van den Boogaard, M., Schoonhoven, L., Maseda, E., Plowright, C., Jones, C., Luetz, A., Sackey, P. V., Jorens, P., Aitken, L. M., van Haren, F. M. P., Donders, R., van der Hoeven, J. G. & Pickkers, P. (2014). Recalibration of the delirium prediction model for ICU patients (PRE-DELIRIC): a multinational observational study. Intensive Care Medicine, 40(3), pp. 361-369. doi: 10.1007/s00134-013-3202-7

Recalibration of the Delirium Prediction Model for ICU Patients (PRE-DELIRIC); a multinational observational study.

M. van den Boogaard[1], L. Schoonhoven[2-3], E. Maseda[4], C. Plowright[5], C. Jones[6], A. Luetz[7], P.V. Sackey[8], P.G. Jorens[9], L.M. Aitken[10-11], F.M.P. van Haren[12], R. Donders[13], J.G. van der Hoeven[1], P. Pickkers[1]

Keywords: delirium, prediction model, recalibration, critical care

Corresponding author
Dr. M. van den Boogaard
Department of Intensive Care Medicine
Radboud University Medical Center
P.O. 6101, internal post 710. Zipcode 6500 HB, Nijmegen, The Netherlands
Tel.: +31-24-3617273, Fax: +31-24-3541612
E-Mail: Mark.vandenBoogaard@radboudumc.nl

[1] Radboud University Medical Center
Department of Intensive Care Medicine
Nijmegen, The Netherlands
Mark.vandenBoogaard@radboudumc.nl and Peter.Pickkers@radboudumc.nl and
Hans.vanderHoeven@radboudumc.nl

[2] Radboud University Medical Center
Scientific Institute for Quality of Healthcare
Nijmegen, The Netherlands
Lisette.Schoonhoven@radboudumc.nl

[3] University of Southampton
Faculty of Health Sciences
Southampton, United Kingdom
Lisette.Schoonhoven@radboudumc.nl

[4] Hospital Universitario La Paz,
Department of Intensive Care Medicine
Madrid, Spain
emilio.maseda@gmail.com

[5] Medway Maritime Hospital
Anaesthetic department
Kent, United Kingdom
Catherine.plowright@nhs.net

[6] Whiston Hospital
Ward 4E (Critical Care)
Prescot, United Kingdom
Christina.Jones@sthk.nhs.uk

[7] Charité – Universitaetsmedizin Berlin
Department of Anesthesiology and Intensive Care Medicine
Berlin, Germany
Alawi.Luetz@charite.de

[8] Karolinska University Hospital Solna
Department of Anesthesiology, Surgical Services and Intensive Care Medicine and Department of Physiology and Pharmacology, Karolinska Institute
Stockholm, Sweden
Peter.Sackey@karolinska.se

[9] Antwerp University Hospital, University of Antwerp
Department of Critical Care Medicine
Edegem (Antwerp), Belgium
Philippe.Jorens@uza.be

[10] Princess Alexandra Hospital
Intensive Care Unit
Brisbane, Australia
l.aitken@griffith.edu.au

[11] Griffith University
NHMRC Center for Research Excellence, Center for Health Practice Innovation, Griffith Health Institute
Brisbane, Australia
l.aitken@griffith.edu.au

[12] Canberra Hospital
Department of Intensive Care
Canberra, Australia
frank.vanharen@act.gov.au

[13] Radboud University Medical Center
Department for Health Evidence
Nijmegen, The Netherlands
Rogier.Donders@radboudumc.nl

Abstract

Purpose: Recalibration and determining discriminative power, internationally, of the existing delirium prediction model (PRE-DELIRIC) for intensive care patients.

Methods: A prospective multicenter cohort study was performed in eight intensive care units (ICUs) in six countries. The 10 predictors (age, APACHE-II, urgent and admission category, infection, coma, sedation, morphine use, urea level, metabolic acidosis) were collected within 24 hours after ICU admission. The confusion assessment method for the Intensive Care Unit (CAM-ICU) was used to identify ICU delirium. CAM-ICU screening compliance and inter-rater reliability measurements were used to secure the quality of the data.

Results: 2,852 adult ICU patients were screened of which 1,824 (64%) were eligible for the study. Main reasons for exclusion were length of stay <1day (19.1%) and sustained coma (4.1%). CAM-ICU compliance was mean (SD) 82±16% and inter-rater reliability 0.87±0.17. The median delirium incidence was 22.5% (IQR 12.8%–36.6%). Although the incidence of all ten predictors differed significantly between centers, the area under the receiver operating characteristic (AUROC) curve of the 8 participating centers remained good: 0.77 (95%CI: 0.74-0.79). The linear predictor and intercept of the prediction rule were adjusted and resulted in improved re-calibration of the PRE-DELIRIC model.

Conclusions: In this multinational study we recalibrated the PRE-DELIRIC-model. Despite differences in the incidence of predictors between the centers in the different countries the performance of the PRE-DELIRIC-model remained good. Following validation of the PRE-DELIRIC model it may facilitate implementation of strategies to prevent delirium and aid improvements in delirium management of ICU patients.

Introduction

Delirium, the acute onset of confusion and consciousness disturbances with a fluctuating course [1], occurs frequently in critically ill patients [2-4]. Delirium is associated with a prolonged stay in the intensive care unit (ICU) and hospital, increased morbidity and mortality rate, higher costs [2, 3, 5] and adverse long-term outcome [6, 7]. There are several delirium assessment tools for ICU patients such as the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). Although recent studies [8, 9] showed a lower accuracy of the CAM-ICU than in the original studies [10, 11], this screening tool has the highest sensitivity and specificity [12, 13]. Structured delirium screenings results in better recognition of delirious patients [14] that may facilitate early treatment [15, 16]. Besides adequate delirium treatment, prevention of delirium is crucial. While some preliminary studies have reported effective preventive interventions in both non-critically ill [17, 18] and ICU patients [19], applying these interventions in all ICU patients is time consuming, inefficient and exposes a substantial number of patients to unnecessary risks to possible side-effects of drugs used for delirium prevention. A readily available prediction model to identify high-risk patients would facilitate the use of preventive interventions. Recently, the PRE-DELIRIC prediction model was developed and validated for ICU patients [20] based on identified risk factors for delirium in ICU patients [21]. The development of the prediction model including the relevance of different delirium-associated risk factors in daily ICU practice, such as use of sedatives, morphine and presence of an infection are discussed more extensively in the original article [21]. The discriminative power of the PRE-DELIRIC model was high in predicting delirium with an onset at median day two after ICU admission [20]. Using the PRE-DELIRIC model is effective in predicting delirium and can be used to guide preventive therapy in critically ill patients [22], to stratify patients in testing the effectiveness of any considered intervention and to better inform caregivers and families.

