Intensive care unit benchmarking

Prognostic models for length of stay and presentation of quality indicator values

Verburg, I.W.M.

Publication date
2018

Document Version
Other version

License
Other

Appendix

Summary

The awareness for quality of care has grown among various stakeholders, due to a drive for continuous quality improvement, a pressure on accountability and budgetary constraints. Healthcare institutions compare their quality indicators with their own historical values or their peers in a process called benchmarking to identify opportunities for quality of care improvement.
Nowadays, intensive care unit (ICU) care is very complex and delivered in a highly technical and labor-intensive environment. The costs of ICU care are substantial resulting in a high proportion of the health care budget being spent on ICUs [12]. This makes ICUs a particularly interesting part of the hospital to assess and improve quality of care.
Ideally, quality indicator values represent true values of quality of care and differences between indicator values indicate room for quality of care improvement. However, observed differences in care quality can also arise from noise caused by differences in patient characteristics (case-mix), registration errors, residual confounding, and random variation. This noise may influence the quality indicator values of institutions and could lead to incorrect judgements. Fair and meaningful benchmarking requires correction for differences in patient case-mix, which can be done partially by using prognostic models.
Since costs are strongly related to ICU length of stay [23, 24], it can play an important role in examining the efficiency of care. ICU patients have a wide range of complex health issues, of which each may have a different association with ICU length of stay [25]. Prognostic models for ICU length of stay are not frequently used and little consensus exists on the best method for predicting ICU length of stay and the predictive performance of existing models is modest [26–29].
The first part of this thesis, chapters 2 to 4, addressed the development and performance of prognostic models for ICU length of stay. The second part of this thesis, chapter 5 to 7, addressed the presentation of quality indicators values. As no single quality indicator reflects the whole spectrum of healthcare performance we analyzed in chapter 5 the association between different commonly used ICU quality indicators: in-hospital mortality, readmission to the ICU within 48 hours after ICU discharge, and ICU length of stay. Furthermore, we introduced league tables in chapter 6 and funnel plots in chapter 7 as methods to report values of quality indicators.

Part I: Prognostic models for ICU length of stay

Chapter 2 described a systematic review on the development and validation of prognostic models for ICU length of stay. In total 11 studies were identified. We defined four requirements to examine the suitability of the models for benchmarking: 1) the parameters required to predict ICU length of stay have been published; 2) the model does not include any organizational characteristics; 3) the model has a low level of bias, demonstrated by at least moderate to very good calibration; 4) the model produces accurate predictions.
The performance of the included models were judged on accuracy based on the squared Pearson's correlation coefficient (R^2<36 %), and calibration (minimal moderate calibration as defined by Calster et al.) [62]. The included studies reported a percentage explained variance between 5% and 28% across patients and between 1% and 64 % across ICUs. Only two studies fulfilled our requirements for accuracy on ICU level [26, 28], however these two models did not published moderate calibration.
As none of the models fulfilled all our requirements we concluded that no existing prognostic model is suitable for planning, identifying unexpectedly long ICU length of stay, or benchmarking purposes. Physicians using these prognostic models should interpret them with caution.

Chapter 3 compared the performance of eight regression methods to develop a prognostic model that predict ICU length of stay using patient characteristics only. The predictive performance of the models was assessed by using the bootstrap method to calculate the performance measures R^2, root mean squared prediction error (RMSPE), mean absolute prediction error (MAPE), and bias.
Data of 32,667 unplanned ICU admissions of ICUs participating in the Dutch National Intensive Care Evaluation foundation (NICE) registry during the year 2011 were included. We concluded that the predictive performance of our own developed models was disappointing: R^2 was at most around 20% at patient level and the models had a RMSPE of more than seven days. Even in absolute terms our predictions were, on average, three days different from the observed ICU length of stay. The differences in predictive performance between the models were generally small.
We concluded that it is difficult to predict length of stay of unplanned ICU admissions using patient characteristics at ICU admission only. ICU discharge decisions often do not only depend on a patient's recovery, but also on organizational characteristics of the ICU such as availability of beds on the general ward and the need to free up ICU beds for other patients.

Chapter 4 described the association between ICU organizational characteristics and ICU length of stay after correction for patients characteristics. Additionally, we compared the predictive performance of a model for ICU length of stay correcting for patient and ICU organizational characteristics with that of a model only correcting for patient characteristics.
We performed several mixed-effect regression analyses correcting for patient characteristics and one additional ICU organizational characteristic in each model and used ICU as random intercept. The predictive value of an ICU organizational characteristic was tested by comparing the difference in the residual deviances.
We included data of 78,822 ICU admissions from the NICE registry admitted in 2014 and 2015. The following ICU organizational characteristics were associated with ICU length of stay: number of hospital beds; number of ICU beds; availability of fellows in training for intensivist; full-time equivalent ICU nurses; nurse to patient ratio, and discharged in a shift with 100% bed occupancy. However, we concluded that additional ICU organizational characteristics correction did not significantly improve the performance compared to a model with patient characteristics correction.

To conclude the first part of the thesis we put the results in the context of the objectives for the application of a prognostic model. For capacity planning and identification of patients with unexpected long length of stay, predictions needs to be reliable for individual patients. In this thesis we concluded that the accuracy on patient level of the prognostic models was not sufficient.
Our work has sparked debate in the scientific literature after it was publiced. Straney and co-workers [105] have argued that poor model performance at patient level may not be indicative for poor utility of a model for benchmarking purposes. Similarly, Kramer.[187] has claimed that some of the models included in our review described in chapter 2.
We agree that, for benchmarking purposes, a prognostic model needs to predict average ICU length of stay at ICU level accurately. However if a model fails to also predict patient-level outcomes accurately, we cannot exclude the possibility that there exists significant residual variation in case mix. Therefore we remain cautious with recommending such a model for benchmarking in practice. We examined model performance at ICU level for the model presented in chapter 3. We found an R^2 of 64%, which indicates that the accuracy ICU level predictions would be sufficient for the use of benchmarking. The calibration plot of mean predicted ICU length of stay against mean observed ICU length of stay based on 2% percentiles of predicted ICU length of stay was found to be satisfactory. We do recommend that further research is performed on accuracy for subgroups of patients and the calibration of models for the NICE registry.

Part II: Presentation of quality indicators values

Chapter 5 described the association between case-mix adjusted quality indicators for in-hospital mortality, readmission to the ICU within 48 hours after ICU discharge, and ICU length of stay as outcome measure for the total Dutch ICU population and for subgroups of ICU admissions. We expressed associations through Pearson's correlation coefficients.
We included data of 59,809 ICU admissions from the NICE registry admitted in 2015. For the total ICU population we found no significant associations between the quality indicators. Between the standardized ICU length of stay ratio (SLOSR) and standardized in-hospital mortality ratio (SMR) we found a positive association for admissions with low-mortality risk (i.e. probability of mortality <0.3) and a negative association for admissions with high-mortality risk (i.e. probability of mortality >0.7). We recommended that multiple quality indicators should be used when judging or monitoring ICU quality of care and that the used quality indicators should be accessed across different subgroups of patients (e.g. patients with low and high risks of in-hospital mortality).

Chapter 6 addressed the reliability of a league table of Dutch ICUs. A league table ranks ICUs according to their values of a quality indicator and can be used to identify the worst and best performing ICUs which enables that the worst ICUs can learn from the best ICUs. Rankability expresses the percentage of variation between institutions due to unexplained differences between institutions and differences within institutions due to random variation [50]. Hence, the rankability of a league table should be as high as possible as this expresses the percentage of variation due to differences in quality of care. We examined whether the rankability of a league table could be improved by increasing the period on which the quality indicator is based or by grouping ICUs into clusters with similar performance on the quality indicator.
For this study we used the case-mix adjusted in-hospital mortality as quality indicator. Data of 157,394 ICU admissions in the period 2011 to 2013 from the NICE registry were included. The rankability was 73% for 2013 and 89% for the whole period 2011 to 2013. Rankability over the year 2013 increased until 98% when clustering ICUs.
We concluded that for a one-year period the rankability of a league table of Dutch ICUs based on case-mix adjusted in-hospital mortality was unacceptably low. We believe the clustering approach that we presented could be an useful alternative for registries such as NICE to identify under- and best-performing healthcare institutions. It may form a starting point for staff and directors from the lower clusters to improve clinical practice using information from the best performing clusters.

Chapter 7 provided a workflow-based guidance for statisticians on constructing funnel plots for the evaluation of binary quality indicators in healthcare institutions. The guidelines consist of the following steps: 1) defining policy level input; 2) checking the quality of models used for case-mix correction; 3) examining whether the number of observations per hospital is sufficient; 4) testing for overdispersion of the values of the indicator; 5) testing whether the quality indicator values are associated with organizational characteristics; and 6) specifying how the funnel plot should be constructed.
To assess internal usability of our guidelines, they were tested using data of ICU admissions in 2014 from the NICE registry. Our results showed that it was appropriate to develop funnel plots for case-mix adjusted in-hospital mortality for all ICU admissions, but not for subgroups based on admission type. For these subgroups the number of admissions per ICU was too small (step 3 of the guidelines) or the severity of illness expressed as the expected probability of mortality was associated with case-mix adjusted in hospital mortality (step 5 of the guidelines). We expect that our guidelines will help to strive for consistency in funnel plot construction over projects, employees, and time and are useful for data analysts and registry employees preparing funnel plots. This is particularly true if these people and organizations wish to use standard operating procedures when constructing funnel plots.

