Serial Evaluation of the MODS, SOFA and LOD Scores to Predict ICU Mortality in Mixed Critically Ill Patients

Bodin Khwannimit MD*

** Division of Critical Care, Department of Internal Medicine, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla*

Objective: *To perform a serial assessment and compare ability in predicting the intensive care unit (ICU) mortality of the multiple organ dysfunction score (MODS), sequential organ failure assessment (SOFA) and logistic organ dysfunction (LOD) score.*

Material and Method: *The data were collected prospectively on consecutive ICU admissions over a 24-month period at a tertiary referral university hospital. The MODS, SOFA, and LOD scores were calculated on initial and repeated every 24 hrs.*

Results: *Two thousand fifty four patients were enrolled in the present study. The maximum and Δ-scores of all the organ dysfunction scores correlated with ICU mortality. The maximum score of all models had better ability for predicting ICU mortality than initial or delta score. The areas under the receiver operating characteristic curve (AUC) for maximum scores was 0.892 for the MODS, 0.907 for the SOFA, and 0.92 for the LOD. No statistical difference existed between all maximum scores and Acute Physiology and Chronic Health Evaluation II (APACHE II) score.*

Conclusion: *Serial assessment of organ dysfunction during the ICU stay is reliable with ICU mortality. The maximum scores is the best discrimination comparable with APACHE II score in predicting ICU mortality.*

Keywords: *Multiple organ failure, Organ dysfunction score, Intensive care, Critical illness, Severity of illness*

J Med Assoc Thai 2008; 91 (9): 1336-42
Full text. e-Journal: *http://www.medassocthai.org/journal*

Multiple organ dysfunctions or failure is a clinical pattern of progressive and sequential organ dysfunction, which is common in critically ill patients. Multiple organ failure is the leading cause of morbidity and mortality in patients admitted to an intensive care unit (ICU)[1-4]. ICU mortality has been correlated with the number of organ failures and the degree of organ dysfunction[1-3]. The assessment of organ failure scores during ICU stay may provide important information about a patients' illness, their response to treatment, describe the patient population in clinical trials and aid clinical decision making[5,6].

Commonly used organ failure based systems that have been studied include the multiple organ dysfunction score (MODS)[7], the sequential organ failure assessment (SOFA)[8], and the logistic organ dysfunction (LOD) Score[9]. These three organ failure scores have many similarities that include the same six organ systems, range of scores, and correlated patients' outcome[10-13]. Previous reports showed initial organ dysfunction scores satisfactorily identified the ICU mortality[14]. The LOD score was found to be accurate for predicting hospital mortality in Thailand[15]. However, organ failure is a dynamic process and the degree of dysfunction may vary with time and treatment[6]. Serial or repetitive assessment of organ dysfunction scores allows for a more effective representation of an outcome prediction than does a single measurement[10-13,16-19]. Several papers have reported that maximum, mean or delta scores demonstrated a better correlation with mortality than did an initial or first 24 hrs dysfunction score for each organ[10,11,13,16].

In the present study, the author was interested in evaluating the serial measurement of organ failure scores and comparing their ability at

Correspondence to: Khwannimit B, Division of Critical Care, Department of Internal Medicine, Faculty of Medicine, Prince of Songkla University, Hat Yai, Songkhla 90110, Thailand. Phone: 074-451-452, Fax: 074-429-385, Email: kbordin @medicine.psu.ac.th

predicting the ICU mortality in a group of mixed critically ill patients.

Material and Method

The present study was conducted in the ICU of Songklanagarind Hospital, an 854-bed tertiary referral university teaching hospital at the Prince of Songkla University in Songkhla, Thailand. In this hospital, there are two units in the adult ICU, a ten-bed mixed surgical ICU and a ten-bed mixed medical and coronary care unit. Approval for the project was obtained from the faculty Ethics Committee.

All data were collected concurrently for consecutive ICU admissions, over a two-year period from July1st, 2004 to June 30th, 2006. Patients were excluded if they were younger than 15 years of age, suffered burn injuries, and stayed in the ICU less than 24 hours. Only a patient's first admission to the ICU during the study period was included. Patients were followed up until ICU discharge in order to registrar their survival status.