The PRE-DELIRIC model consists of ten predictors that are readily available within 24 hours following ICU admission and, with an area under the receiver operating characteristic curve (AUROC) of 0.85 [20] has a good performance. Since the PRE-DELIRIC model was developed and validated in the Netherlands, it is unknown what the multinational performance of this model is. In view of relevant differences in case mix and ICU treatment between countries, a good multinational performance of the PRE-DELIRIC model is warranted prior to worldwide implementation.

In the present multinational study we recalibrated the model and determined the discriminative power of the PRE-DELIRIC model.

Methods

Study design

Prospective observational multicenter study carried out in eight general intensive care units for adult patients in six countries (Australia, Belgium, Germany, Spain, Sweden, United Kingdom). The regional Medical Ethical Committee of Arnhem-Nijmegen, The Netherlands (study number 2010/365) approved the study and waived the need for informed consent, since CAM-ICU determinations were part of clinical practice in all centers, no additional interventions were carried out, so data collection was not burdensome to patients, and data were captured and analyzed anonymously. All participating centers obtained ethics approval from the Ethical Committee of their own institution for data collection.

Study population

Each participating center included all eligible ICU patients during a period of three months. The first center started with inclusion in October 2011 and the last center started in June 2012. Patients were excluded if they were: delirious within 24 hours after ICU admission; sustained comatose during complete ICU stay; admitted to the ICU for less than one day; suffering from serious auditory or visual disorders; unable to understand the language of the included center; severely mentally disabled; suffering from a serious receptive aphasia; or if the compliance rate of the delirium screening was <80% during a patients' stay in the ICU. To exclude a potential source of bias, the assessors of the CAM-ICU were not aware of collecting the data of the predictors neither the PRE-DELIRIC score and did not receive the calculated risk to develop delirium for their patient.

Delirium screening

In order to detect delirium, all ICU patients were assessed by well-trained ICU nurses with the validated delirium assessment tool the CAM-ICU [10, 11] at least twice daily. Identical to the original study [20], delirium was defined as at least one positive CAM-ICU screening during a patients' complete intensive care stay. CAM-ICU was part of clinical practice in all participating hospitals.

Data collection

Data relating to delirium screening was collected during patients' complete ICU stay. The ten predictors of the PRE-DELIRIC model as originally defined [20] were collected within the first 24 hours after ICU admission: age, APACHE-II score, coma, urgent admission (unplanned ICU admission), admission category (surgical, medical, trauma, neurology/neurosurgical), infection, coma, use of sedatives, morphine use (three dosages

groups), urea level, and metabolic acidosis [20]. All predictors can objectively be measured and are well defined (Appendix A).

A secured web based electronic clinical report form (E-CRF) was filled out for each screened patient using a unique login and password for each participating center. Consecutive patients received a unique anonymous number. For privacy reasons, only the participating centers were able to identify their patients', based on the E-CRF-numbers.

Data management and quality checks

To ensure the quality of the data, the compliance with the CAM-ICU was calculated monthly. Compliance was calculated as the percentage of assessments performed per day in relation to the total number of assessments that should have been performed. To determine the quality of the performed delirium screenings during the study period, monthly inter-rater reliability measurements were performed for all patients admitted to the ICU on a given day each month. For this, the CAM-ICU screenings assessed by the intensive care nurse were compared with the scores assessed by a dedicated delirium expert nurse or investigator in each center.

We determined *a priori* that a CAM-ICU screening compliance >80% and an inter-rater reliability of >0.80 Cohen's kappa indicated reliable data. The performance of the PRE-DELIRIC model was calculated after excluding data of the center(-s) who did not achieve this compliance or kappa from the analysis. Exclusion and re-analysis was performed per centre. If exclusion of this center did not affect the performance of the PRE-DELIRIC model significantly results centers were included in the final analyses.

Statistical analysis

Missing predictor data were imputed in a similar way as in the original study [20]. We assumed that if a blood value was not determined, most likely the missing variable had a normal value, so the mean normal value of the study population was imputed. For other missing variables we assumed that they had a normal or negative value (i.e. no infection, no metabolic acidosis) or a mean value (e.g. APACHE-II score) of the study population and imputed the mean value of the variable derived from the delirium or non-delirium group, depending on the results of the delirium assessment. The percentage of missing data ranged between <1.0% and 9.8%. We recorded incomplete data for the presence of infection (9.8%), highest urea level (1.8%), APACHE-II score (1.5%) and metabolic acidosis (<1.0%). Data of all other variables were complete.

To determine the performance of the PRE-DELIRIC model for each participating center the original linear predictors were used to calculate the probability of developing delirium for each patient. The estimated

prognostic ability of the model was determined using the area under the receiver operating characteristic curve (AUROC) of the calculated total predicted probability per patient and his/her delirium outcome.

In order to optimize the calibration of the model we used the linear predictors and the intercept in a logistic regression model. For this a generally accepted [23, 24] standard statistical stepwise approach was followed in order to achieve a calibration slope of 1 and an intercept of 0, as a measure of perfect calibration. To test this we used the weights of the linear predictors in a logistic regression analysis resulting in an intercept and a calibration slope. The first approach was to estimate a new intercept and use a fixed calibration slope of 1. Next, we estimated the intercept as well as the calibration slope. Then we estimated the intercept for each center separately with again a fixed calibration slope, followed by estimation of intercept as well as calibration slope per center. With the last approach to optimize the calibration we then applied a general linear mixed model fit by Laplace approximation, using the mean estimated intercept and mean estimated calibration slope. In order to determine if recalibration could be biased by data of the largest group of patients from one center, we also calculated an intercept and linear predictor using weighted data. In order to test the calibration we used the Hosmer-Lemeshow goodness-of-fit statistics before and after recalibration [25], and to judge the calibration and recalibration visually we used calibration belts as described by Finazzi et al. [26].

Statistical analyses were performed using Statistical Package for Social Sciences (SPSS®) 20.01, R statistics version 2.10.1 [27] using the rms package [28]

Sample size

The PRE-DELIRIC model consists of ten predictors. We would needed at least 10-15 patients with delirium and 10-15 patients without delirium per predictor for the validation and re-calibration, so in total at least 300 patients. This formula was based on the recommendation for the development of a new prediction model [23]. With an anticipated delirium incidence of 15-30%, an attrition of 25%, we aimed to enroll at least 1350 patients.