Samenvatting

Het bewustzijn voor de kwaliteit van zorg is bij belanghebbenden gegroeid, gedreven door een motivatie tot continue kwaliteitsverbetering, de druk om resultaten te verantwoorden en budgettaire beperkingen. Zorginstellingen vergelijken kwaliteitsindicatoren met de eigen historie of met andere zorginstellingen om ruimte voor verbetering te kunnen identificeren.
De door intensive care afdelingen (IC's) geleverde zorg is complex en vindt plaats in een technische en arbeidsintensieve omgeving. De hoge kosten voor IC-zorg leiden ertoe dat een groot deel van het budget voor de gezondheidszorg wordt uitgegeven aan IC's [12]. Dit maakt de IC een interessant onderdeel van het ziekenhuis om efficiëntie en effectiviteit van de geleverde zorg te monitoren, te vergelijken en te verbeteren.
De waarden van kwaliteitsindicatoren vormen idealiter een afspiegeling van de werkelijke zorgkwaliteit van een zorginstelling en representeren ruimte voor verbetering in zorgkwaliteit. Echter, verschillen kunnen ook veroorzaakt worden door ruis als gevolg van registratieverschillen, geen of onvolledige correctie voor verschillen in patiëntkarakteristieken en door toeval. De waarden van de kwaliteitsindicatoren kunnen door deze ruis worden beïnvloedt en kunnen leiden tot onjuiste beoordelingen van de zorgkwaliteit van deze instellingen. Een correctie voor verschillen in patiëntkarakteristieken tussen de opgenomen IC-patiënten is nodig voor een eerlijke en zinvolle vergelijking tussen IC's. Hiervoor kunnen prognostische modellen worden gebruikt.
De kosten van IC-opnamen zijn sterk geassocieerd met de behandelduur op de IC [23, 24]. Deze kan dan ook een belangrijke rol spelen bij onderzoek naar de efficiëntie van de geleverde IC-zorg. De IC-patiënten vormen een heterogene populatie met een breed scala aan complexe gezondheidsproblemen waarvan de behandelduur verschilt [25]. Onder experts bestaat weinig consensus met betrekking tot de beste methode om de IC-behandelduur te voorspellen. [26–29]. Het eerste deel van dit proefschrift, hoofdstuk 2 tot en met 4, behandelde de ontwikkeling en het voorspellend vermogen van prognostische modellen voor IC-behandelduur.
Het tweede deel van dit proefschrift behandelde de presentatie van de waarden van kwaliteitsindicatoren. Een enkele kwaliteitsindicator kan niet alle aspecten van zorgkwaliteit en efficiëntie zuiver meten. Vaak wordt gebruik gemaakt van een set kwaliteitsindicatoren om ruimte voor verbetering in zorgkwaliteit te kunnen identificeren en hiermee beleidsbeslissingen te ondersteunen [9]. In dit proefschrift hebben we de associatie tussen de kwaliteitsindicatoren voor ziekenhuissterfte; heropname op de IC binnen 48 uur na IC-ontslag; en IC-behandelduur als uitkomstmaat besproken, in hoofdstuk 5. Tevens werden in dit proefschrift ranglijsten en funnel plots geïntroduceerd als methoden om de waarden van kwaliteitsindicatoren grafisch weer te geven, in respectievelijk hoofdstuk 6 en 7.

Deel I: Prognostische modellen voor intensive care behandelduur

Hoofdstuk 2 beschreef een systematische literatuurstudie naar de ontwikkeling en validatie van prognostische modellen om de IC-behandelduur te voorspellen. In totaal zijn 11 studies geïncludeerd. Voor de beoordeling van de modellen zijn vier eisen gedefinïeerd: 1) de model coëfficiënten zijn gepubliceerd; 2) in het model zijn geen organisatorische karakteristieken opgenomen; 3) de gerapporteerde kalibratie van het model is matig tot zeer goed (gedefinieerd volgens door Calster et al. [62]); 4) de gerapporteerde nauwkeurigheid van de voorspelde waarden uitgedrukt als percentage verklaarde variantie (R^2) is kleiner dan 36%.
Op patiënt niveau varieerde het percentage verklaarde variantie tussen de 5% en de 28%. Op IC-niveau varieerde het percentage verklaarde variantie tussen de 1% en de 64%. Twee studies voldeden aan de eisen met betrekking tot de nauwkeurigheid op IC-niveau [26, 28]. Door deze studies is voor verschillende subgroepen van IC-patiënten de model kalibratie gerapporteerd, maar er werd niet voldaan aan onze eisen. Geen van de modellen voldeed aan onze eisen voor planningsdoeleinden, identificatie van een onverwacht lange behandelduur of het vergelijken van IC's. Bij gebruik van deze prognostische modellen voor deze doeleinden is dus voorzichtigheid geboden.

Hoofdstuk 3 vergeleek acht regressiemethoden voor het voorspellen van de IC-behandelduur, waarbij alleen gecorrigeerd werd voor patiëntkarakteristieken. De modellen zijn geëvalueerd door middel van de bootstrap methode waarbij de volgende maten zijn gebruikt: het percentage verklaarde variantie (R^2), de gemiddelde kwadratische fout, de gemiddelde absolute fout en de bias.
De gegevens van 32.667 ongeplande IC-opnamen over het jaar 2011 en afkomstig van de Nationale Intensive Care Evaluation (NICE) registratie zijn geïncludeerd. De prestaties van de door ons ontwikkelde modellen waren teleurstellend. Het percentage verklaarde variantie was maximaal 20% op patiëntniveau. De modellen hadden een gemiddelde kwadratische fout van meer dan zeven dagen en in absolute termen was de afwijking, afhankelijk van de geobserveerde behandelduur, gemiddeld drie dagen. De gevonden verschillen tussen de regressiemethoden waren over het algemeen klein. We concludeerden dat het moeilijk is om de IC-behandelduur te voorspellen voor ongeplande IC-opnamen, waarbij alleen gecorrigeerd werd voor patiëntkarakteristieken ten tijde van IC-opname. IC-ontslag is mogelijk niet alleen afhankelijk van het herstel van de patiënt, maar ook van organisatorische omstandigheden zoals de beschikbaarheid van bedden op de verpleegafdeling en de noodzaak om bedden voor andere patiënten vrij te houden.

Hoofdstuk 4 beschreef de associatie tussen IC organisatorische kenmerken en de IC-behandelduur na correctie voor patiëntkarakteristieken. Tevens is het voorspellend vermogen van een prognostisch model waarbij alleen gecorrigeerd wordt voor patiëntkarakteristieken vergeleken met een model dat daarnaast ook corrigeerd voor organisatorische kenmerken.
Voor het bepalen van associaties is gebruik gemaakt van mixed effect regressie modellen, waarbij is gecorrigeerd voor patiëntkarakteristieken en voor één organisatorisch kenmerk in elk model. Als random intercept is de IC-afdeling opgenomen. Het voorspellend vermogen van de organisatorische IC-kenmerk is getoetst door de residuen te vergelijken.
We hebben de gegevens van 78,822 IC-opnamen over de jaren 2014 en 2015, afkomstig van de NICE registratie geïncludeerd. We vonden een significante associatie met IC-behandelduur voor de volgende organisatorische kenmerken: aantal ziekenhuis bedden; aantal IC-bedden; aanwezigheid van fellows in opleiding tot intensivist; full-time equivalent IC-verpleegkundigen; verpleegkundige-patiënt ratio; en ontslagen in een shift met 100% bed bezetting. Deze organisatorische IC-kenmerken verbeterden ons prognostisch model voor het voorspellen van de IC-behandelduur nauwelijks.

Ter conclusie zetten we de resultaten van het eerste deel van dit proefschrift in context van de doelstellingen voor het gebruik van prognostische modellen. Voor het plannen van bed en personele capaciteit en de identificatie van patiënten met een onverwacht lange IC-behandelduur moet de voorspelde IC-behandelduur betrouwbaar zijn op patiëntniveau. We hebben in de hoofdstukken 2 en 3 aangetoond dat de nauwkeurigheid van de modellen op patiïtniveau niet voldoende is en de modellen daardoor ongeschikt zijn voor deze doelstellingen.
Ons werk heeft discussie opgeleverd in de wetenschappelijke literatuur. Straney en co-auteurs [105] beargumenteren dat slechte model prestaties op patiënt niveau niet altijd indicatief zijn voor de prestaties voor het vergelijken van IC's. In overeenstemming geeft Kramer [187] aan dat sommige van de door ons in de review, chapter 2 ,geïncludeerde modellen wel geschikt zijn voor het vergelijken van IC's.
We zijn het eens dat voor het vergelijken van IC's het model betrouwbare voorspellingen dient te doen op IC-niveau. Echter, als een model niet accuraat voorspelt op paiënt niveau, kan een significante residuele variatie over blijven. Wij blijven voorzichtig met het aanbevelen van een dergelijk model voor het vergelijken van IC's in de praktijk. Naar aanleiding van deze discussie vonden we voor ons eigen model, beschreven in hoofdstuk 5, een R^2 van 64% op IC-niveau. De kalibratie curve gaf bevredigende resultaten. We adviseren nader onderzoek naar de nauwkeurigheid en kalibratie voor verschillende patiëntgroepen.

