

External validation of SAPS II score reported to the Norwegian Intensive Care and Pandemic Registry (NIPaR)

Eirik Alnes Buanes^{1,2}, Øyvind Bruserud^{1,2}, Sergio Carracedo Huroz³, Andreas Barratt-Due⁴, Hans Flaatten^{1,5} and Reidar Kvåle^{1,2,5}

1) Department of Anaesthesia and Intensive Care, Haukeland University Hospital, Bergen, Norway

2) Norwegian Intensive Care and Pandemic Registry, Haukeland University Hospital, Bergen, Norway

3) Fagsenter for medisinske kvalitetsregister, Seksjon for forskning og innovasjon, Forskings- og utviklingsavdelinga, Haukeland universitetssjukehus, Helse Bergen HF, 5021 Bergen, Norway

4) Department of Anesthesia and Intensive Care Medicine, Rikshospitalet, Oslo University Hospital, Oslo, Norway

5) Department of Clinical Medicine, University of Bergen, Norway

Correspondence: Eirik Alnes Buanes e-mail eirik.alnes.buanes@helse-bergen.no telephone: +47 55975000

ABSTRACT

Background: Simplified Acute Physiology Score II (SAPS II) is a mortality prediction model widely used to compensate for differences between intensive care units (ICU) in benchmarking and research. Accuracy of SAPS II is sparsely documented. We investigate accuracy by comparing patient journal SAPS II values with registry SAPS II values in the Norwegian Intensive Care and Pandemic Registry (NIPaR).

Method: NIPaR personnel collected data from the patient journal during visitations to ICUs in ten different hospitals between 2017 and 2022 while blinded for registry SAPS II data. The patient journal SAPS II values were subsequently compared with the registry SAPS II values.

Results: Difference of means for SAPS II score between patient journal and registry data was 5.2 points (95% CI 2.8–7.6; $p < 0.001$). SAPS II score depended significantly on ICU ($p < 0.001$) and data origin ($p = 0.006$), whereas the interaction term for these two variables was not significant.

Conclusion: We find low accuracy of SAPS II score in a registry setting.

This is an open access article distributed under the Creative Commons Attribution Licence, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

BACKGROUND

One of the main characteristics of intensive care medicine is the large variation between intensive care units (ICU) in terms of organization, diagnostic capabilities, and therapeutic options. As a result, ICUs differ regarding patient groups, severity of illness and patient case mix in terms of age, gender, and comorbidity. Mortality prediction models compensate for some of these differences by weighting risk factors of mortality in a standardized way. This allows for comparison of ICU performance despite differences.

Simplified Acute Physiology Score II (SAPS II) is a mortality prediction model widely used in both intensive care research and intensive care registries (1). There is good documentation for the ability of SAPS II in predicting mortality (2-5). However, there are few publications on the accuracy of SAPS II scoring in a registry setting (6,7). If registry data are inaccurate, the model may not perform optimally in a registry setting despite performing well during development.

In this study we investigate the accuracy of SAPS II scoring by comparing data in the patient journal with SAPS II values in the Norwegian Intensive Care and Pandemic Registry (NIPaR).

MATERIAL AND METHODS

NIPaR (see description at the end of this article) is a national quality registry located in Norway, population

5.39 million (2021) (8). Reporting is mandatory and more than 90% of ICU admissions are included in the registry (9). Trained personnel in each ICU report data to NIPaR based on a detailed description issued by the registry. Additional SAPS II scoring tools have been distributed to registrars and made available on the registry website (10). SAPS II values are reported categorically for patients above 18 years on admission. Bilirubin, bicarbonate, and urea values are voluntary, as they are not routinely measured in all ICUs. Registrars report NIPaR data manually in a web-based case report form (eCRF) or by uploading a file. There is automatic data sampling of a few values in some ICUs, these are available for correction by the registrar before reporting to NIPaR.