Results

A total of 2,852 ICU patients were screened, with 1,824 (64%) patients included in the analysis. The most frequent reason for exclusion was a length of stay on the ICU <1 day (19.1%), followed by sustained coma (4.1%) and development of delirium within 24 hrs (3.5%) (Figure 1). The mean±SD age of patients was 60±17years, the mean APACHE-II score was 19±9, and 57% of the included patients were male. Most patients, over 50%, had a predicted delirium chance between 10-20%, Appendix B.

The median delirium incidence was 22.5% (IQR 12.8%–36.6%). The median time till first positive CAM-ICU occurred was 3 [1-6] days. Of note, apart from the considerable variation in delirium incidence there were important differences between countries concerning the incidence of delirium predictors (Table 1).

CAM-ICU compliance and inter rater reliability

The overall CAM-ICU compliance was 82±16% (minimum 52% and maximum 100%) and the mean inter-rater reliability measurements were 0.87±0.17 Cohen's kappa (minimum 0.57 and maximum 1.00), Appendix C. In total 461 inter rater measurements were performed. There were 10 false negative scores and 11 false positive scores resulting in a sensitivity of 0.93 (95%CI: 0.86-0.95) and a specificity of 0.97 (95%CI: 0.95-0.98).

Discrimination and recalibration of the PRE-DELIRIC

To determine the discriminative power of the PRE-DELIRIC model the AUROC was calculated per center and overall. The mean AUROC of the eight participating centers was 0.77 (95%CI: 0.74-0.79), Appendix D. The AUROC of the model in the early onset delirium group was 0.82 (95%CI 0.79-0.84) and for the late onset delirium group 0.68 (95%CI 0.66-0.71). The sensitivity of the PRE-DELIRIC model in this study was 0.70 and the specificity of 0.73 with a positive and negative Likelihood ratio of 2.43 and 0.39, respectively. After discarding all data of the centers with an overall CAM-ICU compliance below 80% (Appendix B) the AUROC remained similar: 0.79 (95% 0.76-0.82). The mean inter-rater reliability of all centers was >0.80.

To recalibrate the prediction model, four different approaches were used and calculated, as described in the statistical analysis section. None of the first three approaches resulted in a good calibration defined as a calibration slope of nearly 1 and an intercept of nearly 0 (data not shown). Using the general linear mixed model method resulted in an adjustment of the original intercept (-6.31*0.4724 -1.0545). To optimize the calibration slope each linear predictor was then multiplied with 0.4724 resulting in new predicted probabilities (Appendix D) per center. Table 2 reflects the old and new linear predictors and the intercept. This recalibration resulted in improvement of the calibration curve (Figure 2a and 2b), with a calibration slope of 1.09 and an intercept of

0.08. Following adjustment of the calibration slope the AUROC remained similar: 0.76 (95%CI 0.74-0.79). The Hosmer-Lemeshow test improved, Chi-square 797.95 (*p<0.0001)* before recalibration to Chi-square 15.85 (*p 0.045)* after recalibration, indicating a better overall calibration.

Sequentially we calculated a new intercept and linear predictor using a weighted data to determine if the center with the largest sample size biased our results. This resulted in a poorer calibration (data not shown). Importantly, in this center no inter-rater reliability was measured and had the highest APACHE-II score with a relatively low delirium incidence.

Discussion

We previously showed that the prediction of delirium by caregivers is inaccurate and that the PRE-DELIRIC model is of additional value [20]. However, as the model was developed and validated in the Netherlands, the predictive value of the model in other countries was unknown. In this multinational study we determined the discriminative power of the PRE-DELIRIC model for ICU patients was similar to the previous study and the calibration of the model was optimized.

In our study we found important differences between countries regarding the incidence of the ten predictors as well as the delirium incidence, which potentially could be explained by differences in case mix, severity of illness and differences in ICU admission policies, such as sedation protocols, and other ICU treatments. Although remarkably, and in line with other studies [29, 30], in our study the most sedated patients, and patients in coma within the first 24hrs after ICU admission, have the highest rate of delirium (table 1). Nevertheless, since no information was collected that may explain the reason of the observed differences in data entered into the model, we cannot further speculate about this and other differences. The differences in incidence of the predictors, and the already existing slightly overestimation of the model in the Dutch population [20], necessitated halving of the coefficient values of the predictors in order to optimize the calibration of the model in the multinational population. Despite these differences between the countries, the discriminative power of the PRE-DELIRIC model was not affected, indicating that the most important predictors for the development of delirium on the ICU are included in the model.

Furthermore, in this multinational study we only collect data of predictors which are in the PRE-DELIRIC model. Although we feel that others risk factors such as excessive alcohol consumption is clearly a very important risk factor to develop delirium, this risk factor was not in the original PRE-DELIRIC model [20]. Regarding the purpose of this study it is not appropriate to include other/additional risk factors. This would result in a completely new development of a delirium prediction model. Without the risk factor alcohol withdrawal the predictive value of the original model was high, despite we did not measured the prevalence of alcohol withdrawal in this multinational study, the performance of the model remained high. Alcohol withdrawal (acute withdrawal, delirium tremens, and its clinical manifestation) should be clearly distinguished from delirium itself. However, it is important to recognize that alcohol consumption itself is a risk factor for “plain” delirium [31, 32]. We feel that for patients with a high alcohol consumption or withdrawal we do not need a prediction model, these patients have a high risk and delirium preventive measures are anyway indicated for this group.

The increased morbidity and mortality associated with delirium in ICU patients warrant its prevention. There is some evidence that delirium prevention, i.e. by haloperidol is effective in non-cardiac surgery ICU

patients [19]. Importantly, the estimate of the efficacy of haloperidol in this study [19] is likely diluted since the preventive intervention was used in all patients, irrespective of their delirium risk. Theoretically, exclusion of ICU patients with a low risk of developing delirium may better reveal the beneficial effects of preventive measures. Indeed, with use of the PRE-DELIRIC model we previously showed that a low dose of haloperidol was associated with a reduced rate of delirium and mortality among ICU patients with a predicted risk of developing delirium >50%, and seems even more effective in the highest risk (predicted risk >90%) group [22]. However, this study was a pre-post design study that needs to be confirmed in a RCT, it illustrates the need for a delirium prediction model to facilitate the conduct of future prevention studies.

Importantly, the PRE-DELIRIC model is a static prediction model producing a single risk prediction value 24 hours after ICU admission. However, delirium in ICU patients is a complex, dynamic and multi-factorial syndrome. The current PRE-DELIRIC model may require on-going validation as new therapies and interventions emerge. For example, the use of new sedatives or analgesics may affect the development of delirium [33-35] and consequently could affect the performance of the model. Different risk factors may emerge in the future that may need to be investigated and included in the current PRE-DELIRIC model. In addition, since some patients develop delirium within 24 hours after ICU admission, an early delirium prediction model appears necessary in order to facilitate preventive measures in high risk patients immediately after ICU admission.