Deel II: Presentatie van waarde van kwaliteits-indicatoren.

Hoofdstuk 5 behandelde de associatie tussen verschillende voor patiëntkarakteristieken gecorrigeerde kwaliteitsindicatoren voor ziekenhuissterfte, heropname op de IC binnen 48 uur na IC-ontslag en IC-behandelduur als uitkomstmaten. Dit is gedaan voor zowel de totale Nederlandse IC-populatie als voor subgroepen IC-patiënten. De associatie is berekend met behulp van de Pearson's correlatie coëfficiënten.
Voor dit onderzoek zijn 59.809 IC-opnamen over het jaar 2015, afkomstig van de NICE registratie geïncludeerd. Voor de totale IC-populatie hebben we geen significante associatie gevonden tussen de kwaliteitsindicatoren. Voor IC-patiënten met een met een lage sterftekans (sterftekans kleiner dan 0.3) werd een positieve associatie gevonden en voor IC-patië nten met een hoge sterftekans (sterftekans groter dan 0.7) werd een negatieve associatie gevonden. Bij het beoordelen en monitoren van de kwaliteit van IC-zorg kan men het best gebruik maken van verschillende kwaliteitsindicatoren, aangezien verschillende indicatoren verschillende aspecten van IC-zorg reflecteren. Daarnaast is het van belang om de indicatoren voor zowel de totale IC populatie als voor een aantal patiënt subgroepen te beschouwen.

Hoofdstuk 6 behandelde de betrouwbaarheid van een ranglijst van Nederlandse IC's gebaseerd op voor patiëntkarakteristieken gecorrigeerde ziekenhuissterfte. De betrouwbaarheid wordt vastgesteld met behulp van de rankability, gedefinieerd als het percentage van de variatie dat verklaard wordt door kwaliteitsverschillen tussen IC's en niet slechts door random variatie [50]. Aanvullend hebben we onderzocht of de betrouwbaarheid van de ranglijst verbeterd door het verlengen van de verslagperiode en/of door de IC's samen te voegen in clusters van IC's met een vergelijkbare kwaliteit van zorg.
Voor dit onderzoek zijn gegevens van 157.394 IC-opnamen over de jaren 2011 en 2013, afkomstig van de NICE registratie geincludieerd. De rankability van de ranglijst van Nederlandse IC's op basis van de voor patiëntkarakteristieken gecorrigeerde ziekenhuissterfte was 73% indien alleen het jaar 2013 werd gebruikt, dit is onaanvaardbaar laag. De rankability van de ranglijst verbeterde naar 89% door gegevens van 2011 tot en met 2013 te gebruiken. Wanneer IC's werden samengevoegd tot clusters steeg de rankability naar 98% indien alleen het jaar 2013 werd gebruikt.
Wij concluderen dat het samenvoegen van IC's tot clusters van IC's een zinvol alternatief kan zijn voor registraties zoals de NICE om zo betrouwbaar de onderen best presterende zorginstellingen te kunnen identificeren. Dit kan een startpunt vormen voor medewerkers en bestuurders om de klinische praktijk te verbeteren door gebruik te maken van zorgproces informatie van ICŠs uit het best presterende cluster.

Hoofdstuk 7 beschreef een richtlijn voor statistici om een funnel plots te construeren voor binaire uitkomstmaten. Deze richtlijn bestond uit de volgende stappen: 1) het definiëren van beleidsinput; 2) het controleren van de kwaliteit van de prognostische modellen die gebruikt worden voor case-mix correctie; 3) nagaan of het aantal waarnemingen per ziekenhuis voldoende is; 4) testen of de waarden van de kwaliteitsindicator onderhevig zijn aan overdispersie; 5) testen of de waarden van de kwaliteitsindicator geassocieerd zijn met organisatorische IC-kenmerken; en 6) specificeren hoe de funnel plot weergegeven wordt.
We hebben de beschreven richtlijn intern gevalideerd met behulp van data uit de NICE registratie. We hebben voor deze validatie 87.049 IC-opnamen over het jaar 2014 geïncludeerd. We concludeerden dat funnel plots geschikt zijn om IC's met elkaar te vergelijken. Voor subgroepen op basis van opnametype was dit echter niet het geval. Dit kwam voor patiënten met een medisch opnametype door de associatie tussen de SMR en de gemiddelde voorspelde kans op sterfte van een IC (stap 5 van de richtlijn). Voor patiënten opgenomen na spoed- of geplande chirurgie was het aantal opnamen per IC over het algemeen niet voldoende (stap 3 van de richtlijn).
We verwachten dat onze richtlijn nuttig zal zijn voor data analisten en registratie medewerkers die funnel plots willen presenteren. Deze richtlijn zal ook helpen bij het streven naar consistentie in funnel plot constructie over verschillende projecten, werknemers en over tijd.

Overview of cited literature

[1] J. Zimmerman, C. Alzola, and K. von Rueden. The use of benchmarking to identify top performing critical care units: a preliminary assessment of their policies and practices. *J Crit Care* 2003: 18(2), 76–86.

[2] D. Northcott and S. Llewellyn. Benchmarking in UK health: a gap between policy and practice? *BIJ* 2005: 12(5), 419–435.

[3] M. Cole. Benchmarking: contemporary modalities and applications. *EJA* 2011: 11(2), 42–48.

[4] E. van Veen-Berkx, D. de Korne, O. Olivier, et al. Benchmarking operating room departments in the Netherlands. *BIJ* 2016: 23(5), 1171–1192.

[5] H. Lingsma, B. Roozenbeek, B. Li, et al. Large between-center differences in outcome after moderate and severe traumatic brain injury in the international mission on prognosis and clinical trial design in traumatic brain injury (IMPACT) study. *Neurosurgery* 2009: 68, 601–608.

[6] A. Koetsier, N. Peek, and N. de Keizer. Identifying types and causes of errors in mortality data in a clinical registry using multiple information systems. *Stud Health Technol Inform* 2012: 180, 771–775.

[7] C. Pannucci and E. Wilkins. Identifying and avoiding bias in research. *Plast Reconstr Surg* 2010: 126(2), 619–625.

[8] F. Song, L. Hooper, and Y. Loke. Publication bias: what is it? How do we measure it? How do we avoid it? *Open Access J Clin Trials* 2013: 5(1), 51–81.

[9] A. Donabedian. The quality of care. How can it be assessed? *JAMA* 1988: 260(12), 1743–1748.

[10] A. Bottle and P. Aylin. *Statistical methods for healthcare performance monitoring*. United Kingdom: Chapman and Hall, 2016. ISBN: 978-1-4822-4609-4.

[11] G. Smith and M. Nielsen. ABC of intensive care: organisation of intensive care. *BMJ* 1999: 318(7197), 1468–1470.

[12] N. Halpern and S. Pastores. Critical care medicine in the United States 2000-2005: an analysis of bed numbers, occupancy rates, payer mix, and costs. *Crit Care Med* 2010: 38(1), 65–71.

[13] N. van de Klundert, R. Holman, D. Dongelmans, et al. Data resource profile: the Dutch national intensive care evaluation (NICE) registry of admissions to adult intensive care units. *Int J Epidemiol* 2015: 44(6), 1850–1850h.

[14] Dutch National Intensive Care Evaluation (NICE) foundation. Web Page. 2014. URL: http://www.stichting-nice.nl.

[15] D. Arts, N. de Keizer, G. Scheffer, et al. Quality of data collected for severity of illness scores in the Dutch National Intensive Care Evaluation (NICE) registry. *Intensive Care Med* 2002: 28(5), 656–659.

[16] J. Zimmerman, A. Kramer, D. McNair, et al. Acute physiology and chronic health evaluation (APACHE) IV: hospital mortality assessment for today's critically ill patients. *Crit Care Med* 2006: 34(5), 1297–1310.

[17] Australian and New Zealand Intensive Care Society. Web Page. 2014. URL: www.anzics.com.au.

[18] Scottisch Intensive Care Society. Web Page. 2015. URL: www.scottishintensivecare.org.uk.

[19] Dutch National Intensive Care Evaluation (NICE) foundation. *Data in beeld*. Web Page. 2014. URL: http://www.stichting-nice.nl/datainbeeld/public.

[20] A. Koetsier, S. van der Veer, K. Jager, et al. Control charts in healthcare quality improvement. A systematic review on adherence to methodological criteria. *Methods Inf Med* 2012: 51(3), 189–198.

[21] S. Duckett, M. Coory, and K. Sketcher-Baker. Identifying variations in quality of care in Queensland hospitals. *MJA* 2007: 187(10), 571–575.

[22] A. Koetsier, N. de Keizer, and N. Peek. A comparison of internal versus external risk-adjustment for monitoring clinical outcomes. *Stud Health Technol Inform* 2011: 169, 180–184.

[23] N. Halpern, S. Pastores, and R. Greenstein. Critical care medicine in the United States 1985-2000: an analysis of bed numbers, use, and costs. *Crit Care Med* 2004: 32(6), 1254–1259.

[24] J. Kahn, G. Rubenfeld, J. Rohrbach, et al. Cost savings attributable to reductions in intensive care unit length of stay for mechanically ventilated patients. *Med Care* 2008: 46(12), 1226–1233.

