PHILIPS

Critical care

White paper

Philips Critical Care Outcomes Prediction models for ICU Length of Stay

This white paper provides:

- A description of the methods used to develop the Philips Critical Care Outcomes Prediction model (CCOPM) for ICU Length of Stay (LOS)
- A comprehensive performance review of this model in a side-by-side comparison with ICU LOS predictions provided by Acute Physiology and Chronic Health Evaluation (APACHE, from Cerner Corporation, Kansas City, MO).

Executive summary

Background: Philips provides quarterly benchmarking reports for APACHE[1] risk-adjusted models for ICU LOS[2] which customers can use as an indicator of efficiency of care in the ICU. These models are subject to drift over time and must be periodically recalibrated. The latest recalibration for APACHE was performed in 2016[3] and has shown limitations with respect to high-risk/non-surviving cohorts[4].

This white paper introduces Philips CCOPM – LOS, a new ICU LOS model for benchmarking. This model is designed to offer:

- Improved performance over APACHE IVa and IVb with respect to how variations in predicted LOS reflect variations in true LOS, with lower absolute error and improved actual-to-predicted ratios
- Improved calibration performance compared to APACHE IVa and IVb
- Improved predictions for patient subgroups, producing enhanced predictions for surviving and non-surviving cohorts

Methods:

The data used for developing CCOPM – LOS was obtained from the Philips eICU Research Institute (eRI) database comprised of de-identified physiologic, diagnostic and treatment records of all patient stays from participating customers. To develop the model, eRI data from 2017 to 2019 was used. The model selected for this task was the DeepHit model developed by Lee et al[5]. It uses a deep-learning framework for time-to-event prediction. A technical manuscript will describe the model in detail and show the results on a validation set selected from the 2017-2019 data. This white paper provides further validation of the model across the entire eICU benchmarking cohort. CCOPM – LOS predictions were generated for all eICU customers from 2004 to Q1 2022 and a comparative analysis was made with a subset of those customers with active APACHE licenses during the same time period.

This paper examines the behavior and performance of LOS predictions by APACHE IVa, IVb and CCOPM models for different cohorts, segmented by diagnosis, quarter/year, ICU type and ICU admission source. It primarily focuses on the period after 2014, when APACHE IVb was made available, prior to the COVID-19 outbreak in 2020. However, a brief comparison of CCOPM and APACHE IVa is also provided for the years prior to 2014, as well as a summary of observations in the period of years 2020 and 2021 (during the COVID-19 pandemic).

Results:

Compared to APACHE IVa and IVb, CCOPM – LOS demonstrates:

- Better performance in all periods with respect to predicted versus actual length of stay for all metrics considered:
  - actual-to-predicted length-of-stay ratio (A:P)
  - median absolute error (MedAE) in hours
  - coefficient of determination (R^2) of actual-to-predicted ratio, informing the proportion of variation of true LOS that is explained by predicted LOS (possible values from minus infinity to one, where the best possible value is one)
- Improved calibration in all periods assessed, presenting a stable A:P ratio that is consistently close to 1, for small to large LOS values
- Improved results in both surviving and non-surviving cohorts
- Consistently improved performance across multiple segments: ICU type, ICU admission source, admission diagnosis and quarters/years
- Better calibration after the onset of COVID-19, despite an increase in the A:P ratio driven by an increase in diagnostic categories such as respiratory infection and respiratory arrest

Metric	Period	APACHE IVa	APACHE IVb	CCOPM
A:P	2004-2013	0.83	-	**1.04**
	2014-2019	0.78	1.02	**1.00**
	2020-2021	0.87	1.13	**1.09**
MedAE (hours)	2004-2013	46	-	**27**
	2014-2019	47	34	**27**
	2020-2021	52	39	**31**
R^2	2004-2013	0.08	-	**0.27**
	2014-2019	0.03	0.13	**0.29**
	2020-2021	0.10	0.14	**0.28**

Table 1: CCOPM performance compared to APACHE IVa and IVb for a range of periods. Best values for each metric/period are indicated in bold.

Introduction

How Philips uses APACHE models

Philips provides customers with quarterly benchmarking reports of summary statistics covering various relevant ICU metrics. LOS is often used as an indicator of efficiency of care. In this context, predictions of ICU LOS can be used as a means of ICU ranking/benchmarking provided the model is risk adjusted and performs consistently across various subgroups.

Currently, such models are only available in benchmarking reporting to customers licensed to use APACHE models. These models are subject to natural drift due to sustained and generalized changes in ICU clinical practices over time and, as a result, require periodic review and recalibration. APACHE IV was developed on a cohort of patients from 2006-2008. Its latest release was based on recalibration using data from 2014-2015.

This white paper introduces Philips CCOPM – LOS as a new alternative, developed using more recent data from Philips eICU programs. It also presents a comparative performance analysis of Philips CCOPM – LOS with APACHE IVa and IVb LOS predictions.

The Philips Critical Care Outcomes Prediction – LOS model development

The development of Philips Critical Care Outcomes Prediction model for Length of Stay was based on:

- Patient data recorded from ICU admission during a 24-hour time window, used to build features of clinical relevance constructed to be used as model inputs.
- Model feature selection based on clinical relevance, with a preference for highly available features across all health systems. This was also based on feature robustness, reliability of acquisition and documentation, aiming to reduce the dependency on manual data entry.
- The patient status (alive or deceased) by the end of the stay as an outcome, for each set of inputs.
- A deep-learning technique called DeepHit[5], a modern approach to time-to-event modeling strategy, that allows for leveraging the large size and high quality of eRI dataset to produce a risk-adjusted model.
- Model architecture and features designed to be monitored and updated frequently, in order to reflect sustained changes displayed in the eICU installed base that affect ICU performances and practice.

The patient cohort for model development

Model development was based on the Philips eICU Research Institute (eRI) database comprised of de-identified physiologic, diagnostic and treatment records of all stays from participating customers. The dataset used for model training, validation and testing contained all stays in the period from January 2017 to December 2019.

Stays belonging to the same patient that presented an interval shorter than 6 hours between discharge of the previous stay and admission of the following stay were considered to be a single stay. This allows adjacent stays to be merged in a single continuous stay.