Furthermore, the discriminative power of the model remained similar and the calibration was optimized. Regarding the calibration plot there is still some overestimation of the PRE-DELRIC model for patients with a calculated risk of 50% and higher. It appears plausible that for the high risk group the study was underpowered, resulting in the observed overestimation. However, in these patients with a high-risk to develop delirium it is recommendable anyway to take preventive interventions, so the small overestimation of the model would not affect clinical decision making.

Our study has several limitations we wish to address. In our multinational study the discriminative power remained good. During the recalibration process it appeared that the most optimal way to recalibrate the model was to estimate a new intercept and linear predictors for each center separately. The best performance can be achieved when a prediction model is tailored to suit each individual ICU. This would result in the best discriminative power and calibration of such a model, but would impair comparisons between centers. Therefore, we feel it is desirable to have a prediction model that can be used in all hospitals and we chose to use the mean estimated new linear predictors and intercept of all hospitals. In this way the discriminative power remained high. Nevertheless, centers need to take into account that there can be some over- or underestimation in the

prediction of delirium when using the PRE-DELIRIC model in their ICU, especially in the highest risk group. Therefore caution is needed with the use of the model in patient populations with a high probability of delirium. In addition, since the PRE-DELIRIC model is now recalibrated using multinational data, a prospective multinational validation of the recalibration is warranted.

Second, coma represented by RASS level -3 or less, is an important predictor in the PRE-DELIRIC model, but coma can be biased by the effect of sedation which is suggested to be a confounder for delirium [36]. In our study we did not collect data on the duration of coma or on the relation between sedation and the relation with the onset of delirium. However, when excluding the predictor 'sedation or coma' from the model, this did not influence the discriminative power (data not shown), indicating that this did not affect our results to an important extent. Third, for missing data we did not use a specific imputation technique [37], however, in our view, a clinically relevant method to handle missing values. We assumed, similar to the original PRE-DELIRIC study [20], that a missing variable had a normal value, as there were apparently no indications to measure this variable, and consequently imputed the normal value. Since the incidence of missing values was low, our results were not affected importantly using this imputation technique. Fourth, we assessed the presence of delirium using the CAM-ICU. The performance of this assessment tool in daily practice has been re-evaluated recently [8, 9], and also been discussed in sedated patients [38, 39] and may not be as accurate as in the original validation studies [10, 11], however, ongoing bedside education results in a better performance [40]. On the other hand, in the re-evaluation studies the CAM-ICU was measured only once and compared with an expert screening, while in our multinational study the delirium diagnosis was based on all consecutive CAM-ICU screenings during a patients' complete ICU stay, increasing its sensitivity. Finally, we set threshold values for good data quality concerning CAM-ICU compliance and even inter-rater reliability measurements. Although not all centers achieved these thresholds, we demonstrated that this did not affect our results significantly. These issues increase the generalizability of our results, because the lower compliance with CAM-ICU screening may simply reflect real-life clinical practice.

Conclusion

The discriminative value of the PRE-DELIRIC model to predict delirium in ICU patients was confirmed and the predictive value of the model improved after recalibration in this multinational study. However, following recalibration, the model needs to be validated prospectively in order to support its use in clinical practice. Furthermore, caution is needed with the use of the model in patient populations with a high probability of delirium.

Authors' contribution

MvdB carried out the study, performed the statistical analysis (in collaboration with RD), and drafted the manuscript. PP and LS supervised the conduct of the study and writing of the paper. EM, CP, CJ, AL, PS, PJ, LMA, FvH provided the data of the other hospitals and corrected the manuscript. JvdH co-supervised and corrected the manuscript.

All authors read and approved the final manuscript.

The authors declare that they have no competing interests.

Acknowledgement

The authors would like to thank Amanda McCairn and Sue Dowling (research nurses, Whiston Hospital), Anna Schandl (PhD student, Karolinska University Hospital, Stockholm) Lena James and Rod Hurford (Princess Alexandra Hospital, Brisbane, Australia) Walter Verbrugghe, Petra Vertongen (MD/staff member and data management, Antwerp University Hospital, Belgium) for their help in collecting the patient data.

Data sharing: No additional data available

Table 1. Patient characteristics and predictors of included patients of the participating hospitals

	Belgium Antwerp (n=566)	**Germany** Berlin (n=223)	**Spain** Madrid (n=128)	**Sweden** Stockholm (n=77)	**Australia** Brisbane (n=329)	**Australia** Canberra (n=195)	**UK** Prescot (n=235)	**UK** Kent (n=71)
Age, years (mean, SD)	61±15	62±16	60±17	61±17	55±18	63±16	62±17	62±17
APACHE-II points (mean, SD)	26±8	17±8	8±5	14±7	16±6	18±6	17±7	15±7
No coma	499 (88%)	184 (83%)	114 (89%)	47 (61%)	239 (73%)	146 (75%)	138 (59%)	38 (54%)
Coma due to:								
- Medication induced	58 (10%)	37 (17%)	14 (11%)	23 (30%)	31 (9%)	34 (17%)	70 (30%)	28 (39%)
- Miscellaneous	0	2 (1%)	0	1 (1%)	5 (2%)	4 (2%)	4 (2%)	5 (7%)
- Combination	9 (2%)	0	0	6 (8%)	54 (16%)	11 (6%)	23 (10%)	0
No morphine use	347 (79%)	203 (91%)	77 (61%)	25 (42%)	258 (82%)	182 (94%)	175 (75%)	66 (93%)
- Morphine 0.01-7.1mg/day	30 (7%)	10 (5%)	13 (10%)	11 (18%)	7 (2%)	2 (1%)	4 (2%)	0
- Morphine 7.2-18.6mg/day	41 (9%)	8 (4%)	23 (18%)	15 (25%)	19 (6%)	3 (2%)	6 (3%)	0
- Morphine >18.6mg/day	20 (5%)	2 (1%)	13 (10%)	9 (15%)	31 (10%)	7 (4%)	48 (21%)	5 (7%)
Sedated	194 (34%)	35 (16%)	21 (16%)	43 (56%)	271 (82%)	83 (43%)	94 (40%)	33 (47%)
Urgent admission	330 (58%)	114 (51%)	45 (35%)	61 (79%)	159 (48%)	149 (76%)	228 (97%)	61 (86%)
Diagnose group								
- Surgical	286 (51%)	110 (49%)	92 (72%)	26 (34%)	196 (60%)	63 (32%)	65 (28%)	31 (44%)
- Medical	164 (29%)	55 (25%)	8 (6%)	39 (51%)	77 (23%)	112 (57%)	161 (69%)	38 (54%)
- Trauma	1 (0%)	24 (11%)	2 (2%)	12 (16%)	42 (13%)	12 (6%)	4 (2%)	2 (3%)
- Neurology/neurosurgical	115 (20%)	34 (15%)	26 (20%)	0	14 (4%)	8 (4%)	5 (2%)	0
Infection or strong suspicion	92 (16%)	39 (18%)	19 (15%)	51 (66%)	99 (30%)	80 (41%)	97 (41%)	39 (55%)
Metabolic acidosis	205 (36%)	18 (8%)	26 (20%)	29 (38%)	57 (17%)	91 (47%)	90 (38%)	9 (13%)
Highest urea level in mmol/L	4.9±3.7	16.0±11.3	15.5±7.6	11.1±8.7	7.9±6.4	9.3±5.9	11.5±9.6	13.5±12.7
Delirious, n (%)	86 (15%)	60 (27%)	23 (18%)	30 (39%)	42 (13%)	23 (12%)	73 (31%)	26 (37%)

