

## The Inter-Rater Reliability of Simplified Acute Physiology Score 3 (SAPS3) among Intensive Care Unit Nurses

Jun Hyun Kim, M.D., Ji Yeon Kim, M.D., Ph.D., Wonil Kim, M.D., Kyung Woo Kim, M.D., Sang-il Lee, M.D., Kyung-Tae Kim, M.D., Ph.D., Jang Su Park, M.D., Ph.D., Won Joo Choe, M.D., Ph.D., and Jung Won Kim, M.D., Ph.D.

Department of Anesthesiology and Pain Medicine, Ilsan Paik Hospital, Goyang, Korea

**Background:** Simplified acute physiology score 3 (SAPS3) was developed in 2005 to evaluate intensive care unit (ICU) performance and to predict patient mortality or disease severity. The score is usually calculated by doctors, but it requires substantial human resources. And many nurse-lead studies use this scoring system. In the present study, we examined the inter-rater reliability of SAPS3 among nurses in an ICU.

**Methods:** Five ICU nurses who worked in an ICU for a mean length of 7.8 years were educated for 2 hours about SAPS3 score and its components. Each nurse scored 26 patients, and the intraclass correlation coefficient (ICC) of the total scores and each subset were evaluated.

**Results:** The ICC (95% confidence interval) of SAPS3 score was 0.89 (0.82-0.95), that of subset I was 0.90 (0.82-0.95), subset II was 0.54 (0.35-0.73), and subset III was 0.95 (0.91-0.97). The ICC of predicted mortality was 0.91 (0.85-0.96).

**Conclusions:** The ICC of SAPS3 score and predicted mortality among ICU nurses were reliable. According to these ICC values, SAPS3 score is a reliable scale to be used by nurses. The ICC of subset II was lower than those of the other subsets, suggesting that education of SAPS3 should focus on the definition of each subset II component.

**Key Words:** critical care; intensive care units; observer variation; severity of illness index.

### Introduction

The severity of illness in critically ill patients is measured using a variety of scoring systems, including acute physiologic and chronic health evaluation (APACHE) Score, mortality prediction model (MPM) and simplified acute physiology score (SAPS) to predict clinical outcome. In addition to mortality prediction power, these severity scoring systems are also used to evaluate the quality of critical care based on the comparison of predicted mortality to actual mortality. Also, the effectiveness of care bundle practiced in the intensive care unit (ICU) setting is also assessed by comparing mortality predicted after application of bundles against actual mortality.[1,2] Among severity scoring systems, simplified acute physiology score 3 (SAPS3) was developed in 2005 to upgrade the then existing scoring systems. SAPS3 is used to determine disease severity on ICU

admission using 20 clinical variables.[3,4] Although severity scoring systems have been improved, no standards have been established for assessment methods and severity ratings. More importantly, the available scoring systems tend to be too complicated and time consuming. In fact, even in US, scoring systems were used only 10-15% of ICUs.[1]

One of the reasons that severity scoring systems are not

Received on September 25, 2014 Revised on November 3, 2014

Accepted on November 24, 2014

**Correspondence to:** Ji Yeon Kim, Department of Anesthesiology and Pain Medicine, Ilsan Paik Hospital, 170, Juhwa-ro, IlsanSeo-gu, Goyang 411-706, Korea

Tel: +82-31-910-7187, Fax: +82-31-910-7184

E-mail: jy67925@naver.com

\*No potential conflict of interest relevant to this article was reported.

© This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2015 The Korean Society of Critical Care Medicine

commonly used in the ICU can be the need for human resources and time. Some researchers describe SAPS3 scoring is less complicated and time consuming than other scoring systems because it is based on smaller number of variables. [5] However, SAPS3 still takes significant time for inexperienced users to complete. To improve the validity of the predicted mortality, variables should be measured within one hour after ICU admission. Any delay in data collection can overestimate the mortality.[4] Given the importance of timing, ICU nurses can assess severity of illness if ICU physicians are not in a position to perform it in the timely manner. Recently, there are many studies led by nurses, and in some studies, SAPS3 were used as a severity score. We therefore explored in this study if there were any difference in the inter-rater reliability (IRR) of SAPS3 scores assessed by ICU nurses.

Sum score and predicted mortality of one patient can vary depending on raters. Strand K[6] investigated IRR of simplified acute physiology score 2 (SAPS2) and SAPS3 performed by doctors. We evaluated the IRR of SAPS3 performed by ICU nurses using intraclass correlation coefficient (ICC) and compared the results with the previously reported IRR of doctor-assessed SAPS3 score. As a secondary aim of this study, we evaluated the ICC of SAPS3 subscores or boxes to find out what variables cause raters to agree and disagree.

## Materials and Methods

This study was exempted the review of the Medical Research Ethics Committee of the Inje University Ilsan Paik Hospital (IRB No. IB-2-1310-042).

