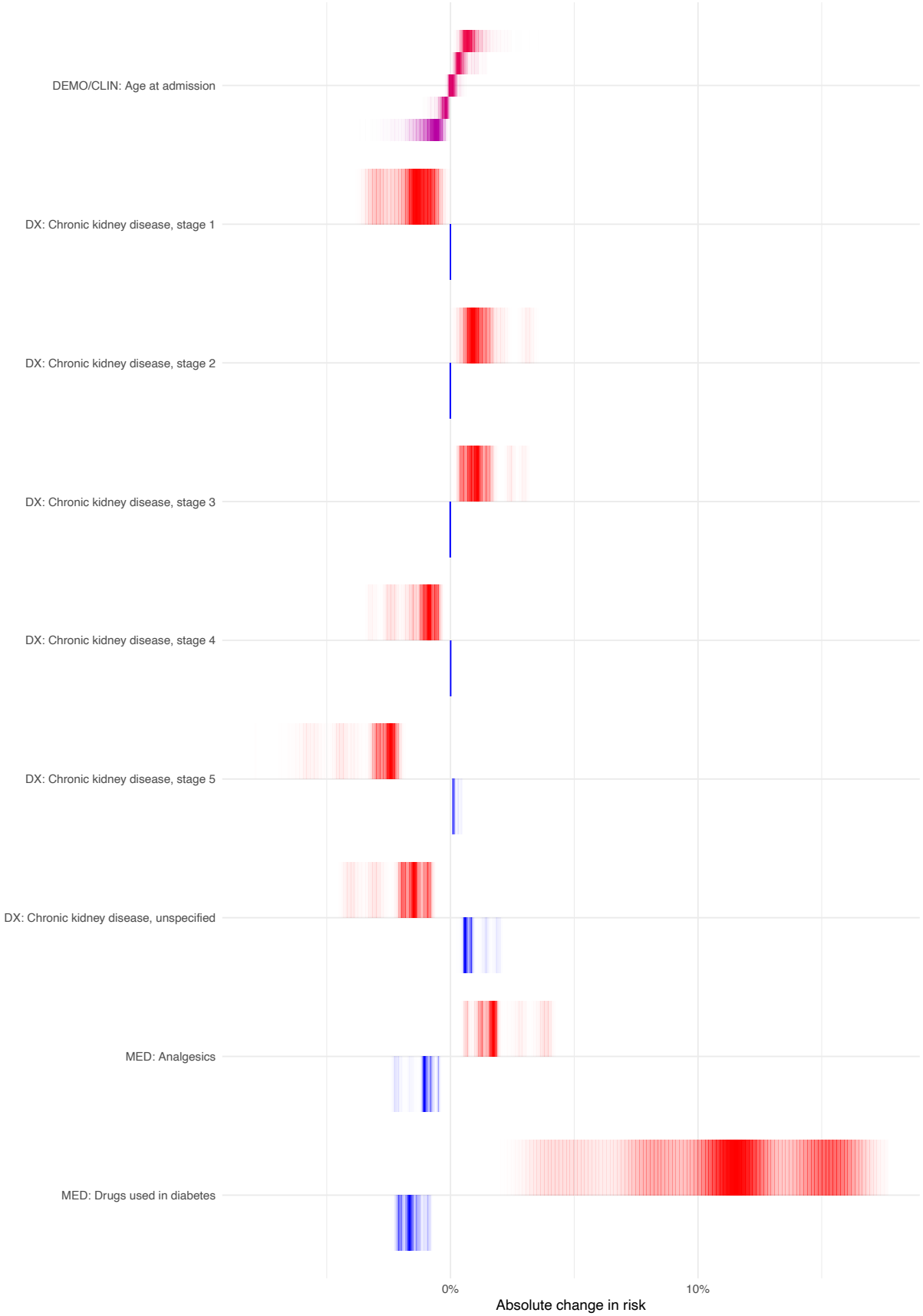


Figure S72: Ref. model, Linear, daily rate >= 1



TRIPOD Checklist: Prediction Model Development

Section/Topic	Item	Checklist Item	Page
Title and abstract			
Title	1	Identify the study as developing and/or validating a multivariable prediction model, the target population, and the outcome to be predicted.	1
Abstract	2	Provide a summary of objectives, study design, setting, participants, sample size, predictors, outcome, statistical analysis, results, and conclusions.	2-3
Introduction			
Background and objectives	3a	Explain the medical context (including whether diagnostic or prognostic) and rationale for developing or validating the multivariable prediction model, including references to existing models.	4-5
	3b	Specify the objectives, including whether the study describes the development or validation of the model or both.	4-5
Methods			
Source of data	4a	Describe the study design or source of data (e.g., randomized trial, cohort, or registry data), separately for the development and validation data sets, if applicable.	5
	4b	Specify the key study dates, including start of accrual; end of accrual; and, if applicable, end of follow-up.	5
Participants	5a	Specify key elements of the study setting (e.g., primary care, secondary care, general population) including number and location of centres.	5
	5b	Describe eligibility criteria for participants.	5
	5c	Give details of treatments received, if relevant.	—
Outcome	6a	Clearly define the outcome that is predicted by the prediction model, including how and when assessed.	6-7
	6b	Report any actions to blind assessment of the outcome to be predicted.	—
Predictors	7a	Clearly define all predictors used in developing or validating the multivariable prediction model, including how and when they were measured.	6-7
	7b	Report any actions to blind assessment of predictors for the outcome and other predictors.	—
Sample size	8	Explain how the study size was arrived at.	Fore-gone
Missing data	9	Describe how missing data were handled (e.g., complete-case analysis, single imputation, multiple imputation) with details of any imputation method.	7
Statistical analysis methods	10a	Describe how predictors were handled in the analyses.	6-7
	10b	Specify type of model, all model-building procedures (including any predictor selection), and method for internal validation.	7-9
	10d	Specify all measures used to assess model performance and, if relevant, to compare multiple models.	9
Risk groups	11	Provide details on how risk groups were created, if done.	—
Results			
Participants	13a	Describe the flow of participants through the study, including the number of participants with and without the outcome and, if applicable, a summary of the follow-up time. A diagram may be helpful.	10 + table 1
	13b	Describe the characteristics of the participants (basic demographics, clinical features, available predictors), including the number of participants with missing data for predictors and outcome.	10 + table 1
Model development	14a	Specify the number of participants and outcome events in each analysis.	Table 1
	14b	If done, report the unadjusted association between each candidate predictor and outcome.	—
Model specification	15a	Present the full prediction model to allow predictions for individuals (i.e., all regression coefficients, and model intercept or baseline survival at a given time point).	Avail. online in repo
	15b	Explain how to the use the prediction model.	—
Model performance	16	Report performance measures (with CIs) for the prediction model.	Table 2 (no CI's)
Discussion			
Limitations	18	Discuss any limitations of the study (such as nonrepresentative sample, few events per predictor, missing data).	14-15
Interpretation	19b	Give an overall interpretation of the results, considering objectives, limitations, and results from similar studies, and other relevant evidence.	12-14
Implications	20	Discuss the potential clinical use of the model and implications for future research.	15
Other information			
Supplementary information	21	Provide information about the availability of supplementary resources, such as study protocol, Web calculator, and data sets.	—
Funding	22	Give the source of funding and the role of the funders for the present study.	16

We recommend using the TRIPOD Checklist in conjunction with the TRIPOD Explanation and Elaboration document.