This may happen, for example, when a patient is discharged from the ICU for an operation and is subsequently readmitted to the ICU.

Inclusion criteria for an ICU stay were:

- Longer than 4 hours and shorter than 1 year
- Patient age greater than 16 years at time of admission
- ICUs used eCareManager and maintained reliable data flow
- All required model features must be non-missing.

Model features

The model features were extracted from patient characteristics on admission. Summarized statistics from vital signs and laboratory data were acquired during the first 24 hours of the ICU stay. Features can be of two types: required or non-required. Required features are those for which the model cannot produce a LOS prediction when missing. Conversely, non-required features may present missing values and the model will still generate a LOS prediction.

Required features are:

- Features acquired at ICU admission such as ICU type and admission source, patient characteristics, preadmission lead time, admission diagnosis.
- Features built from vital signs including average values and the coefficients of variation for mean, diastolic and systolic blood pressure, heart rate, oxygen saturation and respiratory rate.
- Laboratory values such as glucose, white blood cell count and hemoglobin

Non-required features are:

- Ventilation status at the end of the 24-hour window
- A flag indicating if the admission is preceded by an elective surgery.
- Glasgow Coma Score (GCS),
- Additional laboratory values such as PaCO2 and lactate.

These features are not as frequently or reliably acquired on admission to the ICU. However, in cases where they are recorded, they can be of significant predictive value.

Modeling technique

For this study, Deephit[5], a deep-learning framework for time-to-event was used with a competing risk-modeling technique. In addition to producing a risk-adjusted model, this approach accounts for competing risks.
In this case, competing risks are defined as patients surviving or non-surviving their ICU stay.

Model validation

The Philips CCOPM – LOS model was validated and tested in random samples, each composed of 20% of the eRI patient stays between January 2017 and December 2019. Details of model training, validation, and testing in eRI data will be described in a manuscript currently under peer review.

This study focuses on a broader validation expanding from eRI data into the entire eCareManager archived databases used for quarterly benchmarking.
Results are divided into three sections.

- The first and main section provides a comparison of CCOPM – LOS with APACHE IVa and IVb for the period 2014-2019, prior to COVID-19 onset and after APACHE IVb became available.
- The second section focuses on the earlier years of 2004 to 2014, when APACHE IVa was provided but APACHE IVb was not yet available.
- The third section provides a succinct outline of results from 2020 to 2021 which included the COVID-19 outbreak, highlighting changes in behaviors of the three models.

Results

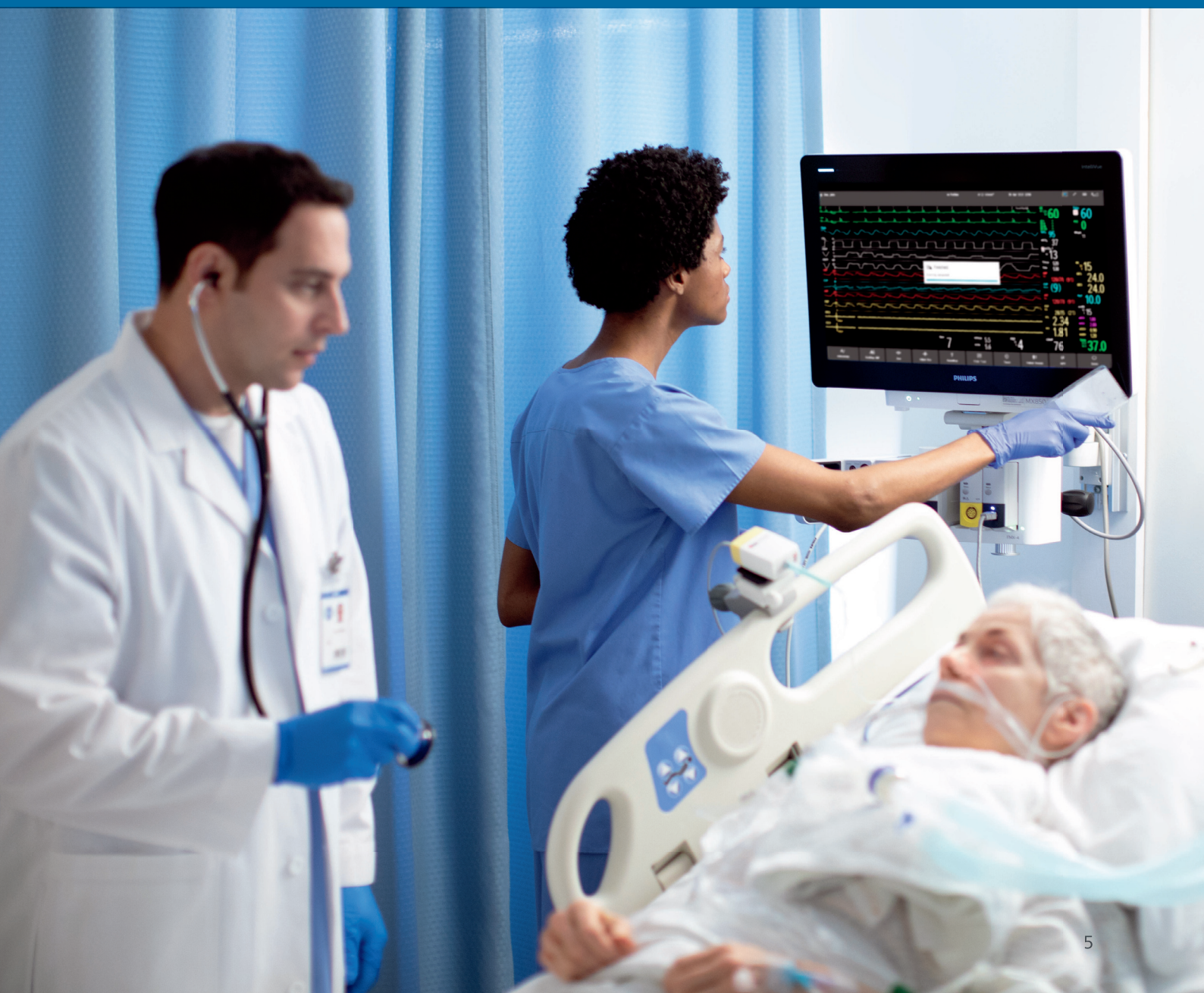
ICU stay cohorts

We identified 5,229,874 reconciled patient unit stays during the period from 2004 to 2021 from 4,151,669 different patients in 421 hospitals and 1154 ICUs.