[25] J. Moran and P. Solomon. A review of statistical estimators for risk-adjusted length of stay: analysis of the Australian and New Zealand intensive care adult patient data-base, 2008-2009. *BMC Med Res Methodol* 2012: 12, 68–85.

[26] E. Vasilevskis, M. Kuzniewicz, B. Cason, et al. Mortality probability model III and simplified acute physiology score II: assessing their value in predicting length of stay and comparison to APACHE IV. *Chest* 2009: 136(1), 89–101.

[27] R. Becker, J. Zimmerman, W. Knaus, et al. The use of APACHE III to evaluate ICU length of stay, resource use, and mortality after coronary artery by-pass surgery. *J Cardiovasc Surg (Torino)* 1995: 36(1), 1–11.

[28] J. Zimmerman, A. Kramer, D. McNair, et al. Intensive care unit length of stay: Benchmarking based on Acute Physiology and Chronic Health Evaluation (APACHE) IV. *Crit Care Med* 2006: 34(10), 2517–2529.

[29] A. Woods, F. MacKirdy, B. Livingston, et al. Evaluation of predicted and actual length of stay in 22 Scottish intensive care units using the APACHE III system. Acute physiology and chronic health evaluation. *Anaesthesia* 2000: 55(11), 1058–1065.

[30] K. Strand, S. Walther, M. Reinikainen, et al. Variations in the length of stay of intensive care unit nonsurvivors in three scandinavian countries. *Crit Care* 2010: 14(5), R175.

[31] E. Simchen, C. Sprung, N. Galai, et al. Survival of critically ill patients hospitalized in and out of intensive care units under paucity of intensive care unit beds. *Crid Care Med* 2004: 32(8), 1654–1661.

[32] H. Rothen, K. Stricker, J. Einfalt, et al. Variability in outcome and resource use in intensive care units. *Intensive Care Med* 2007: 33(8).

[33] M. Prin and H. Wunsch. The role of stepdown beds in hospital care. *Am J Respir Crit Care Med* 2014: 190(11), 1210–1216.

[34] T. Williams and G. Leslie. Delayed discharges from an adult intensive care unit. *Aust Health Rev* 2004: 28(1), 87–96.

[35] P. Marik and L. Hedman. What's in a day? Determining intensive care unit length of stay. *Crit Care Med* 2008: 28(6), 2090–2093.

[36] J. Rapoport, D. Teres, Y. Zhao, et al. Length of stay data as a guide to hospital economic performance for ICU patients. *Med Care* 2003: 41(3), 386–397.

[37] A. Rosenberg, J. Zimmerman, C. Alzola, et al. Intensive care unit length of stay: recent changes and future challenges. *Crit Care Med* 2000: 28(10), 3465–3473.

[38] P. Metnitz, F. Fieux, B. Jordan, et al. Critically ill patients readmitted to intensive care units–lessons to learn? *Intensive Care Med* 2003: 29(2), 241–248.

[39] G. Cooper, C. Sirio, A. Rotondi, et al. Are readmissions to the intensive care unit a useful measure of hospital performance? *Med Care* 1999: 37(4), 399–408.

[40] A. Kramer, T. Higgins, and J. Zimmerman. Intensive care unit readmissions in U.S. hospitals: patient characteristics, risk factors, and outcomes. *Crit Care Med* 2012: 40(1), 3–10.

[41] N. Kolfschoten, J. Kievit, G. Gooiker, et al. Focusing on desired outcomes of care after colon cancer resections; hospital variations in textbook outcome. *Eur J Surg Oncol* 2013: 39(2), 156–163.

[42] P. Kaboli, J. Go, J. Hockenberry, et al. Associations between reduced hospital length of stay and 30-day readmission rate and mortality: 14-year experience in 129 Veterans Affairs hospitals. *Ann Intern Med* 2012: 157(12), 837–845.

[43] L. Horwitz, Y. Wang, M. Desai, et al. Correlations among risk-standardized mortality rates and among risk-standardized readmission rates within hospitals. *J Hosp Med* 2012: 7(9), 690–696.

[44] M. Shwartz, A. Cohen, J. Restuccia, et al. How well can we identify the high-performing hospital? *Med Care Res Rev* 2011: 68(3), 290–310.

[45] R. Gibberd, S. Hancock, P. Howley, et al. Using indicators to quantify the potential to improve the quality of health care. *Int J Qual Health Care* 2004: 16 Suppl 1(3), 37–43.

[46] O. Lemmers, M. Broeders, A. Verbeek, et al. League tables of breast cancer screening units: worst-case and best-case scenario ratings helped in exposing real differences between performance ratings. *J Med Screen* 2009: 16(2), 67–72.

[47] H. Goldstein and D. Spiegelhalter. League Tables and Their Limitations: Statiatical Issues in Comparisons of Institutional Performance. *J R Stat Soc* 1996: 159(3), 385–443.

[48] S. Siregar, R. Groenwold, E. Jansen, et al. Limitations of ranking lists based on cardiac surgery mortality rates. *Circ Cardiovasc Qual Outcomes* 2012: 5(3), 403–409.

[49] H. van Houwelingen and R. Brand. Empirical Bayes methods for monitoring health care quality. 2005. URL: http://www.stat.fi/isi99/proceedings/arkisto/varasto/vanh0228.pdf.

[50] A. van Dishoeck, H. Lingsma, J. Mackenbach, et al. Random variation and rankability of hospitals using outcome indicators. *BMJ Qual Saf* 2011: 20(10), 869–874.

[51] D. Spiegelhalter. Funnel plots for comparing institutional performance. *Stat Med* 2005: 24(8), 1185–1202.

[52] T. Rakow, R. Wright, D. Spiegelhalter, et al. The pros and cons of funnel plots as an aid to risk communication and patient decision making. *Br J Psychol* 2015: 106(2), 327–348.

[53] E. Mayer, A. Bottle, C. Rao, et al. Funnel plots and their emerging application in surgery. *Ann Surg* 2009: 249(3), 376–383.

[54] N. Halpern and S. Pastores. Critical care medicine beds, use, occupancy, and costs in the United States: a methodological review. *Crit Care Med* 2015: 43(11), 2452–2459.

[55] M. Niskanen, M. Reinikainen, and V. Pettila. Case-mix-adjusted length of stay and mortality in 23 Finnish ICUs. *Intensive Care Med* 2009: 35(6), 1060–1067.

[56] I. Verburg, N. de Keizer, E. de Jonge, et al. Comparison of regression methods for modeling intensive care length of stay. *PLoS One* 2014: 9(10), e109684.

[57] K. Moons, J. de Groot, W. Bouwmeester, et al. Critical appraisal and data extraction for systematic reviews of prediction modelling studies: the CHARMS checklist. *PLoS Med* 2014: 11(10), e1001744.

[58] E. Wallace, M. Uijen, B. Clyne, et al. Impact analysis studies of clinical prediction rules relevant to primary care: a systematic review. *BMJ open* 2016: 6(3), e009957.

[59] M. Walsh, N. Horgan, C. Walsh, et al. Systematic review of risk prediction models for falls after stroke. *J Epidemiol Community Health* 2016: 70(5), 513–519.

[60] F. Kunath, A. Spek, K. Jensen, et al. Prognostic factors for tumor recurrence in patients with clinical stage I seminoma undergoing surveillance-protocol for a systematic review. *Syst Rev* 2015: 4, 182–189.

[61] S. Medlock, A. Ravelli, P. Tamminga, et al. Prediction of mortality in very premature infants: a systematic review of prediction models. *PLoS One* 2011: 6(9), e23441.

[62] B. van Calster, D. Nieboer, Y. Vergouwe, et al. A calibration hierarchy for risk models was defined: from utopia to emperical data. *J Clin Epidemiol* 2016: 74, 167–176.

[63] K. Divaris, W. Vann, A. Baker, et al. Examining the accuracy of caregivers' assessments of young children's oral health status. *J Am Dent Assoc* 2012: 143(11), 1237–1247.

[64] W. Knaus, D. Wagner, J. Zimmerman, et al. Variations in mortality and length of stay in intensive care units. *Ann Intern Med* 1993: 118(10), 753–761.

[65] A. Perez, W. Chan, and R. Dennis. Predicting the length of stay of patients admitted for intensive care using a first step analysis. *Health Serv Outcome Res Meth* 2006: 6(3-4), 127–138.

[66] J. Moran, P. Bristow, P. Solomon, et al. Mortality and length-of-stay outcomes, 1993-2003, in the binational Australian and New Zealand intensive care adult patient database. *Crit Care Med* 2008: 36(1), 46–61.

[67] A. Kramer and J. Zimmerman. A predictive model for the early identification of patients at risk for a prolonged intensive care unit length of stay. *BMC Med Inform Decis Mak* 2010: 10, 27–43.

[68] G. Clermont, V. Kaplan, R. Moreno, et al. Dynamic microsimulation to model multiple outcomes in cohort of critically ill patients. *Intensive Care Med* 2004: 30(12), 2237–2244.

[69] M. Al Tehewy, M. El Houssinie, N. El Ezz, et al. Developing severity adjusted quality measures for intensive care units. *Int J Health Care Qual Assur* 2010: 23(3), 277–286.

[70] Public Health Observatories (PHOs). Web Page. 2013. URL: http://www.apho.org.uk/resource.

