Online Resource 1: Supplementary tables

Sefunmi Ajayi[1] and Tertsegha Joseph Anande[1*]

[1]Department of Computing, Ulster University, Birmingham, United Kingdom.

*Corresponding author(s). E-mail(s): t.anande@ulster.ac.uk;

Abstract

Supplementary tables for "Hybrid recommendation of healthcare interventions for personalised patient engagement: a leakage-free offline evaluation on synthetic electronic health records". All values are generated by the analysis code in Online Resource 2 from the Synthea April 2020 sample (1,171 patients, 300 interventions). Model abbreviations follow the main text.

Table S1 Random split (primary protocol): all metrics. Pers., personalisation (1 − mean pairwise Jaccard overlap of top-10 lists).

	$K = 5$				$K = 10$				$K = 20$					
Model	P	R	NDCG	HR	P	R	NDCG	HR	P	R	NDCG	HR	Cov.@10	Pers.@10
Random	0.009	0.010	0.010	0.042	0.011	0.030	0.017	0.096	0.011	0.064	0.029	0.188	1.000	0.982
Popularity	0.131	0.290	0.282	0.550	0.100	0.368	0.314	0.682	0.092	0.484	0.369	0.741	0.093	0.448
CBF	0.193	0.320	0.267	0.626	0.158	0.469	0.319	0.764	0.102	0.548	0.354	0.787	0.773	0.914
CBF-D	0.209	0.398	0.349	0.669	0.175	0.604	0.430	0.860	0.128	0.791	0.503	0.940	0.990	0.963
User-kNN CF	0.374	0.576	0.617	0.856	0.246	0.733	0.643	0.935	0.142	0.840	0.681	0.973	0.700	0.854
EASE	0.434	0.693	0.735	0.923	0.278	0.827	0.754	0.971	0.153	0.895	0.779	0.985	0.747	0.890
Hybrid	0.398	0.633	0.668	0.899	0.257	0.783	0.692	0.960	0.145	0.871	0.723	0.980	0.877	0.886
Hybrid-D	0.418	0.672	0.704	0.919	0.266	0.812	0.724	0.969	0.150	0.896	0.755	0.984	0.973	0.916
Hybrid-E	0.452	0.732	0.773	0.953	0.284	0.854	0.787	0.976	0.157	0.922	0.813	0.987	0.987	0.935

Table S2 Temporal leave-latest-out split: all metrics. Pers., personalisation (1 − mean pairwise Jaccard overlap of top-10 lists).

	$K = 5$				$K = 10$				$K = 20$					
Model	P	R	NDCG	HR	P	R	NDCG	HR	P	R	NDCG	HR	Cov.@10	Pers.@10
Random	0.016	0.016	0.019	0.077	0.015	0.030	0.024	0.132	0.016	0.068	0.037	0.254	1.000	0.982
Popularity	0.089	0.162	0.169	0.361	0.068	0.222	0.186	0.479	0.060	0.307	0.221	0.569	0.110	0.452
CBF	0.171	0.274	0.246	0.553	0.134	0.376	0.283	0.654	0.089	0.439	0.310	0.701	0.727	0.915
CBF-D	0.192	0.290	0.278	0.534	0.159	0.450	0.339	0.712	0.118	0.615	0.402	0.840	0.910	0.959
User-kNN CF	0.208	0.317	0.332	0.594	0.149	0.438	0.357	0.721	0.109	0.602	0.417	0.838	0.427	0.794
EASE	0.211	0.327	0.342	0.625	0.153	0.461	0.374	0.755	0.111	0.630	0.436	0.867	0.477	0.808
Hybrid	0.233	0.370	0.379	0.651	0.176	0.524	0.421	0.829	0.122	0.658	0.472	0.912	0.763	0.878
Hybrid-D	0.265	0.414	0.418	0.722	0.192	0.561	0.456	0.846	0.133	0.712	0.515	0.925	0.907	0.902
Hybrid-E	0.268	0.423	0.428	0.747	0.197	0.577	0.471	0.862	0.135	0.726	0.528	0.936	0.893	0.907

Table S3 Hyperparameters selected on inner validation data (grids: $k \in \{5, 10, 15, 20, 30, 50, 75, 100, 150\}$; $\lambda \in \{10, 50, 100, 250, 500, 1000\}$; $\alpha \in \{0, 0.1, \ldots, 1\}$; criterion NDCG@10).

Run	k (kNN)	λ (EASE)	α Hybrid	α Hybrid-D	α Hybrid-E
Random split, seed 2026 (primary)	50	50	0.3	0.5	0.5
Temporal split	150	250	0.4	0.5	0.5
10 further seeds (range)	50–100	10–50	0.3–0.4	0.5–0.6	0.4–0.5

Table S4 Split-seed robustness: random-split protocol repeated with 10 further seeds (2027–2036), each with its own inner tuning. Mean (SD) across seeds.