The MODS, SOFA and LOD scores were calculated on initial (first 24 hrs) and daily until discharge from the ICU. The worst physiological values of each organ failure in the 24 hrs following ICU admission and those subsequent within 24 hrs were used for the present calculations as outlined in the original literature[7-9]. For patients who were sedated a Glasgow Coma Score (GCS) was determined either from their medical records before sedation or through interviewing the physician who ordered the sedation. However, if a variable could not be measured, the GCS was assumed normal. Acute Physiology and Chronic Health Evaluation II (APACHE II) score was calculated using the worst values for the first 24 hrs following ICU admission[20].

The maximum score was defined as maximum score at a certain time point during the ICU stay. The final score was referred to the score at the ICU discharge date or deaths. The delta (Δ) score was defined as the difference between the score at the reference time and the initial score such as the ΔLOD-48, which was the difference between the LOD at 48 hrs and the initial LOD score.

Data are presented as mean $\pm$ SD, when indicated. The Chi-square statistic was used to test for the statistical significance of categorical variables. A p-value less than 0.05 was considered statistically significant. The ability of the models for predicting ICU mortality was determined by examining their discrimination, that is to say the ability of the model to distinguish between a patient who will live and one who will die. Discrimination was measured by examining the area under the receiver operating characteristics curve (AUC). An AUC of one is a perfect discrimination and an AUC of 0.5 is a random chance. The model has good discrimination when the AUC is more than 0.8. The difference in the AUC was analyzed by the z statistic as described by Hanley and McNeil[21]. The correlation of the MODS, SOFA, and LOD was evaluated statistically with the Pearson's correlation coefficients. Statistics analysis was performed using the Stata 7 software (Stata Corporation, College Station Tx, USA.).

Results

Two thousand fifty four patients were including during the study period. One thousand two hundred and eighteen patients were admitted for at least 48 hrs and 375 for at least 96 hrs. Overall 332 (16.2%) died in the ICU and 437 patients (21.3%) died in the hospital. The patients' demographic and clinical characteristics are shown in Table 1. In comparison with those surviving patients, patients who died were

Table 1. Demographic and clinical characteristics of patients in this study

Age (year)	55.7 $\pm$ 18.0
Male	1178 (57.4)
APACHE II score	17.7 $\pm$ 9.5
APACHE II predicted hospital mortality (%)	28.8 $\pm$ 27.5
ICU LOS (day)	4.2 $\pm$ 5.3
Hospital LOS (day)	23.2 $\pm$ 26.2
Diagnostic Category	
Nonoperation	1126 (54.8)
Cardiovascular disease	404 (19.7)
Sepsis	322 (15.6)
Respiratory disease	110 (5.4)
Gastrointestinal disease	50 (2.4)
Neurological disease	37 (1.8)
Other	203 (9.9)
Elective operation	602 (29.3)
Emergency operation	326 (15.9)
Comorbidities	
Liver cirrhosis	30 (1.5)
Hematologic malignancy	76 (3.7)
Metastasis carcinoma	49 (2.5)
Immunocompromised	48 (2.4)
AIDS	25 (1.3)

APACHE II; acute physiology and chronic health evaluation II, LOS; length of stay, AIDS; acquired immune deficiency syndrome

nonoperative, cardiovascular disease, sepsis, gastrointestinal, and patients with comorbidities. In contrast, the survivor group significantly had respiratory problems and non-comorbidities. Age, gender, and patients with neurological disease were not significantly different between the survivor and non-survivor groups.

The maximum (Fig. 1) organ dysfunction scores were associated with ICU mortality. Initial all model scores of less than six predicted a mortality of less than 10% while initial all model scores greater than 14 predicted a mortality rate of 73.5%, 82.9%, and 93.9% for the MODS, SOFA, and LOD scores, respectively.

Trends in the organ dysfunction scores during 48 hrs and ICU mortality are presented in Fig. 2. When scores decreased or remained unchanged, the ICU mortality was less than 15%. However, ICU mortality was more than 30% and 45% when the score increased from 2-4 points and more than four points, respectively. The LOD score showed the best significant increase in ICU mortality when the score increased.

The AUC of the MODS, SOFA, and LOD scores during ICU admission, the Δ-scores and APACHE II score are shown in Table 2. The receiver operating characteristics curves and AUC of all maximum scores are shown in Fig. 3. In addition to pair-wise, calculated statistical differences of AUCs are presented in Table 2. Pearson's correlation coefficients for the maximum scores were 0.912 for MODS and SOFA ($p < 0.001$), 0.816 for MODS and LOD ($p < 0.001$), and 0.841 for SOFA and LOD ($p < 0.001$). However, Pearson's correlation coefficients for Δ-48 scores were 0.767 (Δ MOD and SOFA-48, $p < 0.001$), 0.572 (Δ MOD and LOD-48, $p < 0.001$) and 0.556 (Δ SOFA and LOD-48, $p < 0.001$).