[71] E. Steyerberg. Evaluation of performance. *Clinical prediction models: a practical approach to development, validation, and updating.* USA: Springer, 2009. Chap. 5, 83–99. ISBN: 978-0-387-77244-8.

[72] S. Mallett, P. Royston, S. Dutton, et al. Reporting methods in studies developing prognostic models in cancer: a review 2010: 8, 20.

[73] S. Mallett, P. Royston, R. Waters, et al. Reporting performance of prognostic models in cancer: a review. *BMC Med* 2010: 8, 21.

[74] S. Lemeshow, D. Teres, J. Avrunin, et al. Refining intensive care unit outcome prediction by using changing probabilities of mortality. *Crit Care Med* 1988: 16(5), 470–477.

[75] J. Le Gall, S. Lemeshow, and F. Saulnier. A new Simplified Acute Physiology Score (SAPS II) based on a European/North American multicenter study. *JAMA* 1993: 270(24), 2957–2963.

[76] N. Messaoudi, J. de Cocker, B. Stockman, et al. Prediction of prolonged length of stay in the intensive care unit after cardiac surgery: the need for a multi-institutional risk scoring system. *J Card Surg* 2009: 24(2), 127–133.

[77] P. Austin and E. Steyerberg. The number of subjects per variable required in linear regression analyses. *J Clin Epidemiol* 2015: 68(6), 627–636.

[78] M. Babyak. What you see may not be what you get: a brief, nontechtnical introduction to overfitting in regression-type models. *Psychosom Med* 2004: 66(3), 411–421.

[79] K. Kelley and E. Maxwell. Sample size for multiple regression: obtaining regression coefficients that are accurate, not simply significant. *Psychol Methods* 2003: 8(3), 305–321.

[80] P. Peduzzi, J. Concato, A. Feinstein, et al. Importance of events per independent variable in proportional hazards regression analysis. II. Accuracy and precision of regression estimates. *J Clin Epidemiol* 1995: 48(12), 1503–1510.

[81] F. Harrell, K. Lee, and D. Mark. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med* 1996: 15(4), 361–387.

[82] R. Henderson, P. Diggle, and A. Dobson. Joint modelling of longitudinal measurements event time data. *Biostatistics* 2000: 1(4), 465–480.

[83] M. Wolbers, M. Koller, J. Witteman, et al. Prognostic models with competing risks: methods and application to coronary risk prediction. *Epidemiology* 2009: 20(4), 555–561.

[84] Z. Zhou. United Kingdom: Chapman and Hall, 2012. ISBN: 978-1-439-83003-1.

[85] J. Shawe-Taylor and N Cristiano. *Kernel methods for pattern analysis.* United Kingdom: Cambridge University Press, 2004. ISBN: 978-0-521-81397-6.

[86] R. Koenker and G. Bassett. Regression Quantiles. *Econometrica* 1978: 46(1), 33–50.

[87] W. Knaus, E. Draper, D. Wagner, et al. APACHE II: a severity of disease classification system. *Crit Care Med* 1985: 13(10), 818–829.

[88] J. Zimmerman and A. Kramer. Outcome prediction in critical care: the acute physiology and chronic health evaluation models. *Curr Opin Crit Care* 2008: 14(5), 491–497.

[89] B. Nathanson, T. Higgins, D. Teres, et al. A revised method to assess intensive care unit clinical performance and resource utilization. *Crit Care Med* 2007: 35(8), 1853–1862.

[90] W. Manning and J. Mullahy. Estimating log models: to transform or not to transform? *J Health Econ* 2001: 20(4), 461–494.

[91] P. Austin, D. Rothwell, and J. Tu. A comparison of statistical modeling strategies for analyzing length of stay after CABG surgery. *Health Serv Outcomes Res Methodol* 2002: 3(2), 107–133.

[92] A. Stolwijk, H. Straatman, and G. Zielhuis. Studying seasonality by using sine and cosine functions in regression analysis. *J Epidemiol Community Health* 1999: 53(4), 235–238.

[93] Dutch National Intensive Care Evaluation (NICE) foundation. *Anual Journal: focus IC*. 2013. URL: http://www.stichting-nice.nl/doc/jaarboek-2013-web.pdf.

[94] S. Brinkman, F. Bakhshi-Raiez, A. Abu-Hanna, et al. External validation of acute physiology and chronic health evaluation IV in Dutch intensive care units and comparison with acute physiology and chronic health evaluation II and simplified acute physiology score II. *J Crit Care* 2011: 26(1), e11–18.

[95] T. Hastie, R. Tibshirani, and J. Friedman. *The elements of statistical learning: data mining, inference, and prediction.* USA: Springer, 2001, 115–163. ISBN: 978-0-387-84858-7.

[96] E. Steyerberg. Overfitting and optimism in prediction models. *Clinical prediction models: a practical approach to development, validation, and updating.* USA: Springer, 2009, 175–187. ISBN: 978-0-387-77244-8.

[97] Development Core Team, R. *A Language and Environment for Statistical Computing*. Web Page. 2005.

[98] R. Becker and J. Zimmerman. ICU scoring systems allow prediction of patient outcomes and comparison of ICU performance. *Crit Care Clin* 1996: 12(3), 503–514.

[99] Y. Widyastuti, R. Stenseth, A. Wahba, et al. Length of intensive care unit stay following cardiac surgery: is it impossible to find a universal prediction model? *Interact Cardiovasc Thorac Surg* 2012: 15(5), 825–832.

[100] A. Lee, K. Wang, K. Yau, et al. Maternity length of stay modelling by gamma mixture regression with random effects. *Biom J* 2007: 49(5), 750–764.

[101] M. Faddy, N. Graves, and A. Pettitt. Modeling length of stay in hospital and other right skewed data: comparison of phase-type, gamma and log-normal distributions. *Value Health* 2009: 12(2), 309–314.

[102] L. Straney, A. Clements, J. Alexander, et al. Quantifying variation of paediatric length of stay among intensive care units in Australia and New Zealand. *Qual Saf Health Care* 2010: 19(6), e5.

[103] N. Halpern, S. Pastores, H. Thaler, et al. Critical care medicine use and cost among Medicare beneficiaries 1995-2000: major discrepancies between two United States federal Medicare databases. *Crit Care Med* 2007: 35(3), 692–699.

[104] I. Verburg, A. Atashi, S. Eslami, et al. Which models can I use to predict adult ICU length of stay? A systematic review. *Crit Care Med* 2017: 45(2), e222–231.

[105] L. Straney, A. Udy, A. Burrell, et al. Modelling risk-adjusted variation in length of stay among Australian and New Zealand ICUs. *Plos one* 2017: 12(5), e0176570.

[106] S. Walther and U. Jonasson. Outcome of the elderly critically ill after intensive care in an era of cost containment. *Acta Anaesthesiol Scand* 2004: 48(4), 417–422.

[107] J. Adamski, R. Goraj, D. Onichimowski, et al. The differences between two selected intensive care units located in central and northern Europe - preliminary observation. *Anaesthesiol Intensive Ther* 2015: 47(2), 117–124.

[108] S. Dara and B. Afessa. Intensivist-to-bed ratio: association with outcomes in the medical ICU. *Chest* 2005: 128(2), 567–572.

[109] D. Gruenberg, W. Shelton, S. Rose, et al. Factors influencing length of stay in the intensive care unit. *Am J Crit Care* 2006: 15(5), 502–509.

[110] A. Oliveira, O. Dias, M. Mello, et al. Factors associated with increased mortality and prolonged length of stay in an adult intensive care unit. *Rev Bras Ter Intensiva* 2010: 22(3), 250–256.

[111] T. Higgins, W. McGee, J. Steingrub, et al. Early indicators of prolonged intensive care unit stay: impact of illness severity, physician staffing, and pre-intensive care unit length of stay. *Crit Care Med* 2003: 31(1), 45–51.

[112] E. Asano, I. Rasera, and E. Shiraga. Cross-sectional study of variables associated with length of stay and ICU need in open Roux-En-Y gastric bypass surgery for morbid obese patients: an exploratory analysis based on the Public Health System administrative database (Datasus) in Brazil. *Obes Surg* 2012: 22(12), 1810–1817.

[113] G. Rosenthal, D. Harper, L. Quinn, et al. Severity-adjusted mortality and length of stay in teaching and nonteaching hospitals. Results of a regional study. *JAMA* 1997: 278(6), 485–490.

[114] F. Mallor, C. Azcárate, and J. Barado. Control problems and management policies in health systems: application to intensive care units. *Flex Serv Manuf J* 2016: 28(1-2), 62–89.

[115] I. Verburg, N. de Keizer, R. Holman, et al. Individual and Clustered Rankability of ICUs According to Case-Mix-Adjusted Mortality. *Crit Care Med* 2015: 44(5), 901–909.

[116] H. Brown and R. Prescott. *Applied mixed models in medicine*. United Kingdom: Wiley, 2015. ISBN: 978-1-118-77825-8.

[117] S. Nakagawa and H. Schielzeth. A general and simple method for obtaining R^2 from generalized linear mixed-effect models. *Methods Ecol Evol* 2013: 4(2), 133–142.

[118] D. Bates, M. Maechler, B. Bolker, et al. *Package lme4*. Web Page. 2016. URL: https://github.com/lme4/lme4/http://lme4.r-forge.r-project.org/.

[119] F. Harell. *Package rms*. Web Page. 2017. URL: https://cran.r-project.org/web/packages/rms/rms.pdf.