Model	P@10	R@10	NDCG@10	HR@10
Random	0.013 (0.001)	0.038 (0.005)	0.023 (0.003)	0.117 (0.012)
Popularity	0.098 (0.003)	0.363 (0.009)	0.308 (0.008)	0.675 (0.009)
CBF	0.155 (0.003)	0.463 (0.011)	0.315 (0.008)	0.747 (0.009)
CBF-D	0.173 (0.002)	0.596 (0.007)	0.431 (0.006)	0.852 (0.005)
User-kNN CF	0.244 (0.003)	0.725 (0.010)	0.629 (0.011)	0.926 (0.010)
EASE	0.280 (0.002)	0.830 (0.008)	0.761 (0.010)	0.967 (0.005)
Hybrid	0.254 (0.002)	0.772 (0.006)	0.673 (0.010)	0.950 (0.007)
Hybrid-D	0.266 (0.002)	0.810 (0.007)	0.718 (0.009)	0.962 (0.006)
Hybrid-E	0.287 (0.002)	0.862 (0.006)	0.792 (0.007)	0.980 (0.004)

Table S5 Paired Wilcoxon signed-rank tests on per-patient NDCG@10 (reference minus comparator), Holm-adjusted within each block.

Comparison	Mean difference	W	p_{Holm}
Random split, reference Hybrid			
vs Random	+0.675	611.0	<0.001
vs Popularity	+0.378	18,733.0	<0.001
vs CBF	+0.372	27,318.0	<0.001
vs CBF-D	+0.261	90,282.5	<0.001
vs Hybrid-E	-0.096	125,327.5	<0.001
vs EASE	-0.062	150,000.5	<0.001
vs User-kNN CF	+0.049	152,653.0	<0.001
vs Hybrid-D	-0.032	193,637.5	<0.001
Random split, reference EASE			
vs Random	+0.737	335.5	<0.001
vs Popularity	+0.440	16,334.0	<0.001
vs CBF	+0.434	24,637.0	<0.001
vs CBF-D	+0.323	61,405.0	<0.001
vs User-kNN CF	+0.111	91,111.5	<0.001
vs Hybrid	+0.062	150,000.5	<0.001
vs Hybrid-D	+0.030	194,200.5	<0.001
vs Hybrid-E	-0.034	205,143.5	<0.001
Temporal split, reference Hybrid			
vs Random	+0.397	15,280.5	<0.001
vs Popularity	+0.234	64,373.5	<0.001
vs CBF	+0.137	142,558.5	<0.001
vs User-kNN CF	+0.063	164,929.5	<0.001
vs EASE	+0.047	192,259.5	<0.001
vs CBF-D	+0.082	193,758.5	<0.001
vs Hybrid-E	-0.050	197,046.0	<0.001
vs Hybrid-D	-0.036	212,487.5	<0.001
Temporal split, reference EASE			
vs Random	+0.350	26,198.0	<0.001
vs Popularity	+0.187	70,824.5	<0.001
vs Hybrid-E	-0.097	137,859.0	<0.001
vs Hybrid-D	-0.083	156,483.0	<0.001
vs CBF	+0.090	187,661.0	<0.001
vs Hybrid	-0.047	192,259.5	<0.001
vs User-kNN CF	+0.016	221,668.0	0.002
vs CBF-D	+0.035	223,522.5	0.002

Table S6 Cohort-size analysis (random split). For $n < 1{,}171$, 20 random patient subsamples per size, each fully re-tuned; mean (SD) of NDCG@10. $n = 113$ matches the dissertation cohort size.

Model	$n = 113$	$n = 250$	$n = 500$	$n = 1{,}171$
Random	0.033 (0.007)	0.025 (0.004)	0.023 (0.004)	0.017
Popularity	0.304 (0.036)	0.312 (0.023)	0.309 (0.012)	0.314
CBF	0.338 (0.034)	0.329 (0.015)	0.316 (0.010)	0.319
CBF-D	0.399 (0.040)	0.421 (0.024)	0.435 (0.017)	0.430
User-kNN CF	0.519 (0.035)	0.565 (0.023)	0.594 (0.016)	0.643
EASE	0.591 (0.040)	0.675 (0.025)	0.716 (0.020)	0.754
Hybrid	0.583 (0.037)	0.632 (0.024)	0.649 (0.015)	0.692
Hybrid-D	0.599 (0.035)	0.657 (0.028)	0.691 (0.012)	0.724
Hybrid-E	0.635 (0.034)	0.713 (0.022)	0.755 (0.011)	0.787
Selected α (Hybrid)	0.48 (0.08)	0.40 (0.05)	0.36 (0.05)	0.30

Table S7 Subgroup performance (random split): NDCG@10, mean (95% bootstrap CI).

Attribute	Group	n	Hybrid	Hybrid-D
Sex	Female	551	0.702 (0.680–0.723)	0.731 (0.710–0.751)
Sex	Male	453	0.679 (0.652–0.705)	0.716 (0.689–0.740)
Race	Asian	81	0.736 (0.674–0.792)	0.790 (0.738–0.838)
Race	Black	91	0.731 (0.680–0.779)	0.759 (0.710–0.805)
Race	Other/native	15	0.669 (0.529–0.795)	0.736 (0.612–0.842)
Race	White	817	0.683 (0.664–0.702)	0.713 (0.694–0.731)
Ethnicity	Hispanic	100	0.684 (0.630–0.738)	0.710 (0.659–0.762)
Ethnicity	Nonhispanic	904	0.692 (0.675–0.710)	0.725 (0.708–0.742)
Age band	0-17	166	0.686 (0.641–0.732)	0.742 (0.702–0.780)
Age band	18-44	331	0.698 (0.669–0.726)	0.715 (0.687–0.743)
Age band	45-64	293	0.691 (0.660–0.722)	0.715 (0.685–0.745)
Age band	65+	214	0.686 (0.647–0.724)	0.735 (0.702–0.766)