CHAPTER I

INTRODUCTION

Rationale and background

Severity scoring systems are important tools to provide estimation of patient outcome probabilities. Within the field of intensive care (ICU) medicine, severity scoring systems are often used for mortality predictions which enable physicians to describe ICU populations, and can be helpful in clinical decision making, guiding resource allocation, stratification of patients for enrollment in clinical trials and controlling for severity of illness in auditing quality of care [1].

Well-known severity scoring systems in ICU are the Acute Physiology and Chronic Health Evaluation II (APACHE II) [2], the Simplified Acute Physiology Score II (SAPS II) [3] and the Mortality Probability Model II [4] (MPM_0 II and MPM_{24} II). All These severity scoring systems involve logistic regression models to predict the probability of in-hospital mortality. They use slightly different sets of covariates describing the demography, admission type, co-morbidity, and worst physiological status of the patient in the first 24 hours of ICU admission (or in case of the MPM_0 II, in the first hour of ICU admission).

These severity scoring systems were developed more than 10 years ago on other (American or European) patient populations than those to which they are applied now; their generalizability must be assessed before the systems can be used in clinical practice. Therefore, many studies have been published on validating and comparing these severity scoring systems in external settings, with the aim of choosing the best performing system and assessing its performance. These studies commonly focus on measuring the scoring systems' discrimination using the area under the Receiver Operating Characteristic (ROC) Curve, and their calibration using the Hosmere-Lemeshow goodness-of-fit statistics.

The results of these studies vary considerably. Whereas one study concluded that SAPS II is superior over MPM_{24} II [5], another found MPM_{24} II is superior over SAPS II [6]. Similarly, some studies concluded that the calibration of SAPS II is insufficient [5, 6, 7] whereas another concluded that SAPS II is sufficient [3, 8]. The variation in these results might be caused by temporal or geographical differences between the populations studied. However, only a few studies reported the performance of these scoring systems in Thailand, and there has been only one preliminary study on the area under the ROC curve of MPM_{24} II [9]. Khwannimit et al. [10] found good discrimination of SAPS II (area under the ROC curves of 0.888), but insufficient calibration, however, Lertsithichi et al. found area under the ROC curve of SAPS II of 0.81 in surgical patients [9] and reported excellent discrimination and good calibration of both scoring systems. A limitation of these studies was the relatively small case-mix populations, and suggested that customization or new severity scoring systems should be developed for Thai critically ill patients.

The goal of this study was to validate and compare the performance of the SAPS II and MPM_{24} II scoring systems in predicting hospital mortality in Thai critically ill patients admitted to intensive care units, before accepting these tools for routine use.