[120] A. Peets, P. Boiteau, and C. Doig. Effect of critical care medicine fellows on patient outcome in the intensive care unit. *Acad Med* 2006: 81(10 Suppl), S1–4.

[121] U. Ruttimann and M. Pollack. Variability in duration of stay in pediatric intensive care units: a multi institutional study. *J Pediatr* 1996: 128(1), 35–44.

[122] R. Medeiros, E. NeSmith, J. Heath, et al. Mid-level health providers' impact on ICU length of stay, patient satisfaction, mortality, and resource utilization. *J Trauma Nurs* 2011: 18(3), 149–153.

[123] J. Barado, J. Guergue, L. Esparza, et al. A mathematical model for simulating daily bed occupancy in an intensive care unit. *Crit Care Med* 2012: 40(4), 1098–2104.

[124] S. Logani, A. Green, and J. Gasperino. Benefits of high-intensity intensive care unit physician staffing under the affordable care act. *Crit Care Res Pract* 2011: 2011, 7.

[125] P. Pronovost, D. Angus, T. Dorman, et al. Physician staffing patterns and clinical outcomes in critically ill patients: a systematic review. *JAMA* 2002: 288(17), 2151–2162.

[126] D. Hackner, C. Shufelt, D. Balfe, et al. Do faculty intensivists have better outcomes when caring for patients directly in a closed ICU versus consulting in an open ICU? *Hosp Pract* 2009: 37(1), 40–50.

[127] A. Multz, D. Chalfin, I. Samson, et al. A closed medical intensive care unit (MICU) improves resource utilization when compared with an open MICU. *Am J Respir Crit Care Med* 1998: 157(5 Pt 1), 1468–1473.

[128] P. Pronovost, S. Berenholtz, T. Dorman, et al. Improving communication in the ICU using daily goals. *J Crit Care* 2003: 18(2), 71–75.

[129] S. Russell. Reducing readmissions to the intensive care unit. *Heart Lung* 1999: 28(5), 365–372.

[130] D. Beck, P. McQuillan, and G. Smith. Waiting for the break of dawn? The effects of discharge time, discharge TISS scores and discharge facility on hospital mortality after intensive care. *Intensive Care Med* 2002: 28(9), 1287–1293.

[131] C. Goldfrad and K. Rowan. Consequences of discharges from intensive care at night. *Lancet* 2000: 355(9210), 1138–1142.

[132] J. Benbassat and M. Taragin. Hospital readmissions as a measure of quality of health care: advantages and limitations. *Arch Intern Med* 2000: 160(8), 1074–1081.

[133] A. Rosenberg, T. Hofer, R. Hayward, et al. Who bounces back? Physiologic and other predictors of intensive care unit readmission. *Crit Care Med* 2001: 29(3), 511–518.

[134] A. Koetsier, N. Peek, E. de Jonge, et al. Reliability of in-hospital mortality as a quality indicator in clinical quality registries. A case study in an intensive care quality register. *Methods Inf Med* 2013: 52(5), 432–440.

[135] N. Van Sluisveld, F. Bakhshi-Raiez, N. De Keizer, et al. Variation in rates of ICU readmissions and post-ICU in-hospital mortality and their association with ICU discharge practices. *BMC Health Serv Res* 2017: 17.

[136] S. Brown, S. Ratcliffe, J. Kahn, et al. The epidemiology of intensive care unit readmissions in the United States. *Am J Respir Crit Care Med* 2012: 185(9), 955–964.

[137] Z. Hashmi, J. Dimick, D. Efron, et al. Reliability adjustment: a necessity for trauma center ranking and benchmarking. *J Trauma Acute Care Surg* 2013: 75(1), 166–172.

[138] J. Dimick, D. Staiger, and J. Birkmeyer. Ranking hospitals on surgical mortality: the importance of reliability adjustment. *Health Serv Res* 1986: 45(6), 1614–1629.

[139] E. Steyerberg, A. Vickers, N. Cook, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology* 2010: 21(1), 128–138.

[140] S. Davies. *Fitting Generalized Linear Models*. Web Page. 1992. URL: https://stat.ethz.ch/R-manual/R-devel/library/stats/html/glm.html.

[141] P. Savicky. *Test for association / correlation between paired samples*. Web Page. 2014. URL: https://cran.r-project.org/web/packages/pspearman/pspearman.pdf.

[142] A. Kramer, T. Higgins, and J. Zimmerman. The association between ICU readmission rate and patient outcomes. *Crit Care Med* 2013: 41(1), 24–33.

[143] M. Pouw, L. Peelen, K. Moons, et al. Including post-discharge mortality in calculation of hospital standardised mortality ratios: retrospective analysis of hospital episode statistics. *BMJ* 2013: 347, f5913.

[144] G. Mulvey, Y. Wang, Z. Lin, et al. Mortality and readmission for patients with heart failure among U.S. News and World Report's top heart hospitals. *Circ Cardiovasc Qual Outcomes* 2009: 2(6), 558–565.

[145] I. Verburg, R. Holman, N. Peek, et al. Guidelines on constructing funnel plots for quality indicators: A case study on mortality in intensive care unit patients. *Stat Methods Med Res* 2017, epub.

[146] S. Brinkman, A. Abu-Hanna, E. de Jonge, et al. Prediction of long-term mortality in ICU patients: model validation and assessing the effect of using in-hospital versus long-term mortality on benchmarking. *Intensive Care Med* 2013: 39(11), 1925–1931.

[147] A. van Dishoeck, M. Koek, E. Steyerberg, et al. Use of surgical-site infection rates to rank hospital performance across several types of surgery. *Br J Surg* 2013: 100(5), 628–637.

[148] D. Henneman, A. van Bommel, A. Snijders, et al. Ranking and rankability of hospital postoperative mortality rates in colorectal cancer surgery. *Ann Surg* 2014: 259(5), 844–849.

[149] F. Bakhshi-Raiez, N. Peek, R. Bosman, et al. The impact of different prognostic models and their customization on institutional comparison of intensive care units. *Crit Care Med* 2007: 35(11), 2553–2560.

[150] E. Marshall and D. Spiegelhalter. Reliability of league tables of in vitro fertilisation clinics: retrospective analysis of live birth rates. *BMJ* 1998: 316(7146), 1701–1705.

[151] K. Cios, W. Pedrycz, R. Swiniarsk, et al. *Data mining: a knowledge discovery approach*. USA: Springer science, 2007. ISBN: 978-0-387-33333-5.

[152] H. Lingsma, M. Eijkemans, and E. Steyerberg. Incorporating natural variation into IVF clinic league tables: The Expected Rank. *BMC Med Res Methodol* 2009: 9(1), 53–59.

[153] H. Lingsma, E. Steyerberg, M. Eijkemans, et al. Comparing and ranking hospitals based on outcome: results from The Netherlands Stroke Survey. *QJM* 2010: 103(2), 99–108.

[154] M. Coory, S. Duckett, and K. Sketcher-Baker. Using control charts to monitor quality of hospital care with administrative data. *Int J Qual Health Care* 2008: 20(1), 31–39.

[155] D. Dover and D. Schopflocher. Using funnel plots in public health surveillance. *Popul Health Metr* 2011: 9(1), 58–69.

[156] S. Few and K. Rowell. Variation and Its Discontents: Funnel Plots for Fair Comparisons. *Visual Business Intelligence Newsletter* 2013. URL: https://www.perceptualedge.com/articles/visual_business_intelligence/variation_and_its_discontents.pdf.

[157] S. Grant, A. Grayson, M. Jackson, et al. Does the choice of risk-adjustment model influence the outcome of surgeon-specific mortality analysis? A retrospective analysis of 14 637 patients under 31 surgeons. *BMJ* 2015: 16(I), i37–43.

[158] D. Griffen, C. Callahan, S. Markwell, et al. Application of statistical process control to physician-specific emergency department patient satisfaction scores: a novel use of the funnel plot. *Acad Emerg Med* 2012: 19(3), 348–355.

[159] D. Tighe, I. Sassoon, A. Kwok, et al. Is benchmarking possible in audit of early outcomes after operations for head and neck cancer? *Br J Oral Maxillofac Surg* 2014: 52(10), 913–921.

[160] Canadian Institute for Health Information. Web Page. 2015. URL: https://www.cihi.ca/en.

[161] S. Brinkman, A. Abu-Hanna, A. van der Veen, et al. A comparison of the performance of a model based on administrative data and a model based on clinical data: effect of severity of illness on standardized mortality ratios of intensive care units. *Crit Care Med* 2012: 40(2), 373–378.

[162] P. Stevens and R. Pooley. Essentials of state and activity diagrams. *Using UML: software engineering with objects and components*. USA: Addison-Wesley, 1999. Chap. 11, 239. ISBN: 978-0321269676.

[163] B. Manktelow and S. Seaton. Specifying the probability characteristics of funnel plot control limits: an investigation of three approaches. *PLoS One* 2012: 7(9), e45723.

[164] M. Faddy, N. Graves, and A. Pettitt. Theorem for random variables with infinite moments. *Am J Math* 1946: 68(2), 257–262.

[165] T. Mehdi, N. Bashardoost, and M. Ahmadi. Kernel Smoothing For ROC Curve And Estimation For Thyroid Stimulating Hormone. *Am J Public Health Res* 2011: Special Issue, 239–242.