To investigate accuracy, NIPaR personnel have collected data from the patient journal during visitations to ICUs in ten different hospitals between 2017 and 2022. The chosen ICUs were of different sizes and from all health regions in Norway. In each of the ten ICUs we selected up to 20 random admissions previously reported to NIPaR for validation. With the assistance of local registrars, a patient journal SAPS II score for each ICU stay was determined based on data from patient records while blinded for the previously reported registry SAPS II values. In seven of the ten ICUs the values of individual SAPS II variables were documented as well as the total SAPS II score.

The patient journal SAPS II values were subsequently compared with the registry SAPS II values. Having

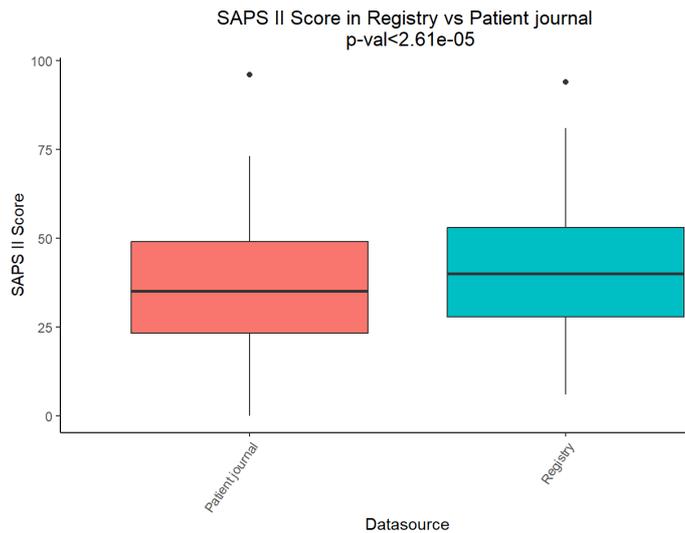


Figure 1. SAPS II score values by data source. Median, inter-quartile range and spread. p-value for paired t-test.

over 150 SAPS II Score values per group and upon confirming homoscedasticity between both groups of samples with a Bartlett test, we compared SAPS II means by paired t-test with significance level $\alpha = 0.05$. All calculations and graphics were performed in R (version 4.2.2 (2022-10-31 Universal C Runtime)). As missing values are not allowed for mandatory variables in the eCRF, value “0” may represent missing data in the patient journal as per scoring instructions. Also, missing values are allowed when uploading files. Values “0” and “-1” are therefore considered a match in all combinations. For practical reasons during validation, variable “age” was only validated in ICU 4,6 and 8, and variable “MvOrCpap” (PaO₂/FiO₂ ratio) was not validated in ICU 1 and 5.

RESULTS

Difference of means for SAPS II score between patient journal and registry data was 5.2 points (95% CI 2.8–7.6; $p < 0.001$) (Figure 1). Mean SAPS II score was 41.3 (95% CI 38.3–44.2) in the registry sample and 36.0 (95% CI 33.3–38.7) in patient journal. Two-way ANOVA analysis was used to compare dependency of SAPS II on ICU and/or data origin (registry vs patient journal) with and without their corresponding interaction term. SAPS II score depended significantly on ICU ($p < 0.001$) and data origin ($p = 0.006$), whereas the interaction term for these two variables was not significant. The trend towards higher registry values was found in most ICUs, except one where patient journal values were higher and one where patient journal values were equal to registry values (Figure 2). The difference in mean values is mainly due to overestimation of SAPS II values below 40 in the registry (Figure 3). Although individual SAPS II variables have a high level of accuracy in most ICUs, SAPS II score based on all 15 variables has low accuracy. We found no apparent pattern regarding which individual variables or types of variables which might explain the discrepancy (Figures 4 and 5). In a Q-Q plot, the SAPS II patient journal sample and the SAPS II registry sample appear to have similar distributions with a stable separation (Figure 6). ICU 9 and 10 had the largest differences between patient journal and registry values. A subgroup analysis excluding ICU 9 and 10 still revealed no significant interaction term for ICU and data origin in two-way ANOVA analysis, while SAPS II score depended significantly on ICU ($p = 0.02$) but not on data origin ($p = 0.16$). The difference between registry and patient journal means in the subgroup was 3.0 points (95% CI 0.2–5.8; $p = 0.03$).

