

Supplementary Table S1

Search strategies

Medline: 1946 to January week 2, 2021	
No.	Searches
1	machine learning/or deep learning/or supervised machine learning/or support vector machine/or unsupervised machine learning/or neural networks, computer/or deep learning/ or bayes theorem/or algorithms/or latent class analysis/
2	((machine or deep) adj3 learn*).ti,ab,kf.
3	perceptron*.ti,ab,kf.
4	((algorithm or bayesian) adj3 learn*).ti,ab,kf.
5	classifier*.ti,ab,kf.
6	(random adj3 forest).ti,ab,kf.
7	(PyTorch or TensorFlow or Keras or Theano or Caffe or Fastai or "scikit-learn" or "scikit learn" or caret or WEKA).ti,ab,kf.
8	(Ensemble adj3 (model* or learn*)).ti,ab,kf.
9	(Neural adj3 network*).ti,ab,kf.
10	(multilayer adj3 perceptron*).ti,ab,kf.
11	("K-nearest neighbo*" or "K nearest neighbo*").ti,ab,kf.
12	((L1 or L2) adj3 (regularization or regularisation)).ti,ab,kf.
13	(Elastic adj3 net*).ti,ab,kf.
14	(Support adj3 vector adj3 machine*).ti,ab,kf.
15	(Discriminant adj3 analysi*).ti,ab,kf.
16	(Naive adj3 Bayes*).ti,ab,kf.
17	(Bayes* adj3 network*).ti,ab,kf.
18	(Random adj3 forest*).ti,ab,kf.
19	((Gradient adj3 boost* or Xgboost*).ti,ab,kf.
20	((adaptive adj3 boost* or AdaBoost).ti,ab,kf.
21	(clinical adj2 predict*).ti,ab,kf.
22	or/1-21
23	((Dataset* or "data-set*" or data or distribution or domain* or covariate* or concept*) adj3 (drift* or shift*)).ti,ab,kf.
24	(Calibration adj3 (drift* or shift*)).ti,ab,kf.
25	(Performance adj3 (drift* or drop*)).ti,ab,kf.
26	recalibrat*.ti,ab,kf.
27	or/23-26
28	22 and 27
29	limit 28 to (english language and humans)
30	28 not 29
Medline Epub Ahead of Print: to January 15, 2021	
No.	Searches
1	((machine or deep) adj3 learn*).ti,ab,kf.
2	perceptron*.ti,ab,kf.
3	((algorithm or bayesian) adj3 learn*).ti,ab,kf.
4	classifier*.ti,ab,kf.
5	(random adj3 forest).ti,ab,kf.
6	(PyTorch or TensorFlow or Keras or Theano or Caffe or Fastai or "scikit-learn" or "scikit learn" or caret or WEKA).ti,ab,kf.

(Continued)