From these, 3,457,723 stays met all the inclusion criteria for CCOPM – LOS during the study period. These comprised 3,168,476 different patients in 396 hospitals and 835 ICUs and made up the total dataset used in this study.

All results refer to either the total dataset or its sub-divisions according to APACHE predictions availability and the period analyzed:

- **Subset A:** with 1,751,382 stays, having predictions for all three models in the period from 2014 to 2019.
- **Subset B:** with 743,158 stays, having predictions for both CCOPM – LOS and APACHE IVa from 2004 to 2014.
- **Subset C:** with 581,889 stays, having predictions for all three models for years 2020 and 2021.



Model performance during the period of 2014 to 2019

CCOPM – LOS performance was compared to APACHE IVa and IVb for data subset A in the period from 2014, when APACHE IVb was made available, to 2019. The metrics used – R^2, median absolute error (MedAE) and actual-to-predicted ratios (A:P) of the averages – present distinct target values, for which a better model should have:

- An A:P ratio close to 1, meaning the average actual value is close to the average predicted value. Values of A:P higher (or lower) than 1 mean the model underestimates (or overestimates) total LOS on average.
- Lower median absolute error (MedAE), which indicates a more accurate model.
- Higher R^2 , meaning that the model variability better reflects changes in the true values.

Figure 1 represents the comparative values for the three models for each metric. CCOPM – LOS presents an A:P ratio closer to 1, lower MedAE and higher R^2 when compared to APACHE IVa and IVb in the period.

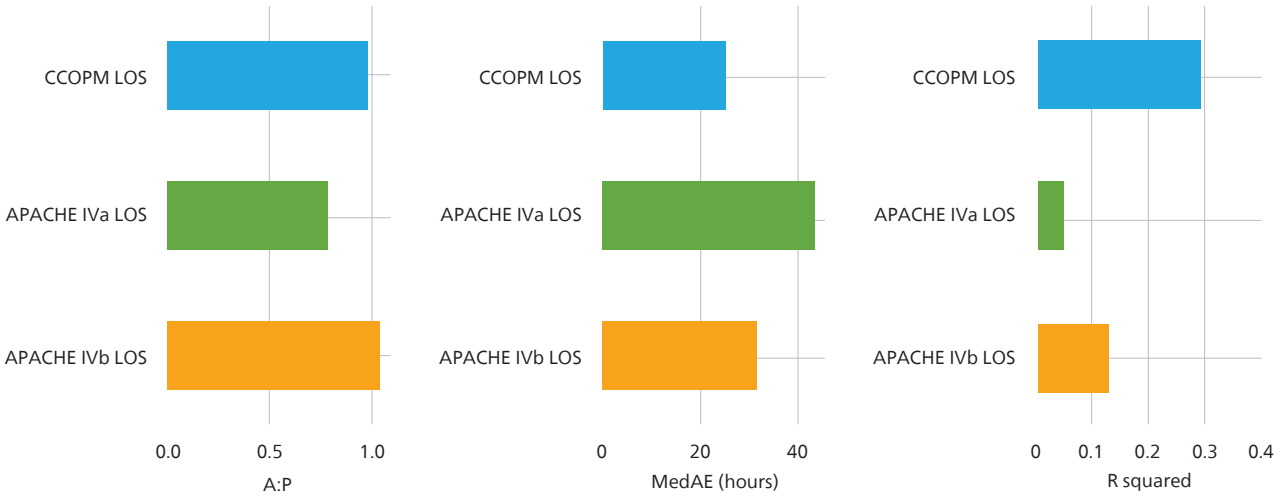


Figure 1: CCOPM – LOS performance compared to APACHE IVa and APACHE IVb length of stay from 2014 until 2019. A:P indicates the ratio of average actual LOS to average predicted LOS, for which a value close to 1 is desirable. Lower median absolute error and higher R^2 are indicative of a better model.

Calibration

Figure 2 depicts the A:P ratio for different ranges of predicted LOS for each model. For each predicted LOS interval, A:P is defined by the average of all actual LOS divided by the average of all predicted LOS that lie within that interval.

CCOPM – LOS shows a stable A:P ratio close to 1 in each interval for predicted values ranging from short LOS (<48 hours) to very long LOS (>288 hours).

APACHE IVa and IVb LOS predictions do not present a stable A:P ratio throughout the intervals, even for APACHE IVb that has an A:P ratio close to 1 overall (Table 1). For APACHE models, on average short LOS predictions underestimate the actual LOS and, conversely, long LOS predictions overestimate the actual LOS on average.

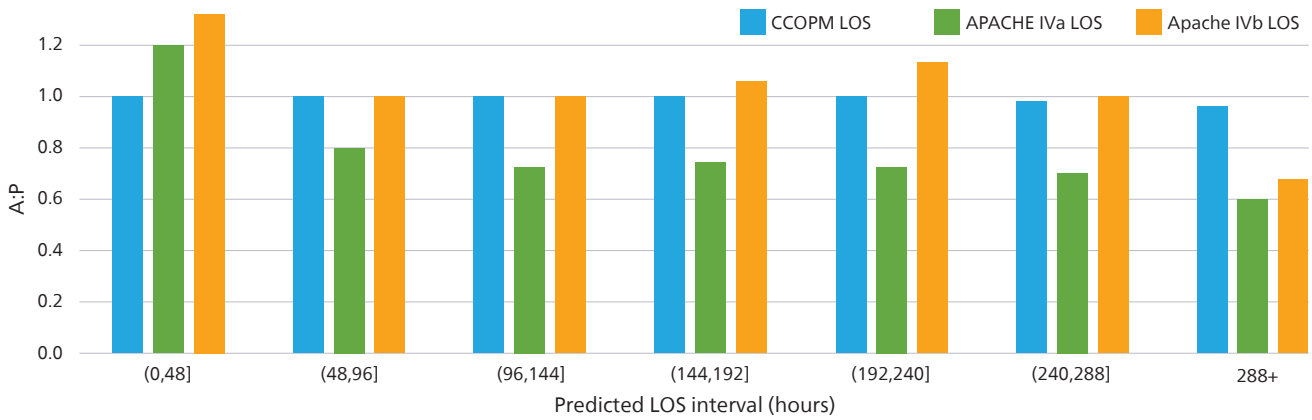


Figure 2: CCOPM comparison with APACHE IVa and IVb segmented by range of predicted LOS in intervals of 48 hours with respect to A:P ratio for years 2014 to 2019.