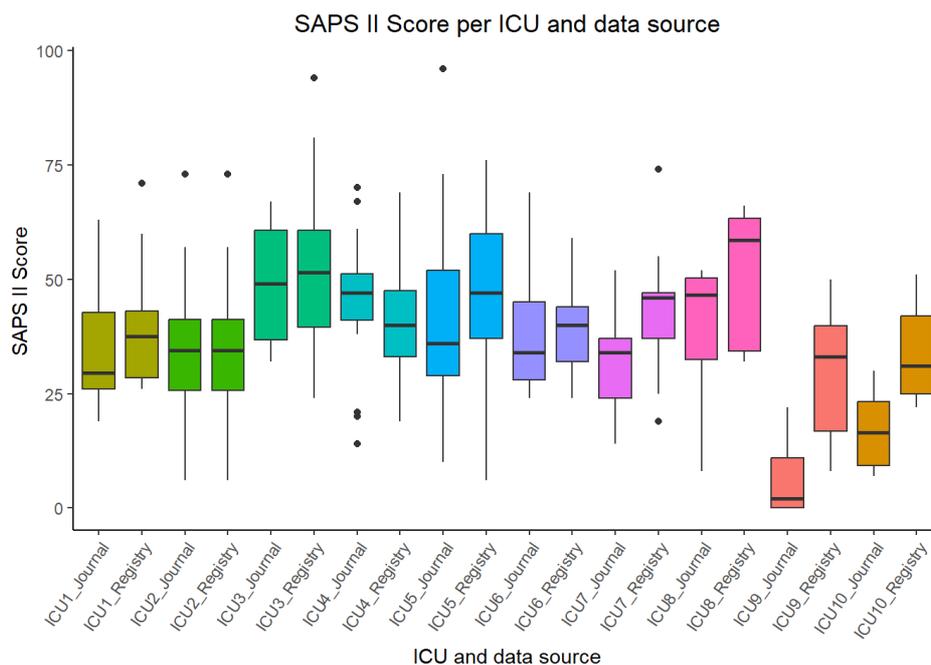


Figure 2. SAPS II score values by hospital and data source.

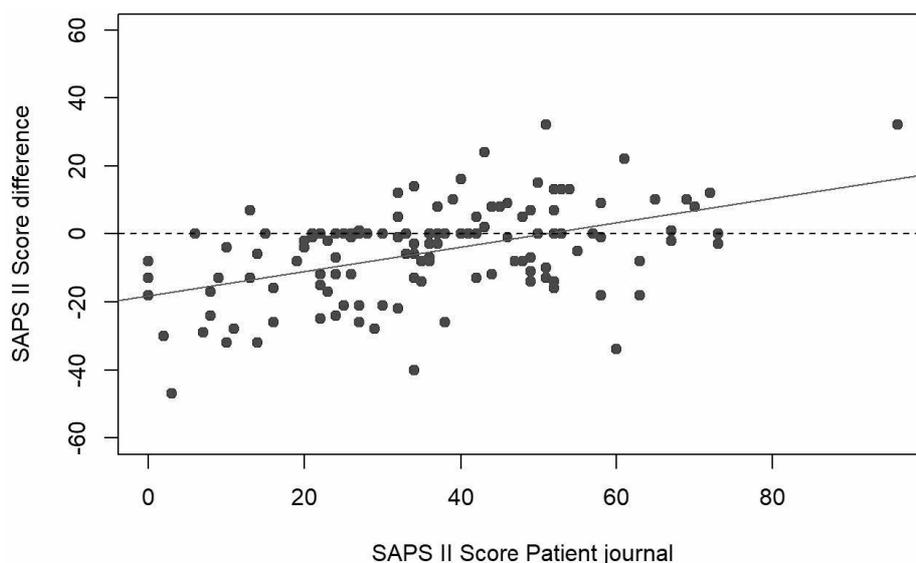


Figure 3. SAPS II score difference between patient journal and registry data by SAPS II value. Dots represent individual ICU stays.

