Generalizability of Risk Stratification Algorithms for Exacerbations in COPD

Joseph Khoa Ho, PharmD; Abdollah Safari, PhD; Amin Adibi, MSc; Don D. Sin, MD; Kate Johnson, PhD; and Mohsen Sadatsafavi, MD, PhD; for the IMPACT Study

Published : December 08, 2022

CHEST JOURNAL

Introduction :

COPD, a highly prevalent chronic pulmonary disease, is one of the most common reasons for hospital admissions and is associated with increased health care use, morbidity and death, and decreased quality of life.

Exacerbation prevention is a cornerstone of contemporary COPD management. According to the Global Initiative for Chronic Obstructive Lung Disease, exacerbation prevention is achieved by directing pharmacotherapy based on patients' prior 12-month history of moderate or severe exacerbations.

Although exacerbation history is the single best predictor for future exacerbations, relying on history alone for risk prediction may be suboptimal because a growing body of evidence suggests that the predictability of exacerbations based on history alone may be less reliable than previously believed.

Clinical prediction tools are multivariable models that combine several patient characteristics to increase the accuracy of risk stratification.

Unlike exacerbation history alone, they can quantify (eg, in risk percentage) and communicate future risk with patients to enable shared decision-making.

Importantly, prediction models are flexible and can be updated to accommodate baseline risk in different settings.

Objectives of this study :

1/ to evaluate the clinical utility of risk stratification algorithms, including multivariable risk prediction models and exacerbation history alone, across cohorts with different exacerbation risks

2/ to determine whether model recalibration with the observed exacerbation risk within each sample can improve the algorithms' clinical utility

MODELS :

We compared the Global Initiative for Chronic Obstructive Lung Disease risk stratification label of "frequent exacerbators" (defined as having >=2 moderate or>=1 severe exacerbations) with two published, validated COPD exacerbation risk prediction models

The Bretens model :

This was the only model that a 2017 comprehensive systematic review of COPD exacerbation risk prediction models considered to have undergone robust development and external validation

The Bertens model uses four predictors :

1/ the number of exacerbations in the previous 12 months

2/ VEMS expressed as % predicted

2 /pack-years of smoking

4/ a history of vascular disease

ACCEPT model :

Acute COPD Exacerbation Prediction Tool (ACCEPT), which was developed to enable individualized predictions of the rate and severity of exacerbations.

ACCEPT uses up to 12 predictors :

1/the number of nonsevere and severe exacerbations in the previous 12 months

2/ age

3/sex

4/BMI

5/smoking status

6/VEMS % predicted after bronchodilator administration

7/ use of statins as a surrogate for cardiovascular disease risk

8/ domiciliary oxygen therapy

9/The St. George’s Respiratory Questionnaire (SGRQ) score (or COPD Assessment Test)

10/current use of inhaled long-acting muscarinic receptor antagonists (LAMAs)

11/long-acting b2 agonists

12/inhaled corticosteroids are optional predictors

Sources of data :

We used data from 3 randomized clinical trials representing three levels of exacerbation risk:

1/the placebo arm of the Study to Understand Mortality and Morbidity in COPD (SUMMIT =2421) .

2/the Long-term Oxygen Treatment Trial (LOTT; n=595)

3/ the placebo arm of the Towards a Revolution in COPD Health (TORCH; n =1,091).

TABLE 1] Baseline Characteristics of the Study Sample of the Included Trials

Characteristic	SUMMIT	LOTT	TORCH
No.	2,421	595	1,091
Follow-up, y	0.83 ± 0.24	0.95 ± 0.15	0.96 ± 0.13
Age, y	65.9 ± 7.9	69.7 ± 7.3	65.5 ± 8.2
Male sex	1,818 (75.1)	434 (72.9)	842 (77.2)
BMI, kg/m^2	28.1 ± 5.7	28.8 ± 6.3	25.6 ± 5.3
Current smoker	1,144 (47.2)	134 (22.5)	462 (42.3)
Smoking history, pack-y	40.6 ± 24.5	60.7 ± 32.7	48.4 ± 26.5
LAMA	624 (25.8)	374 (62.9)	601[a] (55.1)
LABA	1,308 (54.0)	436 (73.3)	372 (34.1)
ICS	1,258 (52.0)	442 (74.3)	518 (47.5)
SGRQ score	43.7 ± 16.2	48.7 ± 18.5	45.7 ± 16.9
FEV_1, % predicted	58.3 ± 11.4	46.7 ± 16.8	44.9 ± 13.9
History of ≥ 1 moderate or severe exacerbation	0.23	0.38	0.48
Observed risk of ≥ 1 moderate or severe exacerbation	0.22	0.38	0.52