Subgroups

When segmented by patient exit status at the end of the ICU stay in surviving and non-surviving cohorts, APACHE IVa and IVb do not present positive R^2 values for the non-surviving cohort, while CCOPM – LOS presents a positive R^2 with a lower value than for the surviving cohort. This is shown in Figure 3.

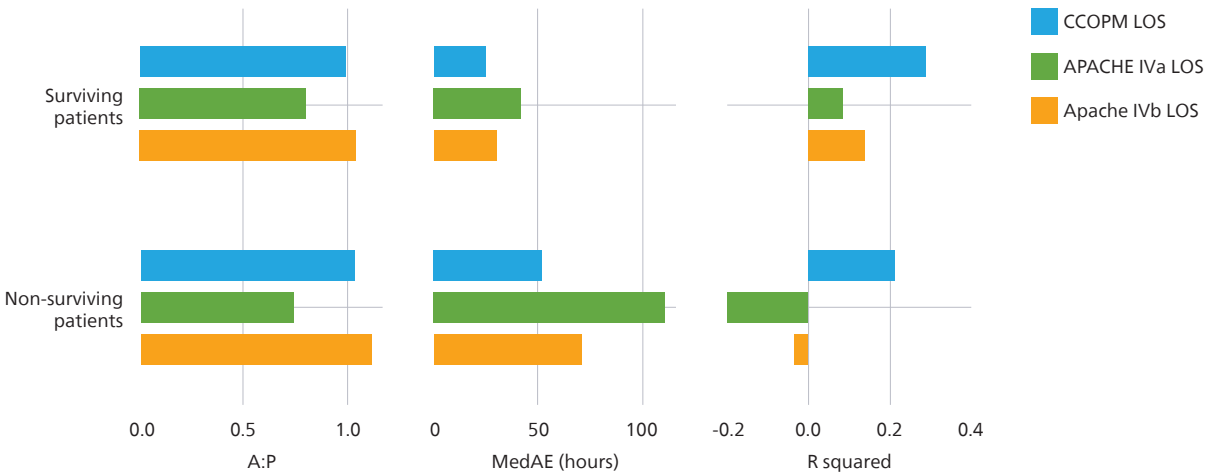


Figure 3: CCOPM – LOS performance compared to APACHE IVa and IVb with respect to surviving and non- surviving cohorts for the years 2014 to 2019.

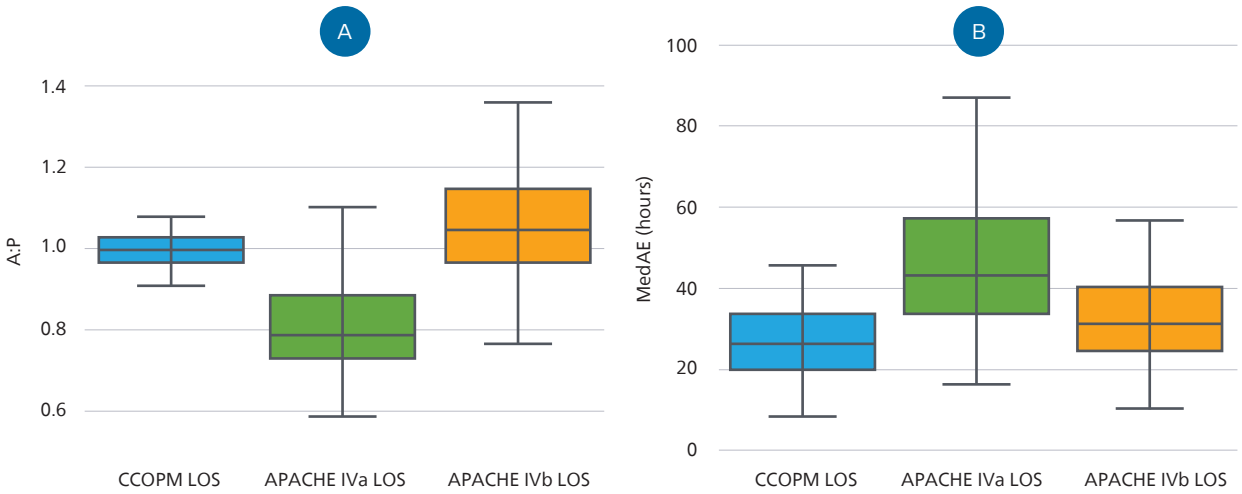


Figure 4: Model performance by diagnostic string with respect to A) actual-to-predicted LOS ratio and B) median absolute error.

Figure 4A depicts a boxplot of the distribution of A:P calculated for each of 54 admission diagnostic groups. A:P is calculated for each diagnostic string as the mean actual LOS divided by the mean predicted LOS for that diagnostic string. The boxplot represents all A:P values calculated for all diagnostic strings. When compared to APACHE IVa and IVb, CCOPM values can be found in a narrow range around the target value of 1. The fact that the box is narrow indicates that the model presents similar performance across the admission diagnostic spectrum. The fact that the median is very close to the target value of 1 indicates that the average predicted values are typically close to the actual predicted values for each diagnosis. Similarly, Figure 4B shows that CCOPM presents a lower MedAE when compared to APACHE IVa and IVb for the diagnostic subgroups.

Figure 5 shows the A:P ratio for CCOPM in a side-by-side comparison with APACHE IVb and IVa for the 10 most common admission diagnostic groups ordered in descending order from the most to the least frequent. It is worth noting that the CCOPM A:P values are close to 1 for these common diagnostic categories.