Data are expressed as mean with standard deviation, unless reported otherwise

Table 2.
PRE-DELIRIC formula, old and new intercept and linear predictors

	Original values of linear predictors	*New values of linear predictors*
Intercept	-6.3131	-4.0369
Age (per year)	0.0387	0.0183
APACHE-II score per point	0.0575	0.0272
Coma; no	0	0
- Drug induced	0.5458	0.2578
- Miscellaneous	2.2695	1.0721
- Combination	2.8283	1.3361
Admission category		
- Surgery	0	0
- Medical	0.3061	0.1446
- Trauma	1.1253	0.5316
- Neurology-surgery	1.3793	0.6516
Presence of Infection	1.0509	0.4965
Presence of Metabolic acidosis	0.2918	0.1378
use of morphine; no	0	0
- 0.01-7.1mg	0.4078	0.1926
- 7.2-18.6mg	0.1323	0.0625
- >18.6mg	0.5110	0.2414
Use of sedatives	1.3932	0.6581
Urea concentration (per mmol/L)	0.0298	0.0141
Urgent admission	0.4004	0.1891

Figure 1. Flowchart of inclusion

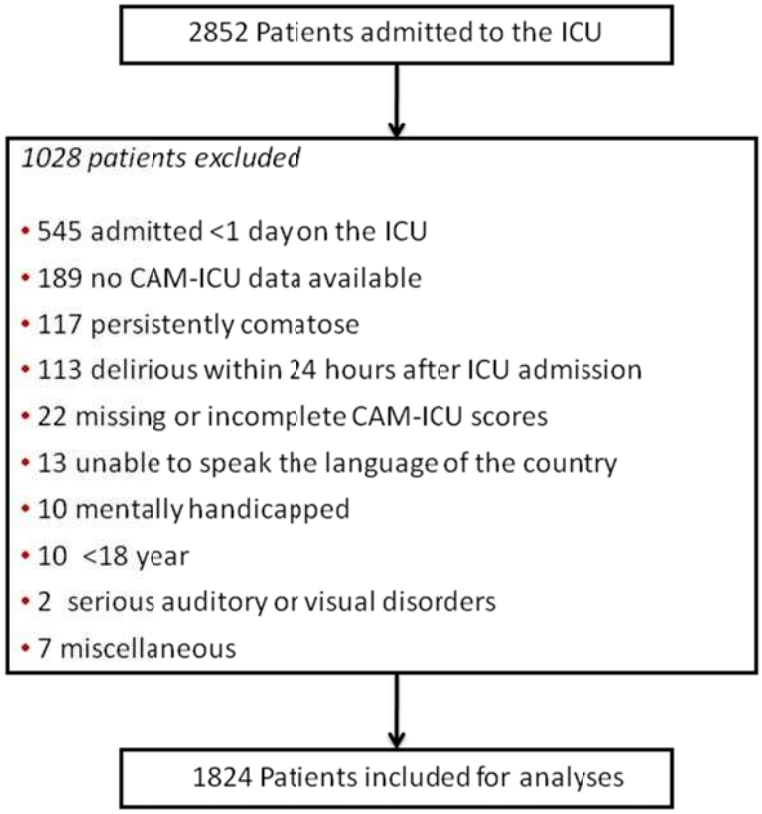


Figure 2a. Calibration belt before recalibration

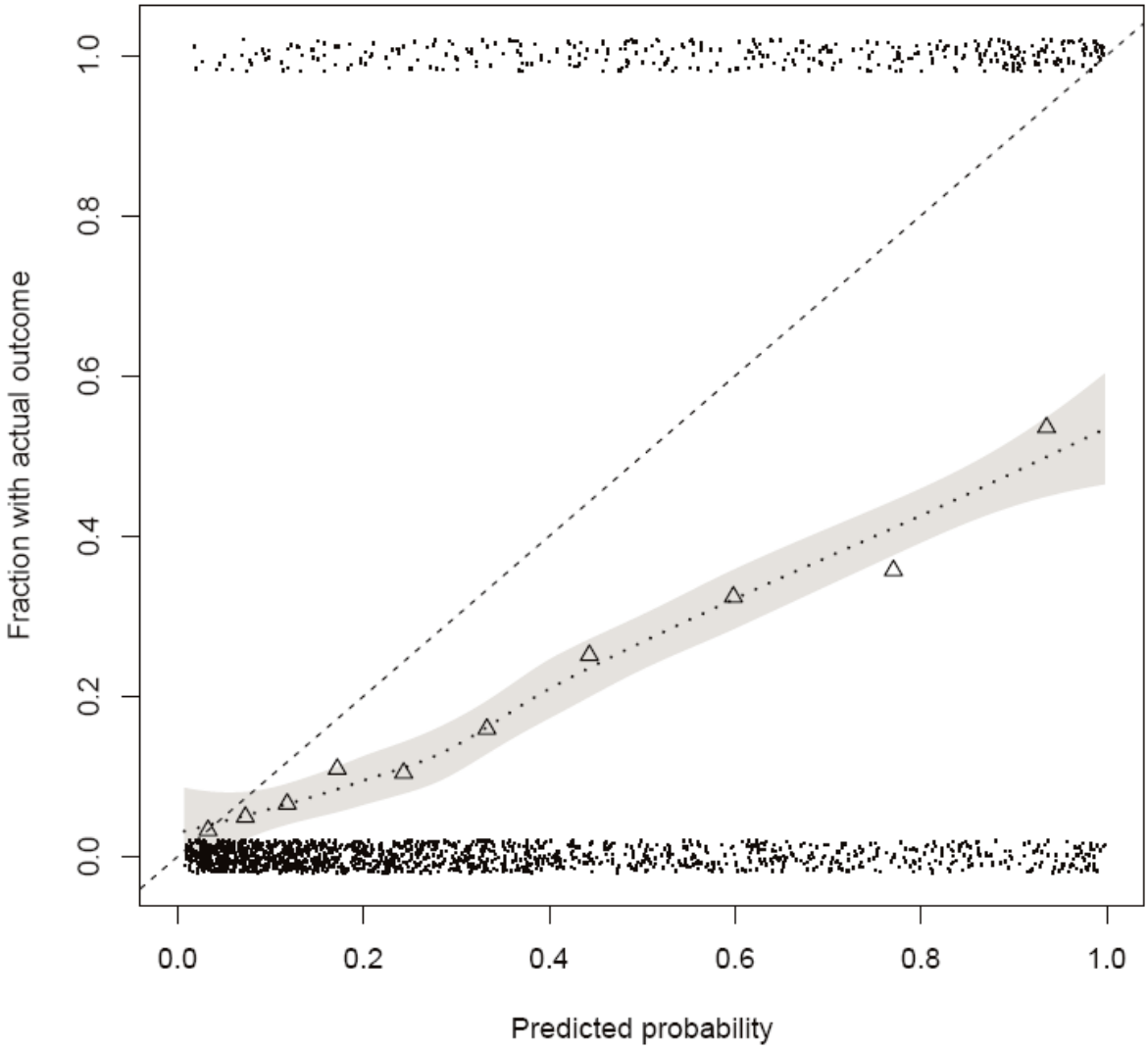
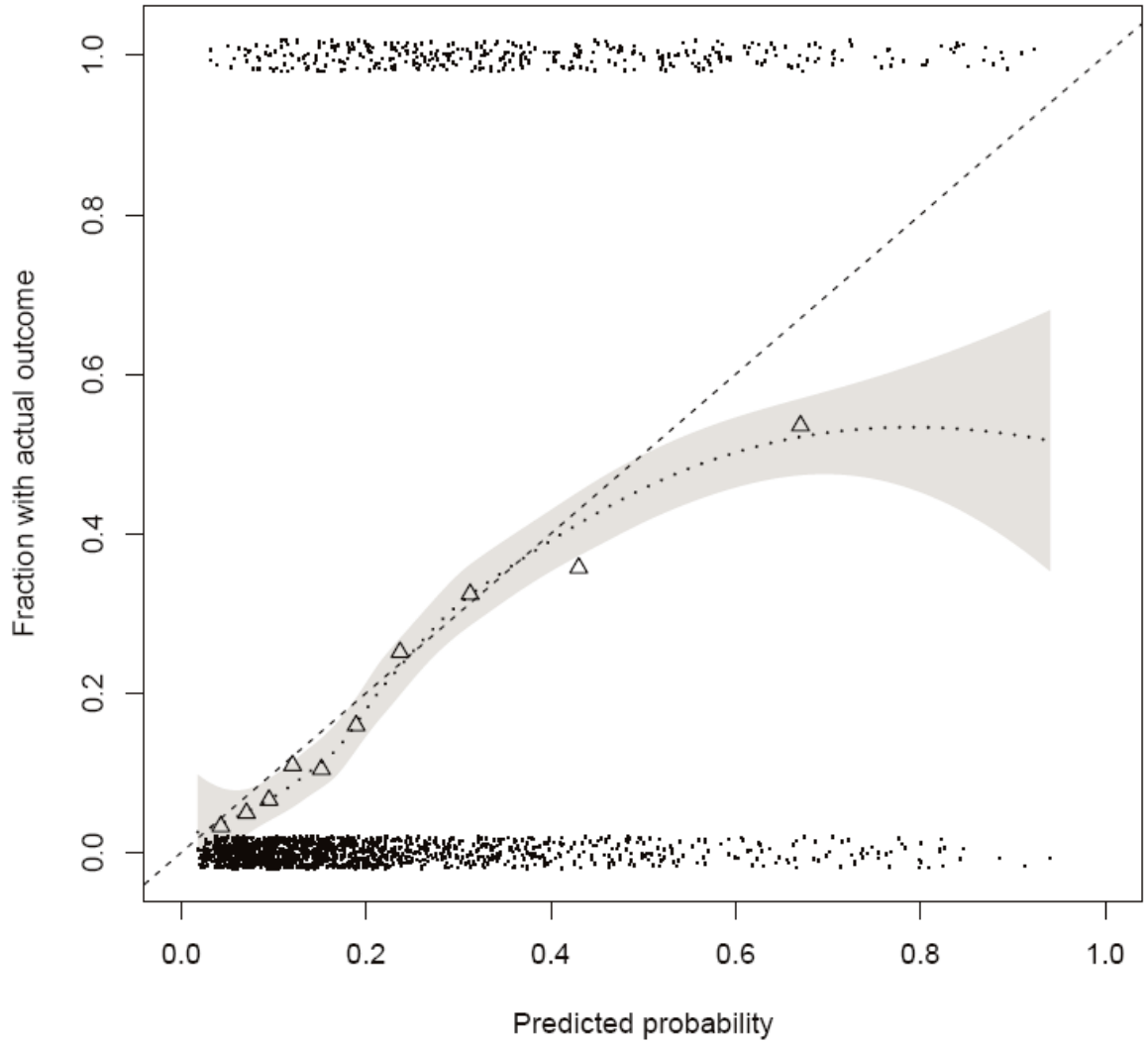


Figure 2b. Calibration belt after recalibration



Appendix A. *Supplement for web-only publication*

Collected delirium predictors within 24 hours after intensive care admission

Variable	**Category**	**Description**
Age *(years)*	C	Continuous variable
APACHE-II score *(per point)*	C	Calculated 24 hours after ICU admission
Coma	Cat	No coma: RASS-4/-5 maximum 8 hours RASS-4/-5 for longer than 8 hours: 1. With use of medication 2. Other (i.e. intra cerebral bleeding, post-resuscitation) 3. Combination (1+2)
Admission category	Cat	1. Surgical 2. Medical 3. Trauma 4. Neurology/neurosurgical
Infection	D	Proven or strong suspicion of infection for which antibiotics were started
Metabolic acidosis*	D	pH <7.35 with bicarbonate <24mmol/L
Morphine use	Cat	No morphine: no use of any morphine Cumulative use of any form of morphine: 1. 0.01-7.1mg 2. 7.2-18.6mg 3. 18.7mg or more
Sedative use	D	Any use of propofol, midazolam, lorazepam or combination
Urgent admission	D	Unplanned intensive care admission
Urea *(mmol/L)*	C	Continuous variable, highest value in blood