Medline: 1946 to January week 2, 2021	
7	(Ensemble adj3 (model* or learn*)).ti,ab,kf.
8	(Neural adj3 network*).ti,ab,kf.
9	(multilayer adj3 perceptron*).ti,ab,kf.
10	("K-nearest neighbo*" or "K nearest neighbo*").ti,ab,kf.
11	((L1 or L2) adj3 (regularization or regularisation)).ti,ab,kf.
12	(Elastic adj3 net*).ti,ab,kf.
13	(Support adj3 vector adj3 machine*).ti,ab,kf.
14	(Discriminant adj3 analysi*).ti,ab,kf.
15	(Naive adj3 Bayes*).ti,ab,kf.
16	(Bayes* adj3 network*).ti,ab,kf.
17	(Random adj3 forest*).ti,ab,kf.
18	((Gradient adj3 boost*) or Xgboost*).ti,ab,kf.
19	((adaptive adj3 boost*) or AdaBoost).ti,ab,kf.
20	(clinical adj2 predict*).ti,ab,kf.
21	or/1–20
22	((Dataset* or "data-set*" or data or distribution or domain* or covariate* or concept*) adj3 (drift* or shift*)).ti,ab,kf.
23	(Calibration adj3 (drift* or shift*)).ti,ab,kf.
24	(Performance adj3 (drift* or drop*)).ti,ab,kf.
25	recalibrat*.ti,ab,kf.
26	or/22–25
27	21 and 26
Medline In-Process: 1946 to January 15, 2021	
Searches	
1	((machine or deep) adj3 learn*).ti,ab,kf.
2	perceptron*.ti,ab,kf.
3	((algorithm or bayesian) adj3 learn*).ti,ab,kf.
4	classifier*.ti,ab,kf.
5	(random adj3 forest).ti,ab,kf.
6	(PyTorch or TensorFlow or Keras or Theano or Caffe or Fastai or "scikit-learn" or "scikit learn" or caret or WEKA).ti,ab,kf.
7	(Ensemble adj3 (model* or learn*)).ti,ab,kf.
8	(Neural adj3 network*).ti,ab,kf.
9	(multilayer adj3 perceptron*).ti,ab,kf.
10	("K-nearest neighbo*" or "K nearest neighbo*").ti,ab,kf.
11	((L1 or L2) adj3 (regularization or regularisation)).ti,ab,kf.
12	(Elastic adj3 net*).ti,ab,kf.
13	(Support adj3 vector adj3 machine*).ti,ab,kf.
14	(Discriminant adj3 analysi*).ti,ab,kf.
15	(Naive adj3 Bayes*).ti,ab,kf.
16	(Bayes* adj3 network*).ti,ab,kf.
17	(Random adj3 forest*).ti,ab,kf.
18	((Gradient adj3 boost*) or Xgboost*).ti,ab,kf.
19	((adaptive adj3 boost*) or AdaBoost).ti,ab,kf.
20	(clinical adj2 predict*).ti,ab,kf.

(Continued)

(Continued)

Medline: 1946 to January week 2, 2021	
21	or/1–20
22	((Dataset* or “data-set*” or data or distribution or domain* or covariate* or concept*) adj3 (drift* or shift*)).ti,ab,kf.
23	(Calibration adj3 (drift* or shift*)).ti,ab,kf.
24	(Performance adj3 (drift* or drop*)).ti,ab,kf.
25	recalibrat*.ti,ab,kf.
26	or/22–25
27	21 and 26
Embase: 1947 to January 15, 2021	
Searches	
1	machine learning/or exp artificial neural network/or bayesian learning/or exp classification algorithm/or classifier/or k nearest neighbor/or exp learning algorithm/or network learning/or exp perceptron/or random forest/or semi supervised machine learning/or supervised machine learning/or exp support vector machine/or unsupervised machine learning/
2	(machine adj2 learn*).ti,ab,kw.
3	(PyTorch or TensorFlow or Keras or Theano or Caffe or Fastai or “scikit-learn” or “scikit learn” or caret or WEKA).ti,ab,kw. [****Scikit learn added to original search statement - Caffe** or “Caffe2” included in this search statement adding the the number 2 would not add additional references****]
4	((machine or deep) adj3 learn*).ti,ab,kw.
5	(Ensemble adj3 (model* or learn*)).ti,ab,kw.
6	(Neural adj3 network*).ti,ab,kw.
7	(multilayer adj3 perceptron*).ti,ab,kw.
8	(“K-nearest neighbo*” or “K nearest neighbo*”).ti,ab,kw.
9	((L1 or L2) adj3 (regularization or regularisation)).ti,ab,kw.
10	(Elastic adj3 net*).ti,ab,kw.
11	(Support adj3 vector adj3 machine*).ti,ab,kw.
12	(Discriminant adj3 analysi*).ti,ab,kw.
13	(Naive adj3 Bayes*).ti,ab,kw.
14	(Bayes* adj3 network*).ti,ab,kw.
15	(Random adj3 forest*).ti,ab,kw.
16	((Gradient adj3 boost*) or Xgboost*).ti,ab,kw.
17	((adaptive adj3 boost*) or AdaBoost).ti,ab,kw.
18	(clinical adj2 predict*).ti,ab,kw.
19	or/1–18
20	((Dataset* or “data-set*” or data or distribution or domain* or covariate* or concept*) adj3 (drift* or shift*)).ti,ab,kw.
21	(Calibration adj3 (drift* or shift*)).ti,ab,kw.
22	(Performance adj3 (drift* or drop*)).ti,ab,kw. [****performance dr* is too broad as a textword it will also pick up drive which you probably don’t want. In addition most search platforms require a minimum of 3 letters before the truncation symbol to run an efficient search****]
23	recalibrat*.ti,ab,kw.
24	or/20–23
25	19 and 24
26	limit 25 to (human and English language)
27	25 not 26
APA PsycInfo: 1806 to January week 2, 2021	
Searches	