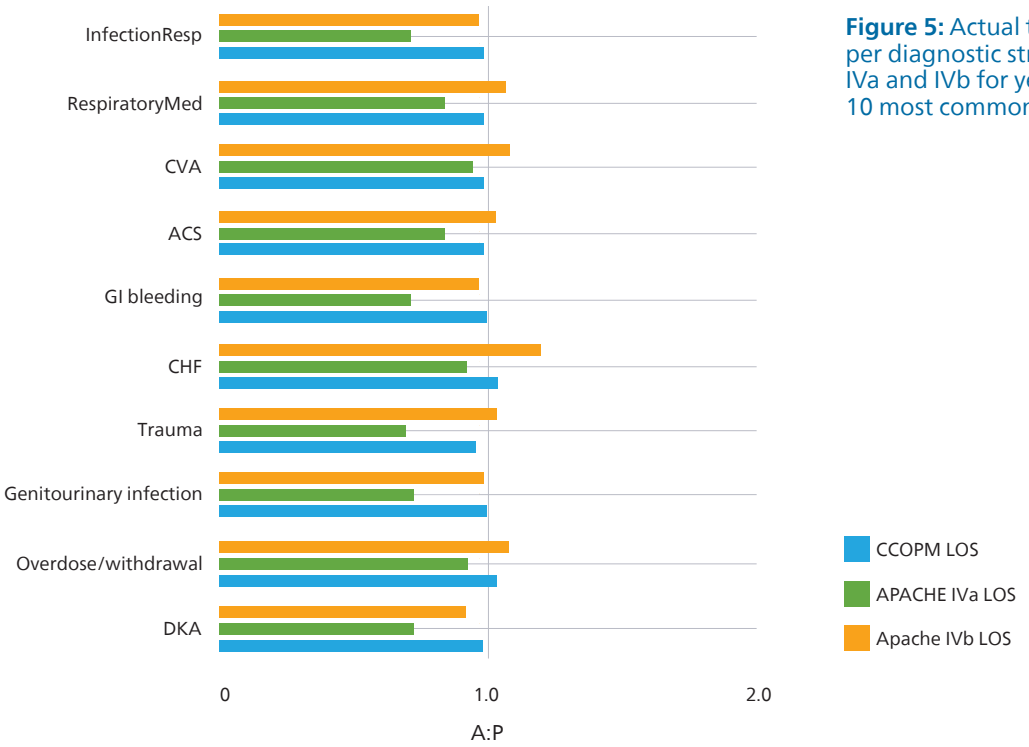


Figure 5: Actual to predicted LOS ratio per diagnostic string for CCOPM, APACHE IVa and IVb for years 2014-2019, for the 10 most common admission diagnoses.

Tables 2 and 3 show different metrics segmented by ICU type and admission source. CCOPM shows better values than APACHE IVa and IVb overall.

		A:P			MedAE			R^2		
ICU Type	**N**	**CCOPM**	**APACHE IVa**	**APACHE IVb**	**CCOPM**	**APACHE IVa**	**APACHE IVb**	**CCOPM**	**APACHE IVa**	**APACHE IVb**
CCU-CTICU	68295	1.07	0.93	1.20	26	38	29	0.27	0.06	0.08
CSICU	55473	0.97	0.80	1.01	20	34	26	0.33	0.01	0.12
CTICU	42493	1.07	1.05	1.30	26	37	30	0.28	0.08	0.07
Cardiac ICU	106543	1.00	0.88	1.15	27	39	30	0.30	0.09	0.12
MICU	232511	1.00	0.75	0.99	26	48	35	0.29	0.02	0.14
Med-Surg ICU	948509	0.98	0.75	1.00	25	44	32	0.28	0.00	0.13
Neuro ICU	96629	1.01	0.96	1.19	33	44	33	0.31	0.16	0.16
SICU	180597	0.99	0.84	1.14	29	46	33	0.30	0.10	0.13
Trauma ICU	19227	0.92	0.85	1.22	38	55	36	0.36	0.17	0.15
Vascular ICU	1105	0.95	1.18	1.48	41	39	31	0.30	0.10	0.05
CCU-CTICU	68295	1.07	0.93	1.20	26	38	29	0.27	0.06	0.08

Table 2. Model performance (actual-to-predicted ratios, median absolute error and R^2 by unit types (2014 -2019)). N indicates number of stays per ICU type within the period.

		A:P			MedAE			R^2		
Adm. Source	**N**	**CCOPM**	**APACHE IVa**	**APACHE IVb**	**CCOPM**	**APACHE IVa**	**APACHE IVb**	**CCOPM**	**APACHE IVa**	**APACHE IVb**
Direct admit	115377	1.002	0.867	1.099	27.8	40.8	31.7	0.297	0.116	0.147
ER/ observation	860461	1.000	0.740	0.979	23.6	43.0	30.7	0.296	0.017	0.146
Floor	335789	0.986	0.835	1.103	36.4	55.2	41.1	0.225	-0.020	0.054
OR	378326	1.005	0.833	1.111	21.9	36.2	26.1	0.323	0.063	0.119
Transfer	60689	0.926	0.925	1.146	50.7	60.0	49.1	0.276	0.157	0.140
Direct admit	115377	1.002	0.867	1.099	27.8	40.8	31.7	0.297	0.116	0.147

Table 3. Model performance (actual-to-predicted ratios, median absolute error and R^2 by admission source (2014 -2019)). N indicates number of stays per admission source within period.

Figure 6 depicts the performance over time of CCOPM – LOS, with little fluctuation per year/quarter from Q1 2014 to Q4 2019, indicating that the model produces comparable and consistent results from quarter to quarter. Moreover, APACHE IVa and IVb also present stable values, indicating that comparisons made between the models within the 5-year window also hold at the quarterly level.