If the authors consider the operative status and case-mix of coronary care and postoperative cardiac surgery, the AUC within the subgroup patients are calculated. When the postoperative patients were excluded, the AUC of the maximum MODS, SOFA, and LOD scores were 0.888, 0.898, and 0.912, respectively. The AUC was 0.888 of the maximum MODS, 0.900 for the maximum SOFA, and 0.912 for the maximum LOD after excluding coronary and postoperative cardiac patients. Not all the AUCs were shown statistically significant when compared pair-wise with each maximum score of all patients.

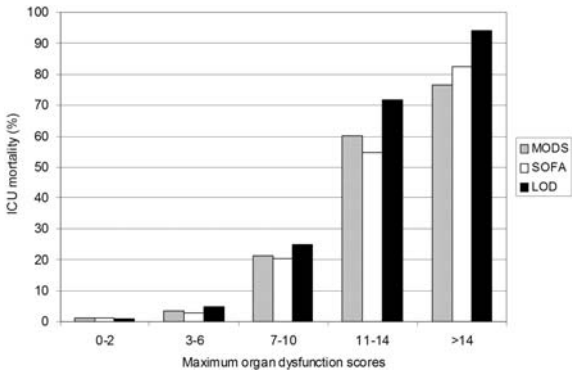

Fig. 1 ICU mortality rate in relation to the maximum score of three organ dysfunction scores

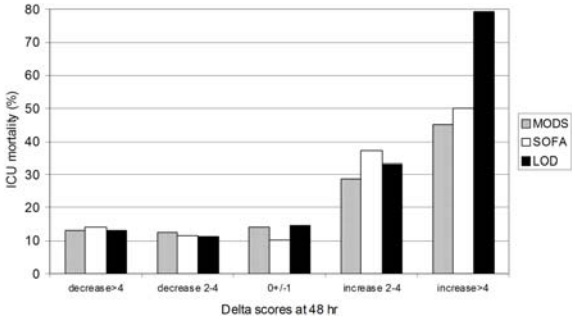

Fig. 2 Correlation of the Δ-48 score of MODS, SOFA and LOD scores with ICU mortality

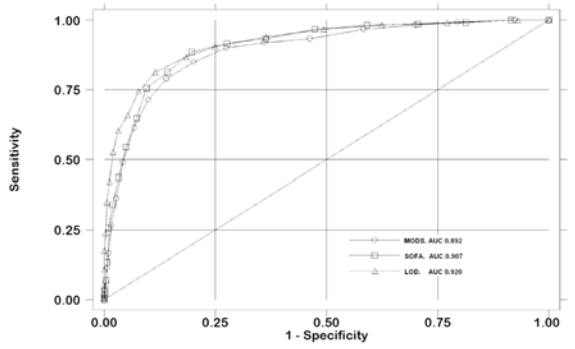

Fig. 3 Comparison of the areas under the receiver operating characteristic curves of the maximum MODS, SOFA and LOD scores for prediction of ICU mortality

Discussion

In the present study, the author evaluated the ability of serial measurement of the MODS, SOFA, and LOD scores to predict ICU mortality in a case-mix of Thai adult ICU patients. All initial and maximum scores of these models gave a good discrimination with the maximum LOD having the best AUC.

Table 2. Comparison of the AUC of serial organ dysfunction scores, delta scores and the APACHE II score as a predictor of ICU mortality