(Continued)

Medline: 1946 to January week 2, 2021	
1	machine learning/or computational reinforcement learning/or inductive logic programming/or machine learning algorithms/or extreme learning machine/ or unsupervised learning/or deep neural networks/or artificial neural networks/or deep neural networks/or “pattern recognition (computer science)”/or feature extraction/or image analysis/or bayesian analysis/
2	((machine or deep) adj3 learn*).ti,ab,id.
3	perceptron*.ti,ab,id.
4	((algorithm or bayesian) adj3 learn*).ti,ab,id.
5	classifier*.ti,ab,id.
6	(random adj3 forest).ti,ab,id.
7	(PyTorch or TensorFlow or Keras or Theano or Caffe or Fastai or “scikit-learn” or “scikit learn” or caret or WEKA).ti,ab, id.
8	(Ensemble adj3 (model* or learn*)).ti,ab,id.
9	(Neural adj3 network*).ti,ab,id.
10	(multilayer adj3 perceptron*).ti,ab,id.
11	(“K-nearest neighbo*” or “K nearest neighbo*”).ti,ab,id.
12	((L1 or L2) adj3 (regularization or regularisation)).ti,ab,id.
13	(Elastic adj3 net*).ti,ab,id.
14	(Support adj3 vector adj3 machine*).ti,ab,id.
15	(Discriminant adj3 analysi*).ti,ab,id.
16	(Naive adj3 Bayes*).ti,ab,id.
17	(Bayes* adj3 network*).ti,ab,id.
18	(Random adj3 forest*).ti,ab,id.
19	((Gradient adj3 boost*) or Xgboost*).ti,ab,id.
20	((adaptive adj3 boost*) or AdaBoost).ti,ab,id.
21	or/1–20 [****machine learning terms****]
22	((Dataset* or “data-set*” or data or distribution or domain* or covariate* or concept*) adj3 (drift* or shift*)).ti,ab,id.
23	(Calibration adj3 (drift* or shift*)).ti,ab,id.
24	(Performance adj3 (drift* or drop*)).ti,ab,id
25	or/22–24
26	21 and 25
	Limits used to divide into 2 groups for screening – include if you decide to only review group 1
27	limit 26 to (human and english language)
28	26 not 27
arXiv: to January 18, 2021	
Set	Query
1	(ti:(machine AND learning) AND abs:drift)
2	(ti:(machine AND learning) AND abs:shift)
3	(ti:(machine AND learning) AND abs:clinical)
4	1 or 2 or 3
Web of Science: to January 18, 2021	
Set	Query
1	TS = ((machine OR deep OR network*) NEAR/3 learn*) OR TS = (artificial NEAR/3 neural NEAR/3 network*) OR TS = (bayes* NEAR/3 learn*) OR TS = (classification NEAR/3 algorithm*) OR TS = (learning NEAR/3 algorithm*) OR TS = perceptron* OR TS = (random NEAR/3 forest) OR TS = (support NEAR/3 vector NEAR/3 machine*) OR TS = (Neural NEAR/3 network*) OR TS = (ensemble NEAR/3 model*) OR TS = (K-nearest neighbo*) OR TS = (K nearest neighbo*) OR TS = ((L1 OR L2) NEAR/3 (regulariz* OR regularis*)) OR TS = (elastic NEAR/3 net*) OR TS = (Support

(Continued)

(Continued)