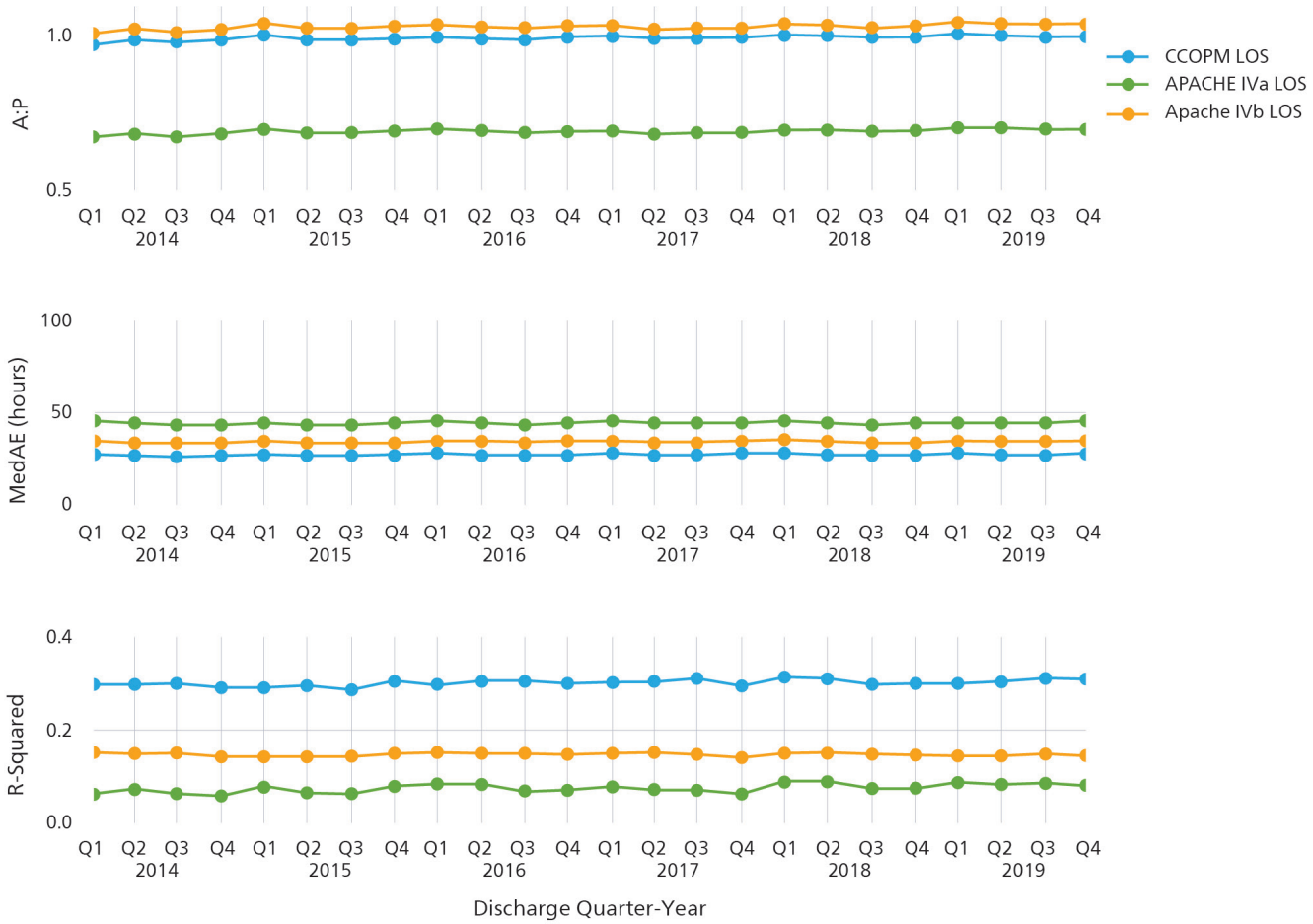


Figure 6: A:P ratio, MedAE and R^2 per quarter for CCOPM (blue), APACHE IVa (green) and IVb (orange) for years 2014-2019.

Model performance during the period of 2004-2013

Comparative values for the main performance metrics of CCOPM – LOS and APACHE IVa LOS are presented in Table 1 for this period and illustrated in Figure 7. APACHE IVb was not available for years prior to 2014.

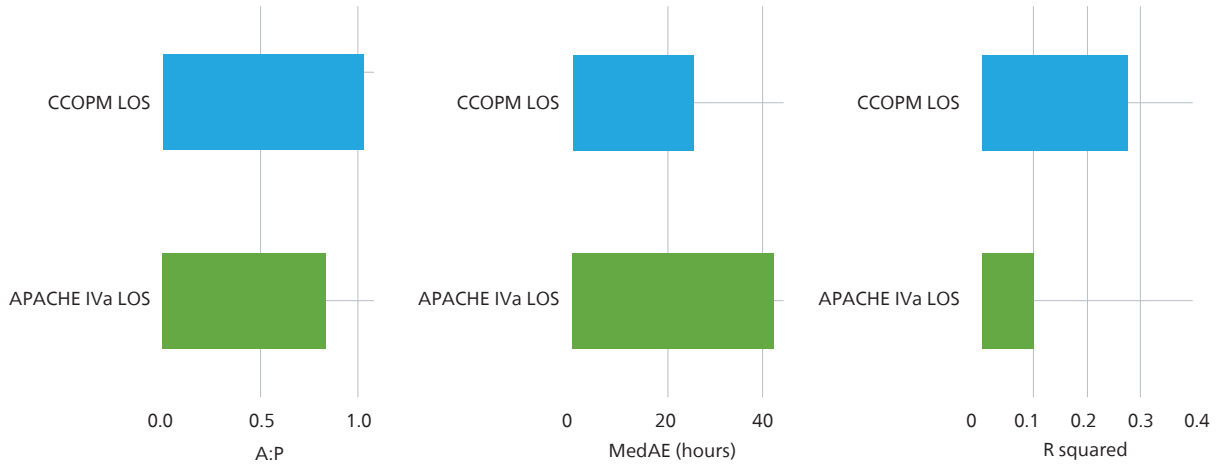


Figure 7: CCOPM – LOS performance compared to APACHE IVa length of stay from 2004 until 2013.

Even though data prior to 2017 was not used during training of the model, CCOPM – LOS presents an A:P ratio closer to 1, smaller MedAE and higher R^2 when compared to APACHE IVa for this period. Performance is comparable with that presented in the 2014-2019 period.

CCOPM has also presented good calibration (Figure 8), having stable A:P ratios for all ranges of predicted LOS in this period.