[166] P. Austin and M. Reeves. The relationship between the C-statistic of a risk-adjustment model and the accuracy of hospital report cards: a Monte Carlo study. *Med Care* 2013: 51(3), 275–284.

[167] S. Seaton and B. Manktelow. The probability of being identified as an outlier with commonly used funnel plot control limits for the standardised mortality ratio. *BMC Med Res Methodol* 2012: 12, 98–106.

[168] S. Seaton, L. Barker, H. Lingsma, et al. What is the probability of detecting poorly performing hospitals using funnel plots? *BMJ Qual Saf* 2012: 22(10), 870–876.

[169] D. Spiegelhalter. Handling over-dispersion of performance indicators. *Qual Saf Health Care* 2005: 14(5), 347–351.

[170] C. Dean. Overdispersion in Poisson and binomial regression models. *JASA* 2015: 87(418), 451–457.

[171] J. Hinde and C. Demétrio. Overdispersion: models and estimation. *Comput Stat Data An* 1998: 27(2), 151–170.

[172] J. IntHout, J. Loannidis, and G. Borm. The Hartung-Knapp-Sidik-Jonkman method for random effects meta-analysis is straightforward and considerable outperforms the standard DerSimonian-Laird method. *Med Res Methodol* 2014: 14(25), 12.

[173] J. Pratt. A normal approximation for binomial, F, Beta, and other common, related tail probabilities, II. *J Am Stat Assoc* 1968: 63(324), 1457–1483.

[174] K. Sidik. Simple heterogeneity variance estimation for meta-analysis. *Appl Stat* 2005: 54(2), 367–384.

[175] D. Spiegelhalter, M. Bardsley, I. Blunt, et al. Statistical methods for healthcare regulation: rating, screening and surveillance. *J R Stat Soc* 2012: 175(1), 1–47.

[176] A. Veroniki and G. Salanti. Methods to estimate the heterogeneity variance, its uncertainty and to draw inference on the meta-analysis summary effect. 2013. URL: https://methods.cochrane.org/statistics/sites/methods.cochrane.org.statistics/files/public/uploads/VeronikiSalantiHeterogeneitySMGMeetingQuebec2013.pdf.

[177] W. Viechtbauer. Bias and efficiency of meta-analytic variance estimators in the random-effects model. *Am Educ Res J* 2005: 30(3), 261–293.

[178] R. DerSimonian and N. Laird. Meta-analysis in clinical trials. *Control Clin Trials* 1986: 7(3), 177–188.

[179] T. Smith, D. Spiegelhalter, and A. Thomas. Bayesian approaches to random-effects meta-analysis: a comparative study. *Stat Med* 1995: 14(24), 2685–2692.

[180] J. Hartung, D. Argac, and K. Makambi. *Homogeneity tests in meta-analysis.* 2003, 361–387.

[181] L. Scrucca. Qcc: An R package for quality control charting and statistical process control. *R news* 1996: 4(1), 11–18.

[182] W. Aelvoet, N. Terryn, A. Blommaert, et al. Community-acquired pneumonia (CAP) hospitalizations and deaths: is there a role for quality improvement through inter-hospital comparisons? *Int J Qual Health Care* 2016: 28(1), 22–32.

[183] S. Goodman. Toward evidence-based medical statistics. 2: The Bayes factor. *Ann Intern Med* 1999: 130(12), 1005–1013.

[184] J. Bland. The tyranny of power: is there a better way to calculate sample size? *BMJ* 2009: 339, b3985.

[185] M. Aregay, Z. Shkedy, and G. Molenberghs. Comparison of Additive and Multiplicative Bayesian Models for Longitudinal Count Data with Overdispersion Parameters: A Simulation Study. *Commun Stat Simul Comput* 2015: 44(2), 454–473.

[186] C. Field and A. Welsh. Bootstrapping clustered data. *J R Stat Soc* 2007: 69(3), 369–390.

[187] A. Kramer. Are ICU Length of Stay Predictions Worthwhile? *Crit Care Med* 2017: 45(2), 379–380.

[188] R. Houthooft, J. Ruyssinck, J. van der Herten, et al. Predictive modelling of survival and length of stay in critically ill patients using sequential organ failure scores. *Artif Intel Med* 2015: 63(3), 191–207.

[189] T. Higgins, N. Starr, J. Lee, et al. Predicting prolonged intensive care unit length-of-stay following coronary artery bypass surgery. *Clin intensive care* 1999: 10(4), 175–182.

[190] R. Huijskes, P. Rosseel, and J. Tijssen. Outcome prediction in coronary artery bypass grafting and valve surgery in the Netherlands: development of the Amphiascore and its comparison with the Euroscore. *Eur J Cardiothorac Surg* 2003: 24(5), 741–749.

[191] I. Borghans, K. Hekkert, L. den Ouden, et al. Unexpectedly long hospital stays as an indicator of risk of unsafe care: an exploratory study. *BMJ Open* 2014: 39(11), 1925–1931.

[192] M. Roos-Blom, D. Dongelmans, M. Arbous, et al. How to Assist Intensive Care Units in Improving Healthcare Quality. Development of Actionable Quality Indicators on Blood use. *Stud Health Technol Inform* 2015: 210, 429–433.

[193] W. Gude, M. Roos-Blom, S. van der Veer, et al. Electronic audit and feedback intervention with action implementation toolbox to improve pain management in intensive care: protocol for a laboratory experiment and cluster randomised trial. *Implement Sci* 2006: 12(1), 68–80.

[194] A. Kramer. A Flock of Birds, a Cluster of ICUs. *Crit Care Med* 2016: 44(5), 1016–1017.

[195] K. Bayley. *Typologies and taxonomies: an introduction to classification techniques.* California: Sage publications, 1994. ISBN: 0-8039-5259-7.

[196] T. Kohonen. Self-organized formation of topologically correct feature maps. *Biol Cybern* 1982: 43(1), 59–69.

Dankwoord

Met veel plezier heb ik de afgelopen jaren aan mijn promotieonderzoek gewerkt. Ik ben dan ook blij met het resultaat. Dit had zeker niet gelukt zonder degene die mij geholpen hebben bij het opzetten van de methodes en het vastleggen van de resultaten van de onderzoeken die ten grondslag liggen aan dit proefschrift.

Om te beginnen wil ik mijn promotoren en copromotoren bedanken. Beste Nicolette ik wil jou bedanken voor je positieve begeleiding. Jouw enthousiasme werkt motiverend en jouw inzichten hebben mijn werk zeker verbeterd. Beste Evert ik wil jou bedanken voor je kritische klinische blik en je scherpe review commentaar. Ik weet zeker dat mijn onderzoeken hierdoor verbeterd zijn. Beste Rebecca en beste Niels, bedankt voor jullie statistisch kritische input, hier heb ik veel aan gehad. Hiernaast wil ik mijn coauteurs bedanken. Beste Ameen, bedankt voor de waardevolle discussies die ik met je gehad heb over funnel plots en maten voor model performance. Beste Dave, bedankt voor je klinische blik als coauteur en je betrokkenheid bij mijn projecten. Ik ben het NICE-bestuur dankbaar voor hun inzet voor de NICE-registratie en alle NICE-deelnemers voor de tijd die zij investeren om de gegevens zo nauwkeurig mogelijk aan te leveren. Deze gegevens zijn het startpunt van mijn onderzoeken. De leden van mijn promotiecommissie bedank ik voor de geïnvesteerde tijd in het beoordelen van mijn proefschrift.

De afgelopen jaren heb ik veel plezier in mijn werk gehad bij de KIK als datamanager en PhD-student. Hiervoor wil ik met name alle NICE collega's bedanken voor de fijne samenwerking. Verder wil ik mijn kamergenoten Twan, Sylvia, Marjolein, Marie-José, Ilse, Rosalie, Rebecca, Charlotte, Ace, Maria, Gaby en Katharina bedanken voor hun bijdrage aan een positieve werksfeer.

Zonder mijn familie had ik nooit zover kunnen komen. Papa en mama, Lindy, Lisette en Micha bedankt voor de fijne jeugd en de motivatie tijdens mijn schoolperiode. Ruud en Hilda bedankt voor het enthousiasme waarmee jullie oppassen en alle logeerpartijtjes van de kinderen.

Tot slot wil mijn gezin bedanken. Vincent, ik ben trots op hoe je werk en gezin weet te combineren. Bedankt voor alle inhoudelijke discussies over ons werk. Annika en Sven, wat ben ik trots op jullie. Jullie regenboog en walvis staan te pronken op deze pagina. Mama is nu eindelijk slimmer dan papa!!!!! Ik hoop dat we nog veel leuke dingen kunnen doen en kunnen genieten van ons nieuwe broertje of zusje.