DISCUSSION

Accuracy was found to be low for overall SAPS II scores in a national quality registry setting, with a mean SAPS II score 5.2 points higher in the registry compared to patient journal. The confidence interval around the mean difference is wide, suggesting variation in the magnitude of difference between patient journal and registry values.

Mortality prediction models gained momentum in intensive care medicine in the 1990s (11). During the next few decades, several studies were published with increasingly large study populations and more variables, improving predictions of mortality on group level (1,12,13). While accuracy of data is regularly investigated in medicine trials, few studies report this dimension of data quality in registry studies on mortality prediction (6,7). This is interesting, since good accuracy is a prerequisite in both the development and use of any mortality prediction model. Moreover, while the precision of a mortality prediction model will increase with the number of variables up to a certain point, accuracy of reported data will likely fall as the number of variables and their complexity increase. This may not substantially affect research projects, where data collection is diligent during a limited period. However, in a registry setting, reporting is continuous over several years involving a high number of registrars. This puts extra demands on variables in quality registries, in that data values need to be easily obtainable and intuitive to the ones who report them. Based on our results, one could speculate that the SAPS II score is too complex for registry use, while better suited for research. NIPaR has taken actions to improve education, continue validation and execute caution when interpreting SAPS II score and standardized mortality ratios based on SAPS II.

While low accuracy in the SAPS II score was expected, the magnitude of the problem was larger than

presumed. A mean SAPS II score five points higher in the registry than in the patient journal will increase the predicted mortality by approximately seven percent. If this is representative for the registry, inaccuracy in reporting could account for a proportion of the previously reported difference between observed and predicted mortality (4).

We could not make out any discernable pattern of variables associated with the inaccuracy. One would expect blood values to have a high level of accuracy throughout, but this does not seem to be the case. We have seen that reporting values outside of the first 24 hours in the ICU, as per the SAPS II definitions, is often the cause of erroneous scoring. We find that low accuracy is a concern throughout the range of SAPS II. However, in ICU stays where SAPS II score is below 40 the discrepancy is largely negative, whereas for SAPS II scores above 40 the discrepancy is both positive and negative. In effect, ICU stays with SAPS II score above 40 will have a similar mean in the registry as compared to the patient journal, whereas ICU stays with a SAPS II score below 40 have a significantly higher mean in the registry compared to the patient journal. Our data suggest that accuracy could be linearly dependent on SAPS II score in the patient journal, but we lack data with high SAPS II values to substantiate the relation (Figure 3). We find no obvious explanation for this pattern.

ICU 9 and 10 seem to have a larger difference between means than other ICUs (Figure 2). Unfortunately, we were not able to investigate this further as individual variables that make up the total score were unavailable. In a subgroup analysis excluding ICU 9 and 10 the difference in means between patient journal and registry decreased to 3 points but was still significant.

Due to legal issues, we were not able to establish patient journal values by two independent evaluators. As a result, one could argue that the present study does