	Medline: 1946 to January week 2, 2021
	NEAR/3 vector NEAR/3 machine*) OR TS = (Discriminant NEAR/3 analysi*) OR TS = (naive NEAR/3 Bayesian) OR TS = (random NEAR/3 forest) OR TS = (Gradient NEAR/3 boost*) OR TS = Xgboost* OR TS = (adaptive NEAR/3 boost*) OR TS = AdaBoost OR TS = (PyTorch OR TensorFlow OR Keras OR Theano OR Caffe OR Fastai OR "scikit-learn" OR caret OR WEKA) OR TS = (scikit NEAR/1 learn) OR TS = (clinical NEAR/3 predict*)
2	TS = ((Dataset* or dataset* or data or distribution or domain* or covariate* or concept*) NEAR/3 (drift* or shift*)) OR TS = recalibrate*
3	TS = (Calibration NEAR/3 (drift* or shift*))
4	TS = (Performance NEAR/3 (drift* or drop*))
5	#2 OR #3 OR 4
6	1 AND 5
7	TS = human* OR TS = people* OR TS = patient OR TS = patients OR TS = medical* OR TS = clinical*
8	6 AND 7
9	TS = Animal* OR TS = environment* OR TS = tree* OR TS = plant* OR TS = vegetation* OR TS = ecolog* OR TS = geolog* OR TS = industr* OR TS = manufactur*
10	6 NOT 9
11	10 NOT 8
12	8 OR 11

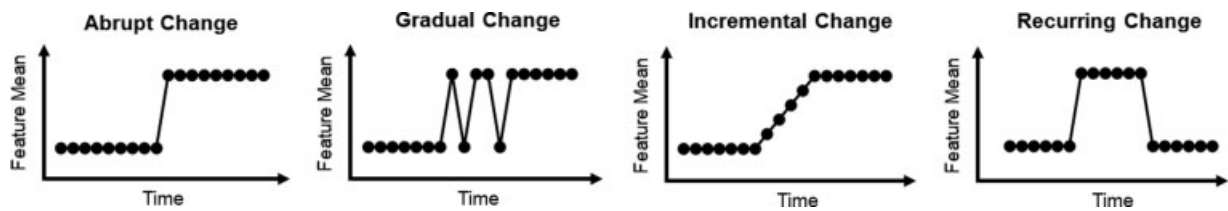
Supplementary Table S2 Details of included studies (n = 15)

Study	Year	Age	Cohort	Label	ML algorithm(s)	Data source	Single or multiple institution	Number of mitigation strategies	Number of time periods	Calibration deterioration	Discrimination deterioration
Feng ²⁸	2020	Adult	ICU admissions	Mortality	Random forest	EHR	Single	6	11	NA	NA
Adam ²²	2020	Adult	ICU admissions	Mortality	3 different	EHR	Single	5	22	NA	Yes
Davis ¹⁶	2019	NA	Inpatient admissions	30-day mortality and acute kidney injury	Logistic regression	Administrative	Multiple	5	3	Yes	NA
Davis ²⁴	2019	NA	Inpatient admissions	30-day mortality	4 different	Administrative	Multiple	5	6	Yes	No
Nestor ²⁵	2019	Adult	ICU admissions	In-ICU mortality and long length-of-stay	4 different	EHR	Single	3	10	NA	Yes
Siregar ¹⁷	2019	Adult	Cardiac surgery	In-hospital mortality	Logistic regression	Registry	Multiple	5	1	Yes	No
Nestor ²⁶	2018	Adult	ICU admissions	Mortality and long length-of-stay	Random forest	EHR	Single	1	10	NA	Yes
Su ²³	2018	Adult	Coronary artery bypass graft surgery	Mortality	Logistic regression	Registry	Multiple	6	1	Yes	No
Davis ³⁰	2017	Adult	Inpatient admissions	Acute kidney injury	7 different	Administrative	Multiple	1	36	Yes	No
Davis ²⁹	2017	Adult	Inpatient admissions	30-day mortality	7 different	Administrative	Multiple	1	28	Yes	No
Siregar ¹⁸	2016	Adult	Cardiac surgery	In-hospital mortality	Logistic regression	Registry	Multiple	6	5	Yes	No
Stroble ²⁷	2015	Adult	Prostate biopsy	Prostate cancer	2 different	Trial	Multiple	5	6 to 14	Yes	Uncertain
Hickey ¹⁹	2013	Adult	Cardiac surgery	In-hospital mortality	Logistic regression	Registry	Multiple	3	5 and 10	Yes	NA
Janssen ²⁰	2008	Adult	Surgery	Severe postoperative pain	Logistic regression	Trial and EHR	Multiple	5	1	Yes	Uncertain
Perry ²¹	2003	Neonates	Premature babies	Mortality	Logistic regression	Trial	Multiple	1	1	Yes	Uncertain