Data are presented as No. (%) or mean ± SD. ICS = inhaled corticosteroid; LABA = long-acting β2 agonist; LAMA = long-acting muscarinic receptor antagonist; LOTT = Long-term Oxygen Treatment Trial; SGRQ = St. George's Respiratory Questionnaire; SUMMIT = Study to Understand Mortality and Morbidity in COPD; TORCH = Towards a Revolution in COPD Health.

[a]Based on imputed values.

The primary outcome was the prospective 12-month risk of a moderate or severe exacerbation :

Moderate exacerbations were those that required treatment with systemic corticosteroids, antibiotics, or both.

Severe exacerbations were those that resulted in ED visits or hospitalizations

Discrimination of Risk Stratification Algorithms Model discrimination (the ability of a risk stratification algorithm to distinguish high-risk vs low-risk patients) was assessed by receiver operating characteristic curves and calculating the area under the receiver operating characteristic curve (AUC).

Timedependent (at 12 months) receiver operating characteristic curves and AUCs were used to account for the variability in follow-up time across patients and were compared using the DeLong test.

Clinical Utility of Risk Stratification Algorithms Clinical utility was measured through net benefit calculations using the decision curve analysis (DCA).

Whereas discrimination evaluates the statistical performance, the DCA provides a comprehensive assessment of the clinical utility of risk stratification to inform treatment decision

Results :

SUMMIT included 2,421 patients (mean age, 65.9 years; 75.1% male) and contributed 636 exacerbations.

LOTT included 595 patients (mean age, 69.7 years; 72.9% male) and contributed 369 exacerbations.

TORCH included 1,091 patients (mean age, 65.5 years; 77.2% male) and contributed 1,074 exacerbations.

=> The annual risk of exacerbations in SUMMIT, LOTT, and TORCH was 0.22, 0.38, and 0.52, respectively

The AUC for exacerbation history alone in predicting future exacerbations in SUMMIT, LOTT, and TORCH was 0.59 (95% CI, 0.57-0.61), 0.63 (95% CI, 0.59-0.67), and 0.65 (95% CI, 0.63-0.68), respectively.

The Bertens model showed a higher AUC compared with exacerbation history alone in SUMMIT (increase of 0.10; P <0.001) and TORCH (increase of 0.05; P <0.001), but not in LOTT (increase of 0.01; P = 0.84).

ACCEPT showed a higher AUC compared with exacerbation history alone in all study samples, by 0.08 (P < .001), 0.07 (P =0.001), and 0.10 (P < 0.001), respectively.

=> Compared with the Bertens model, ACCEPT showed a higher AUC by 0.06 (P =0.001) in LOTT and 0.05 (P < .001) in TORCH, whereas the AUCs were not different in SUMMIT (change of –0.02; P =0.16).

TABLE 2] Time-Dependent AUC at 12 Months

Variable	SUMMIT	LOTT	TORCH
Exacerbation history	0.59 (0.57-0.61)	0.63 (0.59-0.67)	0.65 (0.63-0.68)
Bertens model	0.69[a] (0.66-0.72)	0.64 (0.59-0.69)	0.70[a] (0.67-0.74)
ACCEPT	0.67[a] (0.63-0.70)	0.70[a,b] (0.65-0.74)	0.75[a,b] (0.72-0.78)

Data are presented as AUC (95% CI). ACCEPT = Acute COPD Exacerbation Prediction Tool; AUC = area under the receiver operating characteristic curve; LOTT = Long-term Oxygen Treatment Trial; SUMMIT = Study to Understand Mortality and Morbidity in COPD; TORCH = Towards a Revolution in COPD Health.

[a]Statistically significant compared with exacerbation history.

[b]Statistically significant compared with the Bertens model.