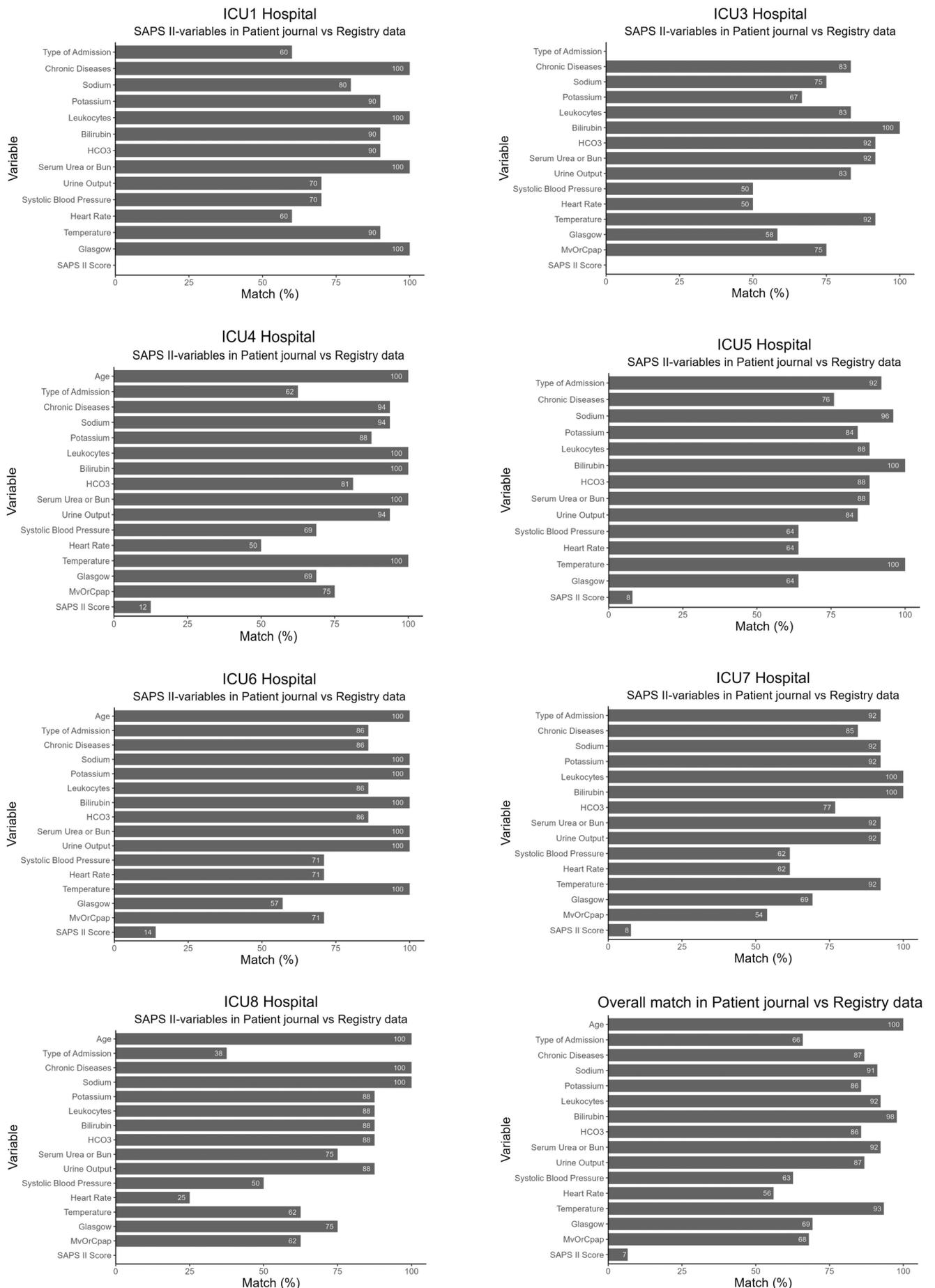


Figure 4. Match between patient journal and registry data by variable and ICU.