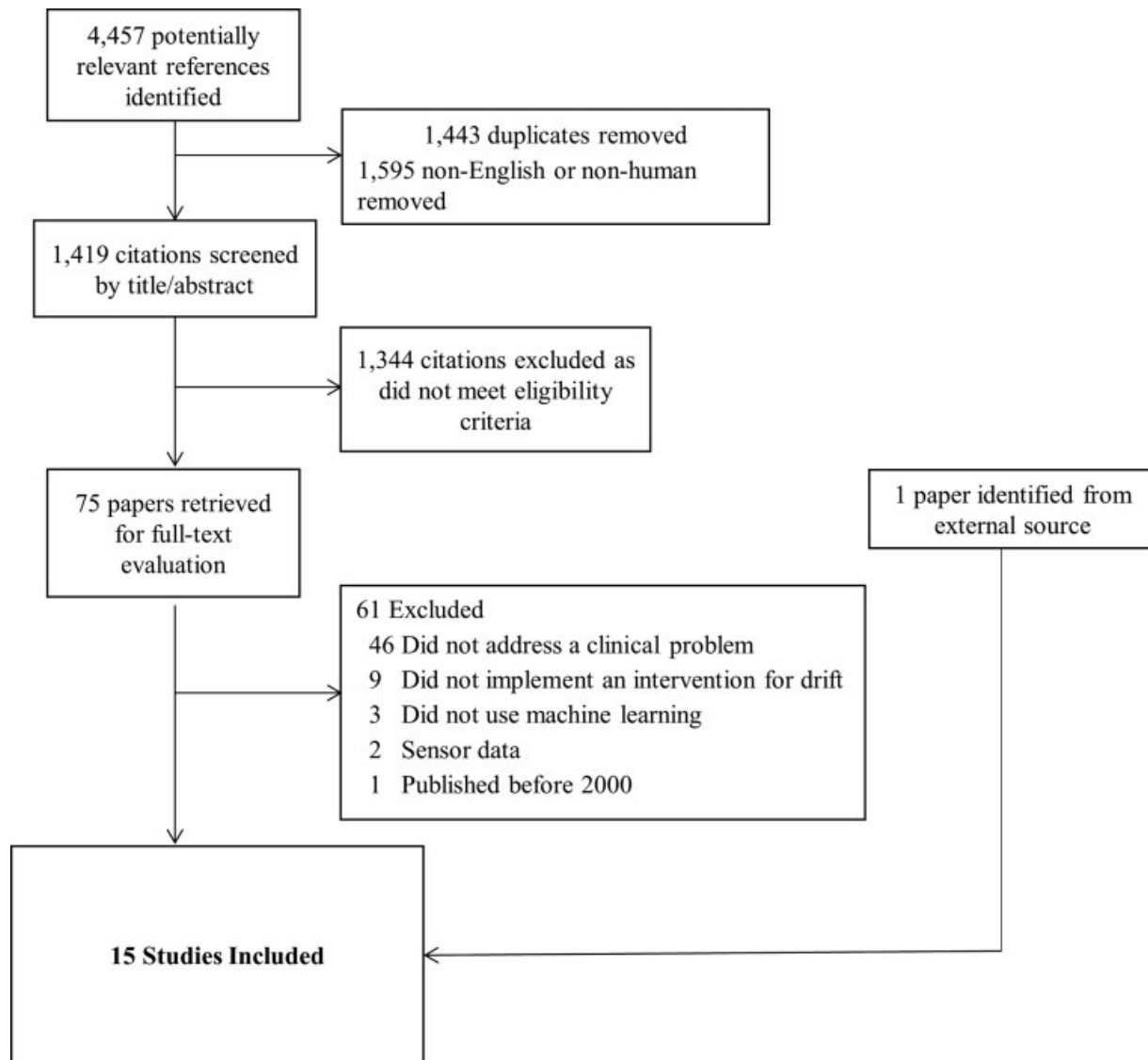
Abbreviations: EHR, electronic health record; ICU, intensive care unit; ML, machine learning; Pub, published.