C= continuously D=dichotomized Cat.=categorical

Appendix B. *Supplement for web-only publication*

Predicted probabilities to develop delirium in decentiles groups

	Number of patients (%)
0 - 10% prediction	356 (19.5)
10-20% prediction	923 (50.6)
20-30% prediction	251 (13.8)
30-40% prediction	105 (5.8)
40-50% prediction	72 (3.9)
50-60% prediction	47 (2.6)
70-80% prediction	32 (1.8)
80-90% prediction	11 (.6)
90-100% prediction	0

Appendix C. *Supplement for web-only publication*

CAM-ICU compliance in %, and inter rater reliability measurements in Cohen's kappa

	CAM-ICU compliance	*Inter rater reliability*[#]
Belgium	78±7	not available
Germany	84±22	0.87±0.04
Spain	93±9	0.89±0.06
Sweden	88±11	0.87±0.13
Australia Brisbane	78±19	0.29±0.23
Australia Canberra	100	not performed[*]
UK_Prescot	61±1	0.79±0.09
UK_Kent	87±6	0.87±0.13
Overall	83±16	0.86±0.03

Data are expressed as mean and standard deviation

[*] *In this center all CAM-ICU were assessed by two dedicated research nurses, making this not applicable*

Appendix D. *Supplement for web-only publication*

AUROC of different hospitals/countries using PRE-DELIRIC en predicted probabilities

	Delirium	Median [IQR] predicted probability *before* recalibration	Median [IQR] predicted probability *after* recalibration	*AUROC*	*95% CI*
Belgium (%)	86 (15.2)	0.12 [0.08-0.25]	0.13 [0.11-0.21]	*0.79*	0.74-0.84
Germany (%)	60 (26.9)	0.09 [0.07-0.17]	0.11 [0.10-0.13]	*0.85*	0.80-0.91
Spain (%)	23 (18.0)	0.08 [0.06-0.11]	0.10 [0.09-0.13]	*0.88*	0.81-0.94
Sweden (%)	30 (39.0)	0.22 [0.09-0.41]	0.19 [0.12-0.32]	*0.71*	0.60-0.83
Australia Brisbane (%)	42 (12.8)	0.13 [0.09-0.33]	0.14 [0.11-0.25]	*0.81*	0.75-0.88
Australia Canberra (%)	23 (11.8)	0.16 [0.09-0.33]	0.16 [0.11-0.26]	*0.80*	0.72-0.89
Australia overall	65 (12.4)			*0.81*	0.76-0.86
UK_Prescot (%)	73 (30.8)	0.19 [0.09-0.36]	0.18 [0.12-0.27]	*0.65*	0.57-0.73
UK_Kent (%)	26 (36.6)	0.20 [0.07-0.41]	0.18 [0.10-0.32]	*0.76*	0.65-0.88
UK_overall	99 (32.4)			*0.68*	0.62-0.74
PRE-DELIRIC overall	**363 (19.9)**			***0.77***	**0.74-0.79**

Data are expressed as median with interquartile (25% and 75%) range, unless reported otherwise

References

1. Association AP (2013) Diagnostic and Statistical Manual of Mental Disorders (DSM-V). American Psychiatric Publishing, Arlington, VA
2. Dubois MJ, Bergeron N, Dumont M, Dial S, Skrobik Y, (2001) Delirium in an intensive care unit: a study of risk factors. Intensive Care Med 27: 1297-1304
3. Ely EW, Gautam S, Margolin R, Francis J, May L, Speroff T, Truman B, Dittus R, Bernard R, Inouye SK, (2001) The impact of delirium in the intensive care unit on hospital length of stay. Intensive Care Med 27: 1892-1900
4. Ouimet S, Kavanagh BP, Gottfried SB, Skrobik Y, (2007) Incidence, risk factors and consequences of ICU delirium. Intensive Care Med 33: 66-73
5. Milbrandt EB, Deppen S, Harrison PL, Shintani AK, Speroff T, Stiles RA, Truman B, Bernard GR, Dittus RS, Ely EW, (2004) Costs associated with delirium in mechanically ventilated patients. Crit Care Med 32: 955-962
6. Girard TD, Jackson JC, Pandharipande PP, Pun BT, Thompson JL, Shintani AK, Gordon SM, Canonico AE, Dittus RS, Bernard GR, Ely EW, (2010) Delirium as a predictor of long-term cognitive impairment in survivors of critical illness. Crit Care Med 38: 1513-1520
7. van den Boogaard M, Schoonhoven L, Evers AW, van der Hoeven JG, van Achterberg T, Pickkers P, (2012) Delirium in critically ill patients: Impact on long-term health-related quality of life and cognitive functioning. Crit Care Med 40: 112-118
8. van Eijk MM, van den Boogaard M, van Marum RJ, Benner P, Eikelenboom P, Honing ML, van der HB, Horn J, Izaks GJ, Kalf A, Karakus A, Klijn IA, Kuiper MA, de Leeuw FE, de MT, van der Mast RC, Osse RJ, de Rooij SE, Spronk PE, van dV, van Gool WA, Slooter AJ, (2011) Routine Use of the Confusion Assessment Method for the Intensive Care Unit: A Multicenter Study. Am J Respir Crit Care Med 184: 340-344
9. Reade MC, Eastwood GM, Peck L, Bellomo R, Baldwin I, (2011) Routine use of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU) by bedside nurses may underdiagnose delirium. Crit Care Resusc 13: 217-224
10. Ely EW, Inouye SK, Bernard GR, Gordon S, Francis J, May L, Truman B, Speroff T, Gautam S, Margolin R, Hart RP, Dittus R, (2001) Delirium in mechanically ventilated patients: validity and reliability of the confusion assessment method for the intensive care unit (CAM-ICU). JAMA 286: 2703-2710
11. Ely EW, Margolin R, Francis J, May L, Truman B, Dittus R, Speroff T, Gautam S, Bernard GR, Inouye SK, (2001) Evaluation of delirium in critically ill patients: validation of the Confusion Assessment Method for the Intensive Care Unit (CAM-ICU). Crit Care Med 29: 1370-1379
12. van den Boogaard M, Pickkers P, Schoonhoven L, (2010) Assessment of delirium in ICU patients; a literature review. Netherlands Journal of Critical Care 14: 10-15
13. Gusmao-Flores D, Figueira Salluh JI, Chalhub RA, Quarantini LC, (2012) The confusion assessment method for the intensive care unit (CAM-ICU) and intensive care delirium screening checklist (ICDSC) for the diagnosis of delirium: a systematic review and meta-analysis of clinical studies. Crit Care 16: R115
14. Spronk PE, Riekerk B, Hofhuis J, Rommes JH, (2009) Occurrence of delirium is severely underestimated in the ICU during daily care. Intensive Care Med 35: 1276-1280
15. Heymann A, Radtke F, Schiemann A, Lutz A, MacGuill M, Wernecke KD, Spies C, (2010) Delayed treatment of delirium increases mortality rate in intensive care unit patients. J Int Med Res 38: 1584-1595
16. van den Boogaard M, Pickkers P, van der Hoeven JG, Roodbol G, van Achterberg T, Schoonhoven L, (2009) Implementation of a delirium assessment tool in the ICU can influence haloperidol use. Crit Care 13: R131
17. Inouye SK, Bogardus ST, Jr., Charpentier PA, Leo-Summers L, Acampora D, Holford TR, Cooney LM, Jr., (1999) A multicomponent intervention to prevent delirium in hospitalized older patients. N Engl J Med 340: 669-676
18. Kalisvaart KJ, de Jonghe JF, Bogaards MJ, Vreeswijk R, Egberts TC, Burger BJ, Eikelenboom P, van Gool WA, (2005) Haloperidol prophylaxis for elderly hip-surgery patients at risk for delirium: a randomized placebo-controlled study. J Am Geriatr Soc 53: 1658-1666
19. Wang W, Li HL, Wang DX, Zhu X, Li SL, Yao GQ, Chen KS, Gu XE, Zhu SN, (2012) Haloperidol prophylaxis decreases delirium incidence in elderly patients after noncardiac surgery: A randomized controlled trial. Crit Care Med 40: 1-9
20. van den Boogaard M, Pickkers P, Slooter AJ, Kuiper MA, Spronk PE, van der Voort PHJ, van der Hoeven JG, Donders R, van AT, Schoonhoven L, (2012) Development and validation of PRE-