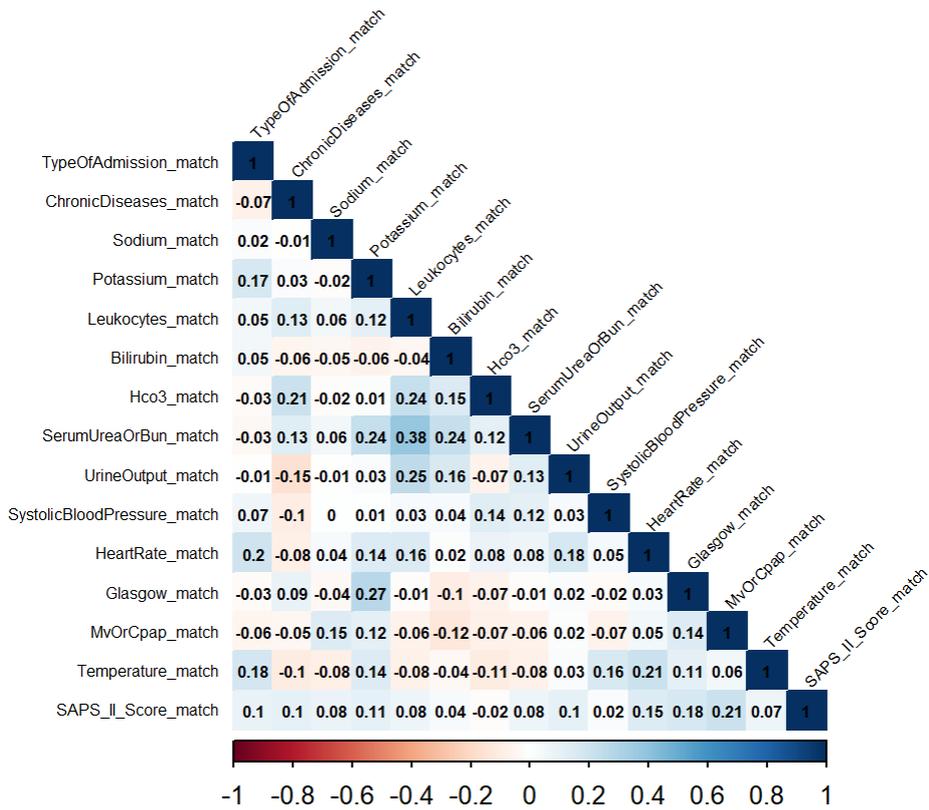


Figure 5. Correlation among variable matches in patient journal versus registry data.

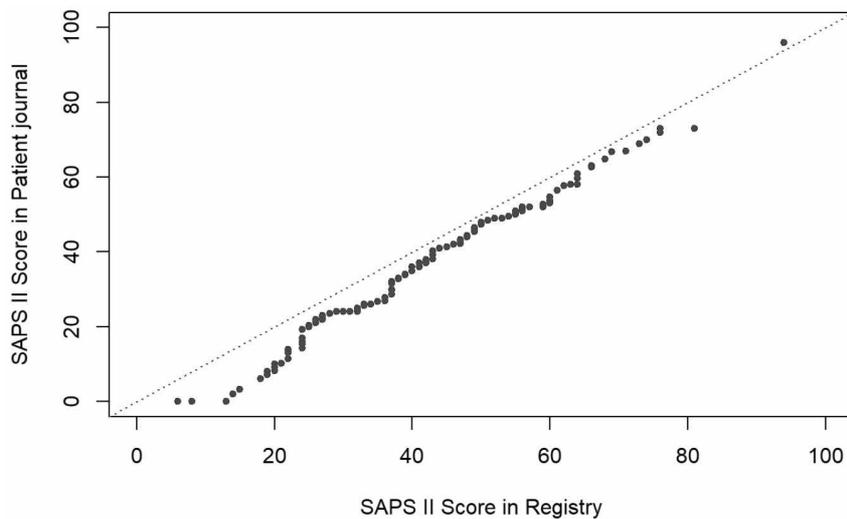


Figure 6. Q-Q plot for SAPS II score using data from all ICUs.

not evaluate accuracy, rather the precision of repeated measurements. In our opinion, the combined evaluation of patient documentation by local and registry personnel is sufficiently similar to independent evaluation for a reasonable estimate of gold standard, and we therefore refer to “accuracy” in this manuscript.

Strengths and limitations

The study investigates accuracy of SAPS II registry data in a mandatory ICU registry on a national level. The data have been randomly sampled from several ICUs across Norway. We lack explanatory variables regard-

ing registration methods (eCRF or upload) and data extraction (manual or automatic). The study sample is small due to the need for visiting individual ICUs to manually sample data. Consequently, studies on the accuracy of SAPS II score in other registry settings are needed. Due to legal issues, we were prohibited from establishing a patient journal gold standard by two independent investigators.