	Organ dysfunction scores			Delta scores	
	Score	AUC	95% CI	AUC	95% CI
MODS					
Initial	5.2 ± 3.9	0.865[a]	0.844-0.887		
Maximum	5.8 ± 4.2	0.892[b,c]	0.872-0.911	0.599	0.562-0.618
48 hr	4.9 ± 3.6	0.837	0.802-0.871	0.604	0.557-0.650
72 hr	5.0 ± 3.7	0.814	0.773-0.855	0.623	0.570-0.673
Final	5.4 ± 4.1	0.768	0.685-0.851	0.632	0.535-0.729
SOFA					
Initial	5.4 ± 4.1	0.879 [d]	0.860-0.899		
Maximum	5.9 ± 4.3	0.907[e]	0.890-0.924	0.599	0.571-0.628
48 hr	4.7 ± 3.7	0.861	0.832-0.889	0.628	0.581-0.674
72 hr	4.8 ± 3.9	0.825	0.787-0.864	0.635	0.583-0.688
Final	4.9 ± 4.1	0.731	0.644-0.818	0.608	0.506-0.709
LOD					
Initial	4.7 ± 3.9	0.883[f]	0.862-0.903		
Maximum	5.1 ± 4.2	0.920[g]	0.903-0.936	0.616	0.587-0.644
48 hr	3.7 ± 3.4	0.874	0.847-0.901	0.634	0.587-0.684
72 hr	3.6 ± 3.5	0.825	0.786-0.864	0.615	0.560-0.669
Final	3.9 ± 3.4	0.753	0.671-0.835	0.646	0.545-0.747
APACHE II	17.7 ± 9.5	0.905[h,i,j]	0.867-0.923		

[a] $p = 0.04$ between initial MODS and initial SOFA, [b] $p = 0.02$ between maximum MODS and maximum SOFA, [c] $p = 0.001$ between maximum MODS and maximum LOD, [d] $p = 0.69$ between initial SOFA and initial LOD, [e] $p = 0.09$ between maximum SOFA and maximum LOD, [f] $p < 0.001$ compared with APACHE II, [g] $p = 0.11$ compared with APACHE II, [h] $p < 0.001$ between APACHE II and initial MODS, APACHE II and initial SOFA, [i] $p = 0.014$ compared with maximum MOD, [j] $p = 0.87$ compared with maximum SOFA

Three previous studies have compared two organ dysfunction scores. Peres Bota et al[11] collected MODS and SOFA scores every 48 hrs, and they found both scores had good discrimination for predicting ICU mortality. All the AUC parameters in this present study were higher than Peres Bota reported; except the maximum MODS score. The second study was published by Zygun and colleagues[24]. The authors found an AUC of 0.63 for initial MODS, 0.64 for maximum MODS, 0.67 for initial SOFA and 0.69 for maximum SOFA[24]. Several differences between this study and Zygun's study warrant discussion. First, they calculated prospective daily results by automated sampling for both scores. Secondly, the MODS score was calculated using physiologic values measured at the same time every morning. Thus, methodological difference or ICU policy and care may have had an effect on the AUC curve of these models. The last report was by Timsit et al[17] and showed an AUC of initial SOFA and LOD of 0.733 and 0.729, respectively. Nevertheless, Timsit's study used logistic regression to customize both models and evaluate hospital mortality predictions as a result it is difficult to compare his results with the present study.

A previous single study that compared the ability of the MODS, SOFA, and LOD scores in predicting mortality was reported by Pettila et al[16]. All scores were collected retrospectively manually for day one, three, five, and seven after ICU admission. The maximum LOD score showed the best AUC to be the same as the present study. The initial and maximum LOD in the present study is better than Pettila reported. It is difficult to make a comparison of the present results to Pettila's study because the authors used hospital mortality as their outcome measurement but the present study has used the ICU mortality statistic as the primary outcome for evaluating the validity of organ dysfunction scores, which has been done for several reasons. First, most of the patients who die in the hospital from multiple organ dysfunction failure do so in the ICU. Second, the most common cause of death in the ICU is multiple organ failure. Finally, hospital

mortality should be altered to several factors including those that occur after discharge from the ICU.

The AUC for all initial organ dysfunction models was less than the APACHE II score because organ dysfunction scores were developed primarily to assess the severity of organ failure rather than predict mortality. Nevertheless, maximum of all scores have been demonstrated to predict ICU mortality, with the best AUC being: AUC > 0.8 and not different from the APACHE II score.

Daily organ score could also be used to determine the baseline severity of illness for patients and to quantify their response to therapy and thus could serve as a basis for evaluation of specific therapeutic interventions. The maximum organ dysfunction scores can identify the cumulative organ dysfunction and reflecting the greatest degree of organ specific life-support during ICU stay. The present study demonstrated a correlation of these scores with ICU outcome and discriminated the survivors from the nonsurvivors very well as has been previously reported[10-13]. It is also the same as the APACHE score[16]. The Δ-organ dysfunction scores exhibited the degree of organ failure developing during ICU stay. The Δ-score of all scores were associated significantly with an increase ICU mortality when the score increased, especially the LOD score. Thus, Δ-scores could be used to reflect patients response to therapy and risk of ICU mortality. Nevertheless, the organ dysfunction score also helps the physician and family to understand the risks and benefits of ICU therapies and clarifies their expectations. However, the author is strongly against withdrawal of therapy based only on high organ dysfunction scores in individual patients, because these models are only accurate for a group of patients.