DELIRIC (PREdiction of DELIRium in ICu patients) delirium prediction model for intensive care patients: observational multicentre study. BMJ 344: e420
21. Van Rompaey B, Elseviers MM, Schuurmans MJ, Shortridge-Baggett LM, Truijen S, Bossaert L, (2009) Risk factors for delirium in intensive care patients: a prospective cohort study. Crit Care 13: R77
22. van den Boogaard M, Schoonhoven L, Van Achterberg T, Van der Hoeven JG, Pickkers P, (2013) Haloperidol prophylaxis in critically ill patients with a high risk for delirium. Critical Care 17: R9
23. Harrell Jr FE (2001) Regression modeling strategies. With applications to linear models, logistic regression, and survival analysis. Springer,
24. Steyerberg EW, Eijkemans MJ, Harrell FE, Jr., Habbema JD, (2001) Prognostic modeling with logistic regression analysis: in search of a sensible strategy in small data sets. Medical decision making : an international journal of the Society for Medical Decision Making 21: 45-56
25. Lemeshow S, Hosmer DW, Jr., (1982) A review of goodness of fit statistics for use in the development of logistic regression models. Am J Epidemiol 115: 92-106
26. Finazzi S, Poole D, Luciani D, Cogo PE, Bertolini G, (2011) Calibration belt for quality-of-care assessment based on dichotomous outcomes. PLoS One 6: e16110
27. Team RDC (2009) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
28. Frank EH, Jr., with contributions from many other u (2009) Hmisc: Harrell Miscellaneous.
29. Pandharipande P, Ely EW, (2006) Sedative and analgesic medications: risk factors for delirium and sleep disturbances in the critically ill. Crit Care Clin 22: 313-327, vii
30. Roberts DJ, Haroon B, Hall RI, (2012) Sedation for critically ill or injured adults in the intensive care unit: a shifting paradigm. Drugs 72: 1881-1916
31. Awissi DK, Lebrun G, Coursin DB, Riker RR, Skrobik Y, (2013) Alcohol withdrawal and delirium tremens in the critically ill: a systematic review and commentary. Intensive Care Med 39: 16-30
32. Reynolds T, Cooke F, Murch N, (2012) Problem based review: alcohol-use disorders on the Acute Medical Unit. Acute medicine 11: 101-106
33. Riker RR, Shehabi Y, Bokesch PM, Ceraso D, Wisemandle W, Koura F, Whitten P, Margolis BD, Byrne DW, Ely EW, Rocha MG, (2009) Dexmedetomidine vs midazolam for sedation of critically ill patients: a randomized trial. JAMA 301: 489-499
34. Pandharipande PP, Pun BT, Herr DL, Maze M, Girard TD, Miller RR, Shintani AK, Thompson JL, Jackson JC, Deppen SA, Stiles RA, Dittus RS, Bernard GR, Ely EW, (2007) Effect of sedation with dexmedetomidine vs lorazepam on acute brain dysfunction in mechanically ventilated patients: the MENDS randomized controlled trial. JAMA 298: 2644-2653
35. Hughes CG, McGrane S, Pandharipande PP, (2012) Sedation in the intensive care setting. Clinical pharmacology : advances and applications 4: 53-63
36. Riker R, Shehabi Y, Wisemandle W, Rocha M (2012) Relationship Between Delirium Incidence Assessed With Cam-Icu and Level of Sedation Assessed By Rass in Adult Icu Patients. In: Editor (ed)^(eds) Book Relationship Between Delirium Incidence Assessed With Cam-Icu and Level of Sedation Assessed By Rass in Adult Icu Patients. City, pp.
37. Donders AR, van der Heijden GJ, Stijnen T, Moons KG, (2006) Review: a gentle introduction to imputation of missing values. J Clin Epidemiol 59: 1087-1091
38. Haenggi M, Blum S, Brechbuehl R, Brunello A, Jakob SM, Takala J, (2013) Effect of sedation level on the prevalence of delirium when assessed with CAM-ICU and ICDSC. Intensive Care Med 39: 2171-2179
39. Can delirium Assessments Be Accurately Labelled Investigators g, Devlin JW, Fraser GL, Joffe AM, Riker RR, Skrobik Y, (2013) The accurate recognition of delirium in the ICU: the emperor's new clothes? Intensive Care Med 39: 2196-2199
40. Vasilevskis EE, Morandi A, Boehm L, Pandharipande PP, Girard TD, Jackson JC, Thompson JL, Shintani A, Gordon SM, Pun BT, Wesley EE, (2011) Delirium and sedation recognition using validated instruments: reliability of bedside intensive care unit nursing assessments from 2007 to 2010. J Am Geriatr Soc 59 Suppl 2: S249-S255