CONCLUSION

We find low accuracy of SAPS II score in a registry setting.

DESCRIPTION OF NIPaR

The Norwegian Intensive Care and Pandemic Registry (NIPaR) was established as an intensive care registry (Norwegian Intensive Care Registry – NIR) in 1998 and has collected individual level data on intensive care stays since 2011. As of March 2023 the current electronic database includes more than 150.000 intensive care stays since 2014. Data include reasons for ICU treatment, risk factors, findings, treatment, diagnoses, and outcomes including patient reported outcome measures (PROM). In 2020, the registry was expanded to include data from all hospital admissions with a positive PCR for Covid-19 in a separate technical installation termed Norwegian Pandemic Registry (NoPaR). The database includes more than 31.000 admissions as of March 2023. Data include risk factors, main cause for admission, disease severity, findings on admission, treatment received, and outcome. NoPaR patients receive questionnaires from NIPaR at 3, 6, 12 and 24 months after admission to hospital. Each questionnaire includes several PROM, and answers are included in the registry. NIPaR is a medical quality registry with national status from the Norwegian Directorate of Health.

REFERENCES

1. Le Gall JR, Lemeshow S, Saulnier F. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993;**270**(24):2957-63.
2. Arabi Y, Al Shirawi N, Memish Z, Venkatesh S, Al-Shimemeri A. Assessment of six mortality prediction models in patients admitted with severe sepsis and septic shock to the intensive care unit: a prospective cohort study. *Crit Care* 2003;**7**(5):R116-22.
3. Czajka S, Ziębińska K, Marczenko K, Posmyk B, Szczepańska AJ, Krzych ŁJ. Validation of APACHE II, APACHE III and SAPS II scores in in-hospital and one year mortality prediction in a mixed intensive care unit in Poland: a cohort study. *BMC Anesthesiol* 2020;**20**(1):296.
4. Haaland OA, Lindemark F, Flaatten H, Kvale R, Johansson KA. A calibration study of SAPS II with Norwegian intensive care registry data. *Acta Anaesthesiol Scand* 2014;**58**(6):701-8.
5. Le Gall JR, Neumann A, Hemery F, Bleriot JP, Fulgencio JP, Garrigues B, et al. Mortality prediction using SAPS II: an update for French intensive care units. *Crit Care* 2005;**9**(6):R645-52.
6. Arts D, de Keizer N, Scheffer GJ, de Jonge E. Quality of data collected for severity of illness scores in the Dutch National Intensive Care Evaluation (NICE) registry. *Intensive Care Med* 2002;**28**(5):656-9.
7. Arts DG, Bosman RJ, de Jonge E, Joore JC, de Keizer NF. Training in data definitions improves quality of intensive care data. *Crit Care* 2003;**7**(2):179-84.
8. Fakta om befolkningen – hvor mange bor det i Norge? Statistics Norway, 2021. Available from: <https://www.ssb.no/befolkning/faktaside/befolkningen>.
9. Buanes EA. Norsk intensiv- og pandemiregister. Årsrapport for 2021 med plan for forbedringstiltak, 2022.
10. Description of registration and scoring tools in NIPaR [Available from: <https://helse-bergen.no/norsk-intensivregister-nir/mal-for-data-og-kodebok>].
11. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. *Crit Care Med* 1985;**13**(10):818-29.
12. Bastos PG, Knaus WA. APACHE III study: a summary. *Intensive Care World* 1991;**8**(1):35-8.
13. Moreno RP, Metnitz PG, Almeida E, Jordan B, Bauer P, Campos RA, et al. SAPS 3 – From evaluation of the patient to evaluation of the intensive care unit. Part 2: Development of a prognostic model for hospital mortality at ICU admission. *Intensive Care Med* 2005;**31**(10):1345-55.