Some limitations of this study should be addressed. First, the MODS score was calculated using the worst value within 24 hrs period subsequently they were not measured at the same point every morning as in the original publication. Second, although, all data was prospective and manually collected, it was not based on an automated computer-based information system. Finally, through studying only a single center it places limitations on the case-mix and quality of ICU care. A multicenter study would have given fewer concerns over the case-mix and also a better external validity.

Conclusions

Serial evaluation of the MODS, SOFA, and LOD scores throughout the ICU stay is a good predictor of ICU mortality in mixed critically ill Thai patients. The initial and maximum of all the scores are well discriminated for the ICU mortality. The maximum of all scores are a reliable outcome predictor as well as the APACHE II score. Increasing organ dysfunction scores during an ICU stay is associated with high ICU mortality.

Acknowledgement

This study was supported through a Faculty of Medicine research grant, Prince of Songkla University.

References

1. Baue AE. Multiple organ failure, multiple organ dysfunction syndrome, and systemic inflammatory response syndrome. Why no magic bullets? Arch Surg 1997; 132: 703-7.
2. Tran DD, Cuesta MA, van Leeuwen PA, Nauta JJ, Wesdorp RI. Risk factors for multiple organ system failure and death in critically injured patients. Surgery 1993; 114: 21-30.
3. Vincent JL, Sakr Y, Sprung CL, Ranieri VM, Reinhart K, Gerlach H, et al. Sepsis in European intensive care units: results of the SOAP study. Crit Care Med 2006; 34: 344-53.
4. Zimmerman JE, Knaus WA, Wagner DP, Sun X, Hakim RB, Nystrom PO. A comparison of risks and outcomes for patients with organ system failure: 1982-1990. Crit Care Med 1996; 24: 1633-41.
5. Cabre L, Mancebo J, Solsona JF, Saura P, Gich I, Blanch L, et al. Multicenter study of the multiple organ dysfunction syndrome in intensive care units: the usefulness of Sequential Organ Failure Assessment scores in decision making. Intensive Care Med 2005; 31: 927-33.
6. Le Gall JR. The use of severity scores in the intensive care unit. Intensive Care Med 2005; 31: 1618-23.
7. Marshall JC, Cook DJ, Christou NV, Bernard GR, Sprung CL, Sibbald WJ. Multiple organ dysfunction score: a reliable descriptor of a complex clinical outcome. Crit Care Med 1995; 23: 1638-52.
8. Vincent JL, Moreno R, Takala J, Willatts S, De Mendonca A, Bruining H, et al. The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine. Intensive Care Med 1996; 22: 707-10.
9. Le Gall JR, Klar J, Lemeshow S, Saulnier F, Alberti C, Artigas A, et al. The logistic organ dysfunction