Supplementary Table S3 Quality assessment ($n = 15$)

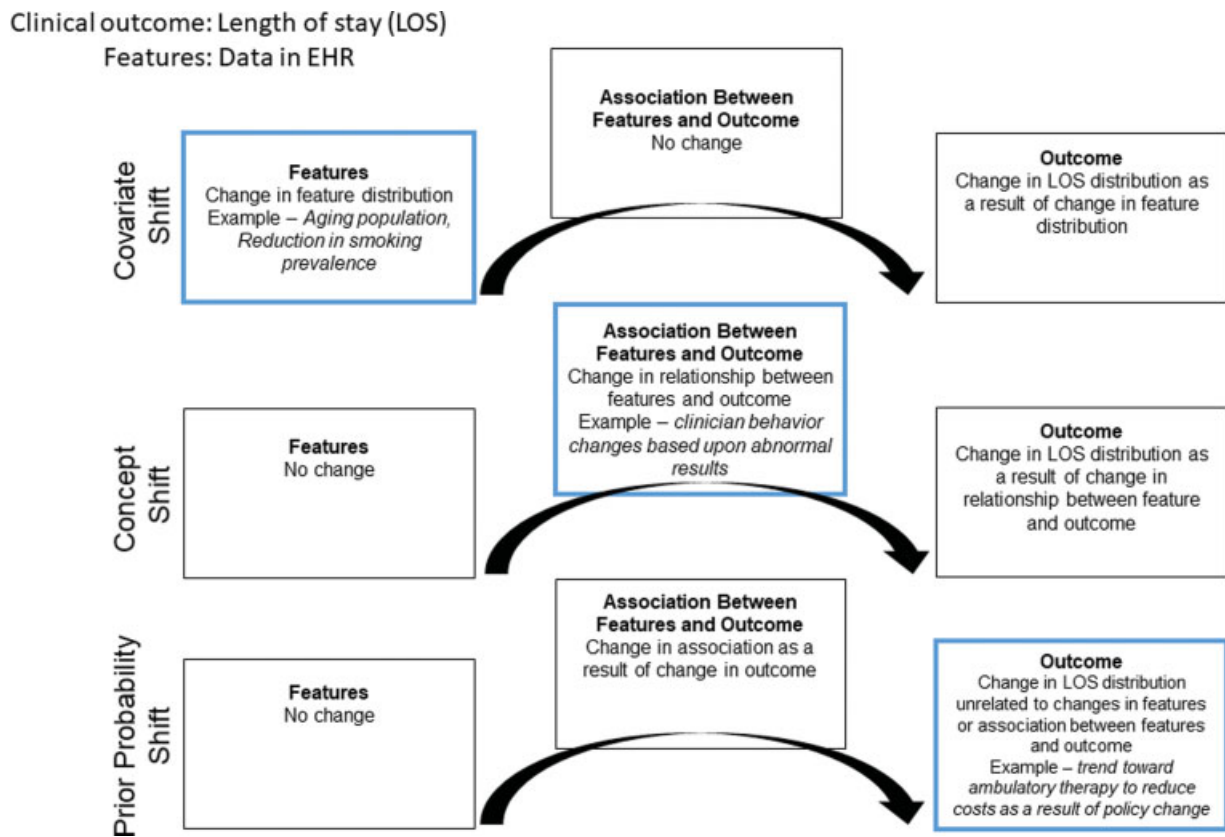
Attribute	Number describing (%)
Inclusion and exclusion criteria	12 (80)
Sample	13 (87)
Response variable	8 (53)
Information leakage prevention	13 (87)
Data preprocessing	8 (53)
Data splitting	13 (87)
Validation metrics	12 (80)
Code availability	2 (13)



Supplementary Fig. S1 Examples of change in data distribution over time. Change in data distribution over time illustrated using the mean of a single feature, for example, age. Adapted from Gama et al.¹²



Supplementary Fig. S2 Flow diagram of study identification, selection, and reasons for exclusion.



Supplementary Fig. S3 Figure illustrates types of dataset shift. The arrows in the diagrams imply the assumed direction of influence from features (X) to clinical outcome (Y), that is, an $X \rightarrow Y$ assumption. Prior probability shift may carry the assumption of a reversed causal relationship in which Y determines X.⁹ Such a relationship is more commonly associated with medical imaging where the disease (Y) determines the pattern of pixels (X). In our prior probability shift diagram (row 3), we underscore that the change in clinical outcome is not due to upstream shifts in feature distribution or the association between features and outcome, but because of hidden covariates not represented in the feature space. This change in outcome distribution can subsequently impact the association between features and outcome and may also impact the distribution of features.