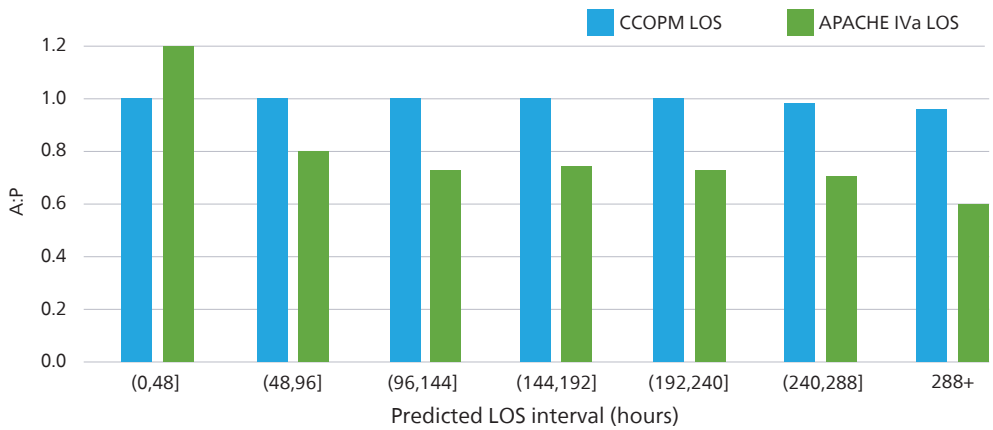


Figure 8. CCOPM comparison with APACHE IVa segmented by range of predicted LOS in intervals of 48 hours with respect to A:P ratio for years 2004 to 2013.

Model performance during the period of 2020-2021

Figure 9 presents the main performance metrics for all three models from January 2020 to end of December 2021 for subset C. This period includes data acquired after the COVID-19 pandemic hit the US (starting March 2020). The most obvious difference when compared with the previous period 2014-2019 is an increase in CCOPM – LOS A:P from 1.00 to 1.09.

When compared with the other APACHE models, CCOPM – LOS still presents A:P closer to 1, smaller MedAE and higher R^2 for this period. Regarding calibration for this period, CCOPM A:P over ranges of predicted LOS (Figure 10) is more stable throughout different ranges and closer to 1 than APACHE IVa and IVb.

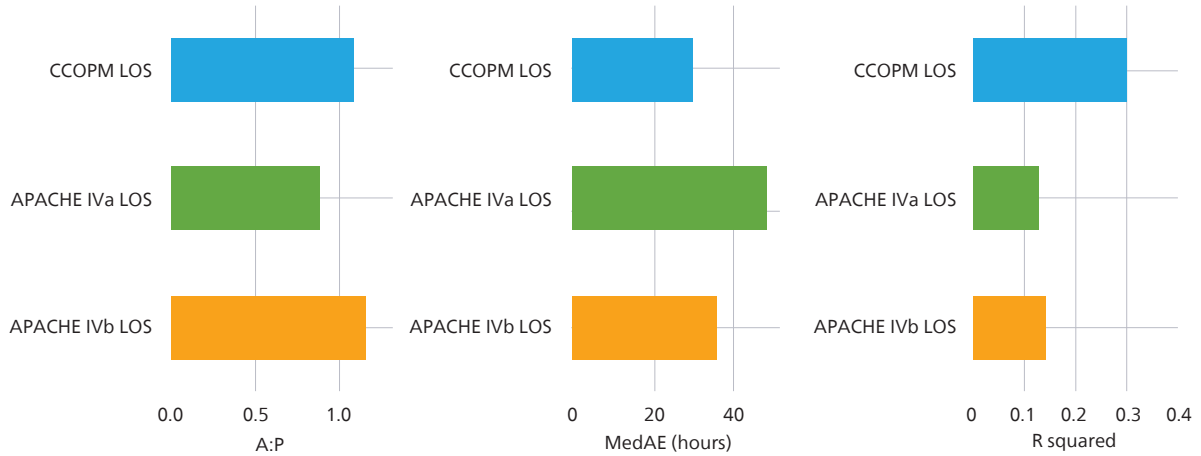


Figure 9. CCOPM – LOS performance compared to APACHE IVa and APACHE IVb length of stay predications for the years 2020 and 2021.

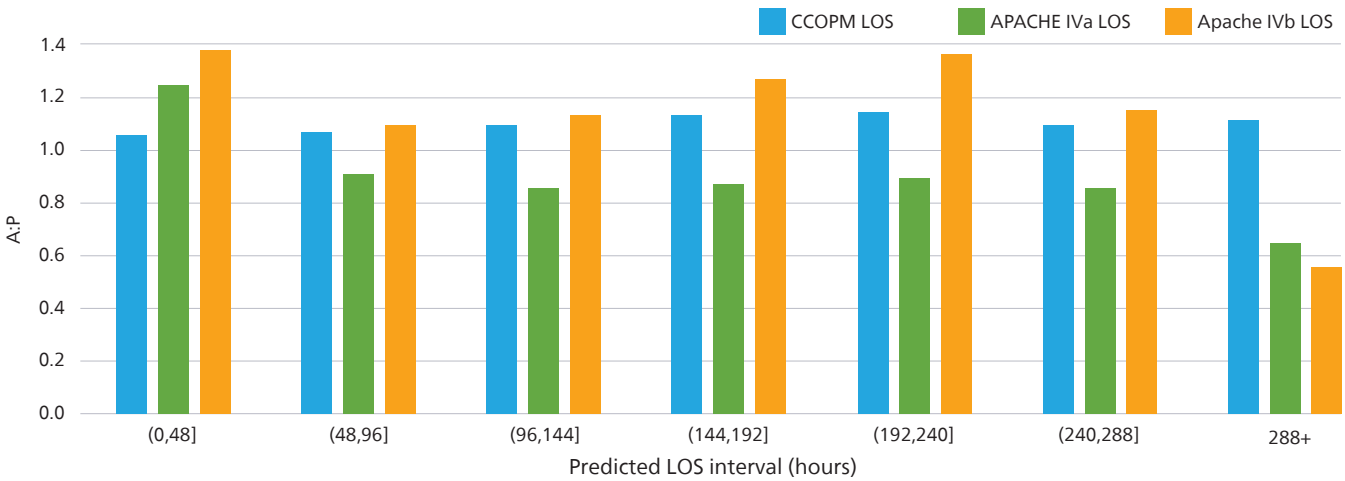


Figure 10. CCOPM comparison with APACHE IVa and IVb segmented by range of predicted LOS in intervals of 48 hours with respect to A:P ratio for the years 2020 and 2021.