Calibration :

In SUMMIT, the Bertens model was well calibrated and showed good agreement between observed and predicted risk (observed risk of 0.22 vs predicted risk of 0.20) but ACCEPT overestimated the risk with a predicted annual risk of 0.34.

In LOTT, the Bertens model underestimated the risk (observed risk of 0.38 vs predicted risk of 0.27), whereas ACCEPT overestimated the risk (predicted risk, 0.53).

In TORCH, the Bertens model underestimated the risk (observed risk of 0.52 vs predicted risk of 0.28), whereas ACCEPT was well calibrated with a predicted risk of 0.51.

After model recalibration, the mean adjusted predicted risk of exacerbation for both the Bertens model and ACCEPT matched the observed risks in the study samples.

Because the Bertens model already was well calibrated in SUMMIT and ACCEPT was well calibrated in TORCH, the improvements were relatively minor for each model in the respective studies.

In SUMMIT, the Bertens model and exacerbation history outperformed ACCEPT. The Bertens model dominated at the low threshold, whereas exacerbation history dominated at the high threshold.

In LOTT, no risk stratification algorithm clearly dominated. ACCEPT was the best at the low threshold, the Bertens model was best at the medium threshold, and exacerbation history was best at the high threshold.

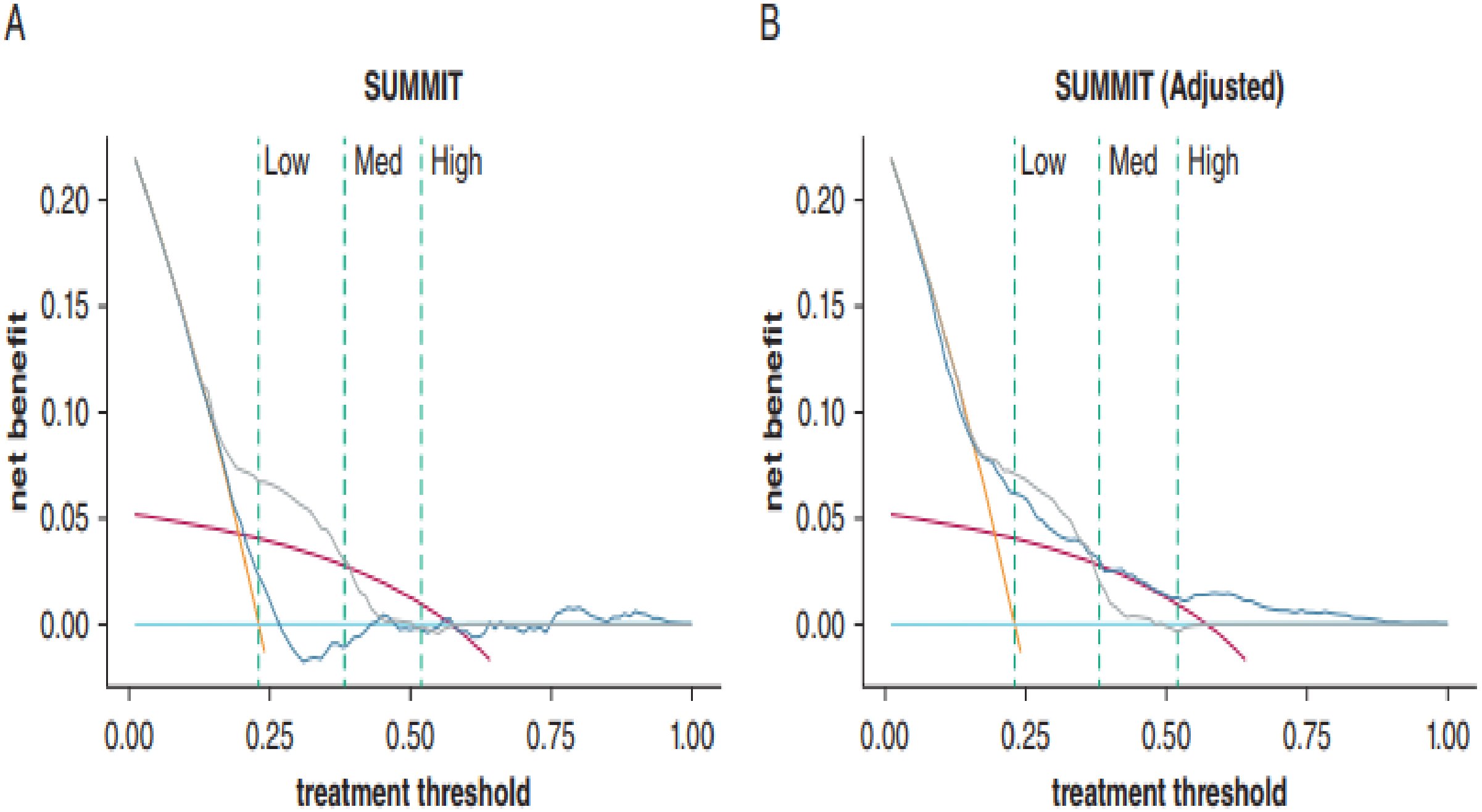
In TORCH, ACCEPT dominated the other algorithms at all three threshold values.