### 1) Methods

A total of five nurses working in the surgical ICU at Ilsan Paik Hospital were trained in the use of SAPS3 scoring for 2 hours. After completion of the training, nurses assessed severity of illness in anonymous 26 patients who admitted to the ICU before the study period. Demographic data of patients are presented in Table 1. The ICC was then calculated by comparing severity scores given to 26 patients by each nurse. The mean ICU working year of nurses was 7.8 years (3-11 years, SD = 3.1). Patients aged 16 or below were excluded.

**Table 1.** Patient demographics

Variables	Values
Age (mean ± SD)	62 ± 17
Gender (male : female)	16 : 10
Admission route (ER : ward : other ICU)	19 : 7 : 0
Reason for admission	
Hemato-oncology	1
Respiratory	3
Gastrointestinal	2
Cardiovascular	3
Others	17
ICU length of stay (days), (median, 25-75% percentiles)	4 (3-10)
Observed hospital mortality	6 (23%)
SAPS3 score (mean ± SD)	61 ± 23
Predicted mortality rate (% , mean ± SD)	35 ± 30

SD: standard deviation; ER: emergency room; ICU: intensive care unit; SAPS3: simplified acute physiology score 3.

### 2) Statistical analysis and sample size

We calculated ICC to determine IRR of collected SAPS3 sum score and subscores and probability of death using the MedCalc software package (MedCalc Software, Ostend, Belgium), version 12.4 for Windows<sup>®</sup>. The probability of in-hospital death for SAPS3 can be calculated by specific customized equations.[4] The required sample size to tell the difference between ICC of 0.8 and 0.9 by five raters was 26 based on the alpha level of 0.05 and the beta level of 0.2.[7] A total of 26 patients were therefore recruited in this study. IRR is measured by the ICC with 0 or less meaning 'no match at all' and 1 meaning 'perfect match.' The ICC value of 0.8 or above is considered preferable.[8]

## Results

When five nurses independently assessed SAPS3 scores in 26 patients, ICC of SAPS3 score was 0.89 (95% confidence interval (CI) was 0.82 to 0.95) and ICC for the mean predicted mortality was 0.91(0.85 to 0.96). Among SAPS3 subscores analyzed in this study, ICC of SAPS3 box I, II and III was 0.90 (0.82 to 0.95), 0.54 (0.35 to 0.73) and 0.95 (0.91 to 0.97), respectively, showing the highest IRR in box III (Table 2). Thus, the IRRs of SAPS3 sum score and predicted in-hospital death were high whereas the IRR of

SAPS3 box II was quite low. The ICC value of SAPS3 box II and corresponding variables are presented in Table 3.

## Discussion

In this study, we found trustworthy ICC of nurse-assessed SAPS3 scores and predicted mortality in ICU patients. Strand K et al[6] established the IRR of doctor-assessed SAPS2 and SAPS3 scores. Their study design was com-

**Table 2.** ICC for SAPS3 score and predicted mortality

Variables	ICC score (95% CI)
SAPS3 Box I	0.90 (0.82 to 0.95)
Box II	0.54 (0.35 to 0.73)
Box III	0.95 (0.91 to 0.97)
SAPS3 sum score	0.89 (0.82 to 0.95)
Predicted mortality	0.91 (0.85 to 0.96)

ICC: intraclass correlation coefficients; SAPS3: simplified acute physiology score 3; CI: confidence interval.

**Table 3.** ICC score of box II components

Variables	ICC score	95% CI
ICU admission: planned or unplanned	0.09	0.04 to 0.29
Reasons for ICU admission		
Cardiovascular	0.51	0.32 to 0.71
Hepatic	0.48	0.29 to 0.68
Digestive	0.36	0.18 to 0.58
Neurologic	0.28	0.11 to 0.51
Surgical status at ICU admission	0.63	0.46 to 0.79
Anatomical site of surgery	0.68	0.51 to 0.82
Acute infection at ICU admission (nosocomial)	0.13	0.01 to 0.34
Acute infection at ICU admission (respiratory)	0.26	0.09 to 0.49

ICC: intraclass correlation coefficients; ICU: intensive care unit; CI: confidence interval.

**Table 4.** Comparison of our study design and Strand K et al[6]'s

	Our study	Strand K et al study[6]
ICU characteristics	Surgical ICU of 27 beds	General ICU of 12 beds
Raters	5 ICU nurses	10 junior anesthesiologists
Experience of raters	3 to 11 years	2 to 6 months of full-time ICU experience
Education	2 hours for SAPS3 score	2.5 hours for SAPS2 and SAPS3 score
Number of scored patients	27	24

ICU: intensive care unit; SAPS3: simplified acute physiology score 3.

pared with that of this study in Table 4.