system. A new way to assess organ dysfunction in the intensive care unit. ICU Scoring Group. JAMA 1996; 276: 802-10.
10. Ferreira FL, Bota DP, Bross A, Melot C, Vincent JL. Serial evaluation of the SOFA score to predict outcome in critically ill patients. JAMA 2001; 286: 1754-8.
11. Peres BD, Melot C, Lopes FF, Nguyen B, V, Vincent JL. The multiple organ dysfunction score (MODS) versus the sequential organ failure assessment (SOFA) score in outcome prediction. Intensive Care Med 2002; 28: 1619-24.
12. Buckley TA, Gomersall CD, Ramsay SJ. Validation of the multiple organ dysfunction (MOD) score in critically ill medical and surgical patients. Intensive Care Med 2003; 29: 2216-22.
13. Moreno R, Vincent JL, Matos R, Mendonca A, Cantraine F, Thijs L, et al. The use of maximum SOFA score to quantify organ dysfunction/failure in intensive care. Results of a prospective, multicentre study. Working Group on Sepsis related Problems of the ESICM. Intensive Care Med 1999; 25: 686-96.
14. Khwannimit B. A comparison of three organ dysfunction scores: MODS, SOFA and LOD for predicting ICU mortality in critically ill patients. J Med Assoc Thai 2007; 90: 1074-81.
15. Khwannimit B. Validation of the LOD score compared with APACHE II score in prediction of the hospital outcome in critically ill patients. Southeast Asian J Trop Med Public Health 2008; 39: 138-45.
16. Pettila V, Pettila M, Sarna S, Voutilainen P, Takkunen O. Comparison of multiple organ dysfunction scores in the prediction of hospital mortality in the critically ill. Crit Care Med 2002; 30: 1705-11.
17. Timsit JF, Fosse JP, Troche G, De Lassence A, Alberti C, Garrouste-Org, et al. Calibration and discrimination by daily logistic organ dysfunction scoring comparatively with daily sequential organ failure assessment scoring for predicting hospital mortality in critically ill patients. Crit Care Med 2002; 30: 2003-13.
18. Pettila V. Sequential assessment of multiple organ dysfunction as a predictor of outcome. JAMA 2002; 287: 713-4.
19. Cook R, Cook D, Tilley J, Lee K, Marshall J. Multiple organ dysfunction: baseline and serial component scores. Crit Care Med 2001; 29: 2046-50.
20. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med 1985; 13: 818-29.
21. Hanley JA, McNeil BJ. A method of comparing the areas under receiver operating characteristic curves derived from the same cases. Radiology 1983; 148: 839-43.
22. Metnitz PG, Lang T, Valentin A, Steltzer H, Krenn CG, Le Gall JR. Evaluation of the logistic organ dysfunction system for the assessment of organ dysfunction and mortality in critically ill patients. Intensive Care Med 2001; 27: 992-8.
23. Timsit JF, Fosse JP, Troche G, De Lassence A, Alberti C, Garrouste-Org, et al. Accuracy of a composite score using daily SAPS II and LOD scores for predicting hospital mortality in ICU patients hospitalized for more than 72 h. Intensive Care Med 2001; 27: 1012-21.
24. Zygun DA, Laupland KB, Fick GH, Sandham JD, Doig CJ. Limited ability of SOFA and MOD scores to discriminate outcome: a prospective evaluation in 1,436 patients. Can J Anaesth 2005; 52: 302-8.

การประเมินคะแนน MODS, SOFA และ LOD อย่างต่อเนื่องในการทำนายอัตราตายในหออภิบาลในผู้ป่วยวิกฤตรวม

บดินทร์ ขวัญนิมิตร

วัตถุประสงค์: *เพื่อประเมินคะแนน MODS, SOFA และ LOD อย่างต่อเนื่องและเปรียบเทียบความสามารถในการทำนายอัตราตายในหออภิบาล*

วัสดุและวิธีการ: *เก็บข้อมูลแบบต่อเนื่องไปข้างหน้าของผู้ป่วยที่เข้ารับการรักษาในหออภิบาลโรงพยาบาลระดับตติยภูมิและโรงเรียนแพทย์ระยะเวลา 24 เดือน คำนวณคะแนน MODS, SOFA และ LOD ทุก 24 ชั่วโมง*

ผลการศึกษา: *ผู้ป่วยเข้าร่วมการศึกษาทั้งหมด 2,054 ราย คะแนนอวัยวะล้มเหลวสูงสุดและผลต่างคะแนนของทั้งสามระบบสัมพันธ์กับอัตราเสียชีวิตในหออภิบาล คะแนนสูงสุดของทั้งสามระบบทำนายอัตราตายในหออภิบาลได้ดีกว่าคะแนนแรกรับและผลต่างคะแนน พื้นที่ใต้ receiver operating characteristic curve ของคะแนนสูงสุดของ MOD เท่ากับ 0.892 ระบบ SOFA เท่ากับ 0.907 และระบบ LOD เท่ากับ 0.92 ไม่พบความแตกต่างของคะแนนสูงสุดของทั้งสามระบบและคะแนน APACHE II ในการทำนายอัตราตายในหออภิบาล*

สรุป: *การติดตามคะแนนอวัยวะล้มเหลวอย่างต่อเนื่องช่วยในการประเมินอัตราตายในหออภิบาล คะแนนอวัยวะล้มเหลวสูงสุดทำนายอัตราตายในหออภิบาลได้เช่นเดียวกับคะแนน APACHE II*