All three risk stratification algorithms were associated with a risk of harm (their net benefit being lower than that of treating no patients or treating all patients).

Exacerbation history showed lower net benefit than treating all patients at the low threshold in LOTT and at the low and medium thresholds in TORCH.

- The Bertens model showed lower net benefit than treating all patients at the low threshold and treating no patients at the high threshold in LOTT and was worse than treating all patients at the low threshold in TORCH.

- ACCEPT showed lower net benefit than treating no patients at the medium threshold in SUMMIT and at the high threshold in LOTT.


A
SUMMIT
Low
Med
High
0.20
0.15
0.10
0.05
0.00
net benefit
0.00
0.25
0.50
0.75
1.00
treatment threshold
B
SUMMIT (Adjusted)
Low
Med
High
0.20
0.15
0.10
0.05
0.00
net benefit
0.00
0.25
0.50
0.75
1.00
treatment threshold

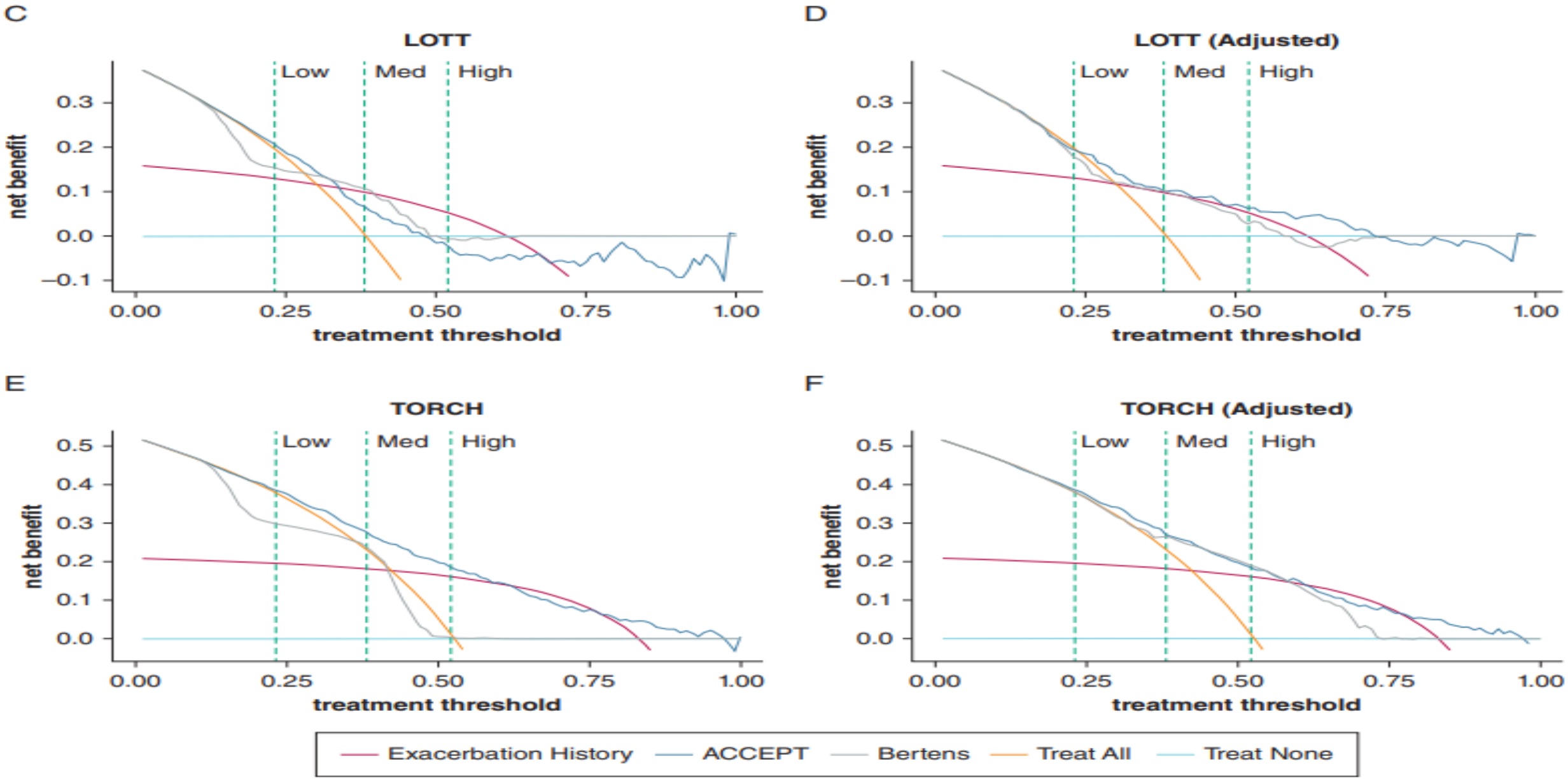


Figure 1 – *A-F, Decision curve analysis comparing the net benefit of the risk stratification algorithms when unadjusted and adjusted with the sample-specific exacerbation risk: unadjusted prediction models in SUMMIT (A), adjusted prediction models in SUMMIT (B), unadjusted prediction models in LOTT (C), adjusted prediction models in LOTT (D), unadjusted prediction models in TORCH (E), and adjusted prediction models in TORCH (F). LOTT = Long-term Oxygen Treatment Trial; SUMMIT = Study to Understand Mortality and Morbidity in COPD; TORCH = Towards a Revolution in COPD Health.*

Compared with exacerbation history, ACCEPT showed better performance in all three samples (change in AUC, 0.08, 0.07, and 0.10, in SUMMIT, LOTT, and TORCH, respectively; P <= 0.001 for all).