As far as the results of both studies are concerned, the ICC value of SAPS3 box II was equally low in this study as that calculated based on doctor-assessed scores in the previous study.[6] However, nurse-assessed SAPS3 score in this study exhibited a high IRR of 0.89, while doctor-assessed SAPS2 and SAPS3 scores exhibited 0.84 and 0.80, respectively.[6]

According to our analysis of SAPS3 subscores, the ICC value of SAPS3 box II was 0.54, showing the lowest IRR for variables dependent on diagnostic information. Coincidentally, the ICC of SAPS3 box II was also 0.54 in Strand K et al's findings.[6] The source of disagreement among raters appears to include that variables are open to different interpretations and that some complex of variable definitions are not easy to remember. Training can focus on such sources of disagreement to improve the reliability among raters. The ICC of SAPS3 box I was slightly lower at 0.90 than 0.94 reported by Strand K et al.[6] However, the reliability of SAPS3 box III was 0.95, showing a significant difference from 0.73 in their study. Overall, this study demonstrated higher ICC of SAPS3 sum score than the study of Strand K. et al[6], which conducted with doctors (0.89 vs 0.80).

As the latest upgraded version of SAPS2, SAPS3 offers a high explanatory power of 50% in SAPS3 box I, which is followed by box II (22.5%) and box III (27.5%).[4] Given smaller explanatory power of Box II to the total sum score, the wide disagreement among nurse raters on box II (where clinical judgment is required) did not affect the overall IRR of SAPS3 score significantly. In our study, IRR of box III was higher (0.95) than previously reported level among doctor raters (0.73), contributing to high ICC of SAPS3 sum score.

Currently, to determine appropriate nurse staffing in the ICU, surgical intensive care unit optimal mobility score

(SOMS) or workload management system for critical care nurses (WMSCN) scoring systems are commonly used for disease severity assessment.[9] However, these tools are less useful to quantify severity of illness and mortality prediction for research purposes. Therefore, SAPS3 can be used for disease severity scoring, by nurses, in studies led by nurses. When SAPS3 scores are provided by multiple raters, IRR across the raters can pose an issue that needs to be resolved. This study demonstrated consistency among nurse raters.

The nurses studied had been working in the ICU for mean length of 7.8 years (range 3-11 years). Participant's relatively long working years as critical care nurses pose a limitation for this study. However, over 5-year work experience may be common among ICU nurses in Korea. Park and Gang[10] stated that 37.9% of ICU nurses in their study had working years of 5-10 years, making up the largest group. It can be said that the participants of this study represents the population to be studied. However, further study is needed to investigate the IRR of new nurses in the ICU. This study ascertains the consistency across nurse raters but does not establish the IRR between nurses and doctors. If both nurses and doctors assess SAPS3, the ICC should be measured again to determine consistency between them.

Despite benefits of severity scoring systems, scoring is not widely performed in the ICU largely due to constraints related to human resources, time and finance: human resources and time for manual data collection and financial resources for a new system that will combine electronic medical records and scoring tool for automatic process.[1] This study aimed to identify the reliability of ICU nurse-assessed severity score and demonstrated high ICC of nurse-assessed SAPS3, which is in agreement with the ICC value calculated from doctor-assessed SAPS3 scores. This study therefore suggests that ICU nurses are able to assess severity of illness in ICU patients and that the resulting scores can be reliable.[6]

In conclusion, this study established the reliability of nurse-assessed SAPS3 scores. Although the established IRR of SAPS3 sum score is satisfactory, compared with the IRR of doctor-assessed SAPS3 score, the overall IRR among can be further improved through better understanding of diagnosis-related variables evaluated for box II.

## Acknowledgements

The authors thank surgical ICU nurses of Ilsan Paik hospital (JO, SE, EJ, YK, MO) for their assistance in preparing this manuscript.

## ORCID

Jun Hyun Kim <http://orcid.org/0000-0002-9466-8399>  
 Ji Yeon Kim <http://orcid.org/0000-0002-4686-5638>  
 Kyung Woo Kim <http://orcid.org/0000-0003-1447-9385>

## References

- 1) Breslow MJ, Badawi O: Severity scoring in the critically ill: part 1-interpretation and accuracy of outcome prediction scoring systems. *Chest* 2012; 141: 245-52.
- 2) Breslow MJ, Badawi O: Severity scoring in the critically ill: part 2: maximizing value from outcome prediction scoring systems. *Chest* 2012; 141: 518-27.
- 3) Metnitz PG, Moreno RP, 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 1: objectives, methods and cohort description. *Intensive Care Med* 2005; 31: 1336-44.
- 4) 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: 1345-55.
- 5) Salluh JI, Soares M: ICU severity of illness scores: APACHE, SAPS and MPM. *Curr Opin Crit Care* 2014; 20; 557-65.
- 6) Strand K, Strand LI, Flaatten H: The interrater reliability of SAPS II and SAPS 3. *Intensive Care Med* 2010; 36: 850-3.
- 7) Walter SD, Eliasziw M, Donner A: Sample size and optimal designs for reliability studies. *Stat Med* 1998; 17: 101-10.
- 8) Altman DG: *Practical statistics for medical research*. 1st ed. London, CRC Press. 1990, p 628.

- 9) Mosenthal AC: Predicting outcome at the bedside: the surgical intensive care unit: optimal mobility score. *Evid Based Nurs* 2013; 16: 86.
- 10) Park HS, Gang EH: A study on job stress and the coping of ICU nurses. *Taehan Kanho Hakhoe Chi* 2007; 37: 810-21.