Figure 11 shows a comparison of CCOPM A:P values per admission diagnosis in the period 2020-2021, (during which COVID-19 was prevalent) with the 2014-2019 period for the 10 most common admission diagnosis categories. The diagnostic category presenting the largest difference in model performance is Respiratory Infection showing 29% increase in A:P ratio. It is worth noting that although the model was not trained on COVID-19 data, its overall performance during the COVID-19 period was acceptable, especially for diagnoses that do not relate to COVID-19.

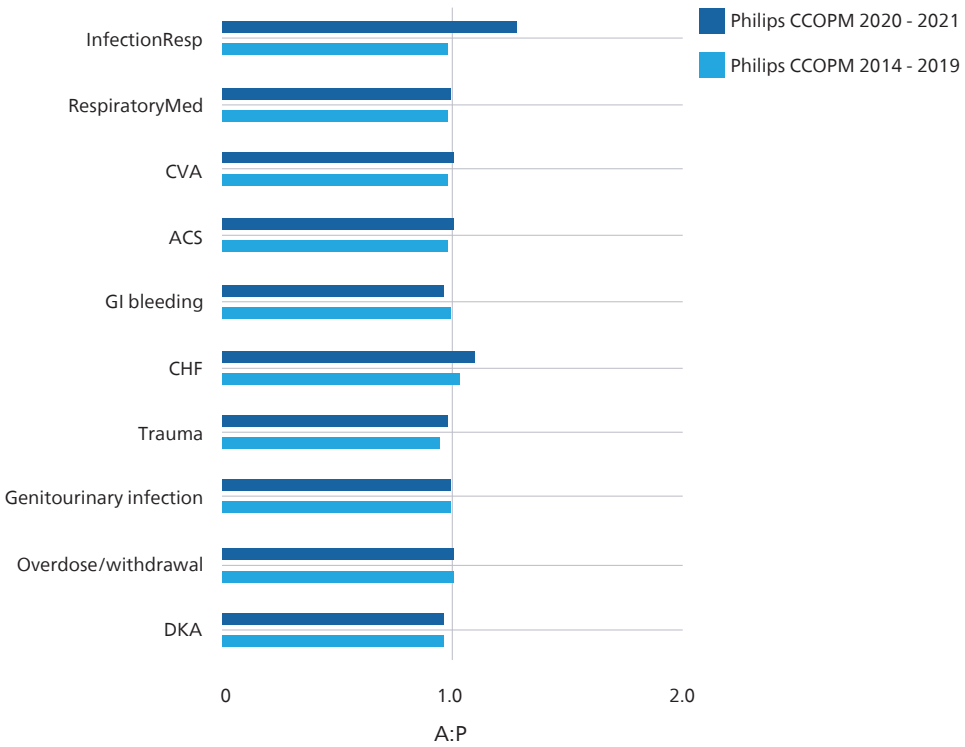


Figure 11. Actual-to-predicted LOS ratio per diagnostic string for CCOPM - LOS for the years 2014-2019 (pre-COVID-19) compared to 2020-2021 (including COVID-19 period).

Discussion

Effective benchmarking of hospitals and ICUs depends on risk-adjusted indicators of quality and efficacy of care. In this context, risk-adjusted models that support benchmarking must be periodically updated in order to keep up with changes in healthcare trends and practices.

In this white paper, we present the novel CCOPM – LOS model that presented better performance than APACHE IVa and IVb for all periods analyzed. The difference in performance might be attributed to the following factors:

- The new model presents a reduced documentation burden when compared to APACHE since manual entries such as comorbidities, active treatments, and urinary output are not required.
- The features were tailored to be more resilient to documentation bias. Changes in data collection methods from manual to automated have provoked categorical data drift. This is treated in the data curation phase, prior to model input. Data collection, monitoring and bias mitigation are discussed in a previous white paper[6].
- The use of a heterogeneous and large training/validation/testing dataset, for a large number of institutions, hospitals and ICUs combined with a deep-learning approach.

As a result, the CCOPM – LOS model was risk-adjusted, allowing patient evolution during the first 24 hours to be associated with different risks according to admission diagnosis. The model showed improved performance compared to APACHE IVa and IVb in all subgroups analyzed, especially the non-surviving cohort for which APACHE models failed to present positive R^2. These differences can be attributed to the deep-learning-based model for competing risks.

Although the model should be recalibrated and revalidated once there is enough data after novel protocols become steady practices, CCOPM – LOS has presented better performance when compared to APACHE IVa and IVb after 2020 even though data from this period was not presented to the model during training.

Conclusion and future steps

The CCOPM – LOS model has presented improved performance and calibration over APACHE IVa and IVb in all periods and subgroups analyzed. Future steps include active monitoring of data distribution drift, further expansion of inclusion criteria and model retraining post COVID-19 outbreak.

References

1 Zimmerman, Jack E., et al. "Acute Physiology and Chronic Health Evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients." Critical care medicine 34.5 (2006): 1297-1310.
2 Zimmerman, Jack E., et al. "Intensive care unit length of stay: Benchmarking based on Acute Physiology and Chronic Health Evaluation (APACHE) IV." Critical care medicine 34.10 (2006): 2517-2529.
3 Cerner. APACHE IVb White Paper Report. Cerner Corporation; 2016.
4 Liu, Xinggang, and Omar Badawi. "369: ICU Length-of-Stay models should account for the interaction between survival and patient severity." Critical Care Medicine 48.1 (2020): 166.
5 Lee, C., Zame, W., Yoon, J. and Van Der Schaar, M., 2018, April. Deephit: A deep learning approach to survival analysis with competing risks. In Proceedings of the AAAI conference on artificial intelligence (Vol. 32, No. 1).
6 Philips. Introduction to the Philips Critical Care Outcomes Prediction models – Mortality prediction; 2020

www.philips.com

00000044-00-00 * FEB 2023