The Bertens model showed better performance compared with exacerbation history in SUMMIT and TORCH (change in AUC, 0.10 and 0.05, respectively; P < .001 for both), but not in LOTT.

No algorithm was superior in clinical utility across all samples.

Before recalibration, the Bertens model generally outperformed the other algorithms in low-risk settings, whereas ACCEPT outperformed others in high-risk settings.

All three algorithms showed the risk of harm (providing lower net benefit than not using any risk stratification).

After recalibration, risk of harm was mitigated substantially for both prediction models

TABLE 3] Dominating Risk Stratification Algorithms at the Three Threshold Levels

Variable	SUMMIT	LOTT	TORCH
Unadjusted			
Low	Bertens model	ACCEPT	ACCEPT
Medium	Bertens model/exacerbation history	Bertens model	ACCEPT
High	Exacerbation history	Exacerbation history	ACCEPT
Adjusted			
Low	Bertens model	ACCEPT	ACCEPT/Bertens model
Medium	ACCEPT/exacerbation history	ACCEPT/Bertens model/exacerbation history	ACCEPT/Bertens model
High	ACCEPT/exacerbation history	ACCEPT	ACCEPT/Bertens model

ACCEPT = Acute COPD Exacerbation Prediction Tool; LOTT = Long-term Oxygen Treatment Trial; SUMMIT = Study to Understand Mortality and Morbidity in COPD; TORCH = Towards a Revolution in COPD Health.

Conclusion

Exacerbation history alone is unlikely to provide clinical utility for predicting COPD exacerbations in all settings and could be associated with a risk of harm.

Prediction models have superior predictive performance, but require setting-specific recalibration to confer higher clinical utility