Liefs, Ilona

Curriculum vitae and portfolio

Curriculum vitae

Ilona W.M. Verburg was born in Schoonhoven, the Netherlands in 1982. In 2000 she finished her pre-university education at 'Het Schoonhovens college', with a special interest in exact sciences. She studied Biomedical Mathematics at the Free University in Amsterdam (VU). Biomedical mathematics combines life science and mathematics to develop new mathematical techniques to create models out of data from recent medical innovations. During here study Ilona had a special interest in statistics and systems biology. She finished her master thesis entitled 'Modeling and Control of Glycolysis in Trypanosoma brucei' in 2006. The aim of this project was to control biochemical reaction networks and combines mathematics with biomedical models.
After obtaining her master's degree Ilona worked from 2006 to 2011 at Statistics Netherlands (Centraal Bureau voor de Statistiek) as statistical researcher in to different functions. At the department of Business Statistics Ilona worked for three years in a project on developing a new methodology and computer program to estimate national turnover rates of companies. At the department of Social and Spatial Statistics Ilona participated for two years in a group of sampling experts. This group performs sampling for most of the major statistics of the department and other government institutions and advices in developing sample designs.
In October 2011 Ilona started as data manager and PhD student at the Dutch National Intensive Care Evaluation (NICE) registry at the Department of Medical Informatics, Academic Medical Center (AMC), University of Amsterdam (UVA), the Netherlands. Under supervision of Prof. Dr. Nicolette F. de Keizer, Prof. Dr. Evert de Jong, Prof. Dr. Niels Peek and Dr. Rebecca Holman she worked on the research described in this Thesis. After finishing here PhD research Ilona continues her work for the NICE registry as a post-doctoral researcher.

Portfolio

Name PhD Student: Ilona W.M. Verburg
PhD Period: October 2011 to September 2017
Promotores: Nicolette F. de Keizer, Evert de Jonge
Co-promotores: Niels Peek, Rebecca Holman

PhD training and courses - (1 of 3)

	Year	Workload (ECTS)
General courses AMC Graduate School		
Evidence based searching in PubMed	2011	0.1
Practical biostatistics	2012	1.1
Clinical epidemiology	2012	0.6
Oral presentation in English	2012	0.8
Reference manager basis	2012	0.1
Scientific writing in English for publication	2013	1.5
Introduction endnote	2014	0.1
Specific courses		
Introductiory course on epidemiology, ERA-EDTA AMC, Amsterdam, The Netherlands	2012	0.6
NIHES: ESP28 Survival analysis, Summer program Erasmus MC, Rotterdam, The Netherlands	2012	1.4
Advanced topics in clinical epidemiology	2014	1.1
Advanced topics in biostatistics	2014	2.1
Anker & Kompas: Persoonlijke kracht via NLP, Driebergen, The Netherlands	2015	1.4
VU Coach Café, Free university (VU), Amsterdam, The Netherlands	2015	0.1
NIHES: EWP13 Advanced analysis of prognosis studies, Winter program Erasmus MC, Rotterdam, The Netherlands	2016	0.7

PhD training and courses - (2 of 3)

	Year	Workload (ECTS)
Seminars, workshops and masterclasses		
Workshop: Absolute risk prediction, Netherlands Cancer Institute (NKI), Amsterdam, The Netherlands	2012	0.25
Symposium: Kwaliteit van data(management) in klinisch wetenschappelijk onderzoek, LUMC, Leiden, The Netherlands	2015	0.1
Symposium: The science of big-data analytics & Visualization, Netherlands eScience center, Utrecht, The Netherlands	2015	0.25
Oral presentations, international and national		
Comparison of different statistical methods to predict intensive care length of stay, ESCTAIC congress 2012, Timisoara, Roemania	2012	0.5
PhD days, Department Medical Informatics AMC, Amsterdam, The Netherlands	2012-2017	1
Lunch presentation: Prognostic models, Department of Gynecologie AMC, Amsterdam, The Netherlands	2016	0.25
Lunch presentation: Practical guidelines for funnel plots for hospital quality indicators, Department KEBB AMC, Amsterdam, The Netherlands	2016	0.25
Poster presentations		
Funnelplots for data quality improvement, Conference of the International Society for Clinical Biostatistics (ISCB 2015), Utrecht, The Netherlands	2015	0.5
Guidelines on constructing funnel plots for quality indicators - a case study on mortality in intensive care units Amsterdam Public Health, Amsterdam, The Netherlands	2015	0.5

PhD training and courses - (3 of 3)

	Year	Workload (ECTS)
Attending (inter)national conferences		
Meeting European Society for Computing and Technology in Anesthesia and Intensive Care (ESCTAIC 2012), Timisoara, Roemania	2012	1
ISCB 2015, Utrecht, The Netherlands	2015	1
ISCB 2016, Birmingham, United Kingdom	2016	1
Medical Informatics PhD days, Department Medical Informatics AMC, Amsterdam, The Netherlands	2012-2017	1
Other PhD training		
Attending research meetings, Department KIK AMC, Amsterdam, The Netherlands	2011-2016	3
Attending book discussions: Modern methods for epidemiology, Y.K. Tu, D.C. Greenwood, Springer Science & Business Media 2012.	2012-2013	1.4
Organizer of Medical Informatics PhD days 2014: No PhD is an island (40 participants), Department Medical Informatics AMC, Amsterdam, The Netherlands	2014	1
Attending KEBB seminar weekly, Department KEBB AMC, Amsterdam, The Netherlands	2016	1.5

Teaching

	Year	Workload (ECTS)
Supervising		
Three months internship: Ashley Duncan, Improving data quality using funnel plots	2015	0.33
One month internship: Ben de Haan, Literature study, Funnel plots in registry reporting	2015	0.33
One month internship: Senate Lesaoana, Literature study, Factors influencing intensive care unit length of stay	2016	0.33
Bachelor internship: Tim Verhagen, Artefactdetectie in hartslagfrequentie metingen op de intensive care	2017	2
Bachelor internship: Tijmen Henrich, Het detecteren van artefacten in ademhalingsfrequentie metingen op de intensive care met behulp van outlier detectie, Department Medical Informatics AMC, Amsterdam, The Netherlands	2017	2
Other teaching		
Health informatics: module kwaliteitsregistraties, Department Medical Informatics AMC, Amsterdam, The Netherlands	2017	1

List of publications

Publications available in this thesis

- **I.W.M. Verburg**, N.F. de Keizer, E. de Jonge and N. Peek
  Comparison of regression methods for modeling intensive care length of stay.
  PLoS One. 2014;9(10):e109684.

- **I.W.M. Verburg**, N.F. de Keizer, R. Holman, D.A. Dongelmans, E. de Jonge and N. Peek
  Individual and Clustered Rankability of ICUs According to Case-Mix-Adjusted Mortality.
  Critical Care Medicine 2016, 44(5):901-909.

- **I.W.M. Verburg**, A. Atashi, S. Eslami, R. Holman, A. Abu-Hanna, E. de Jonge, N. Peek and N.F. de Keizer
  Which models can I use to predict adult ICU length of stay? A systematic review.
  Critical Care Medicine 2017;45(2):e222-e231.

- **I.W.M. Verburg**, R. Holman, N. Peek, A. Abu-Hanna and N.F. de Keizer
  Guidelines on constructing funnel plots for quality indicators: A case study on mortality in intensive care unit patients.
  Statistical Methods in Medical Research 2017, 1:962280217700169.

- **I.W.M. Verburg**, R. Holman, D.A. Dongelmans, E. de Jonge and N.F. de Keizer
  Is patient length of stay associated with intensive care unit characteristics?
  Journal of critical care 2017;43:114-121.

- **I.W.M. Verburg**, E. de Jonge, N. Peek and N.F. de Keizer
  The association between outcome-based quality indicators for intensive care units.
  Submitted.

Other publications

- **I.W.M. Verburg**, M. van Vliet, M. van den Boogaard, N.F. de Keizer, N. Peek, N.M. Blijlevens and P. Pickkers
  Trends in admission prevalence, illness severity and survival of haematological patients treated in Dutch intensive care units.
  Intensive Care Medicine 2014;40(9):1275-84.

- M.M. Bos, **I.W.M. Verburg**, I. Dumaij, J. Stouthard, J.W. Nortier, D. Richel, E.P. van der Zwan, N.F. de Keizer and E. de Jonge
  Intensive care admission of cancer patients: a comparative analysis.
  Cancer Medicine 2015;4(7):966-76.

- A. Atashi, **I.W.M Verburg**, H. Karim, A. Abu-Hanna, E. de Jonge, N. Peek, N.F. de Keizer and S. Eslami
  Reporting and methodological quality of prediction models of prolonged length of stay in the intensive care unit after coronary artery bypass grafting: A systematic review.
  Submitted.

List of abbreviations

ρ	rankability
σ^2	uncertainty
τ^2	heterogeneity
AIC	Akaike Information Criterion
APACHE	Acute Physiology and Chronic Health Evaluation
APS	Acute Physiology Score
CABG	coronary artery bypass graft
CAP	community acquired pneumonia
CHARMS	Checklist for Critical Appraisal and Data Extraction for Systematic Reviews of Prediction Modeling Studies
CPH	Cox proportional hazards
EMBASE	Excerpta Medica database
GCS	Glasgow coma scale
GLM	generalized linear models
GLMMs	generalized linear mixed-fixed models
ICU	intensive care unit
IQR	interquartile range
ISO	International Organization for Standardization
LM	linear models
LMMs	linear mixed-effects models
MAPE	mean absolute prediction error
MEDLINE	Medical Literature Analysis and Retrieval System Online
MeSH	Medical Subject Headings
NICE	Dutch National Intensive Care Evaluation foundation
NVIC	Dutch Society of Intensive Care
OHCA	out of hospital cardiac arrest
OLS	ordinary least square
R^2	squared Pearson's correlation coefficient
RA EWMA	risk-adjusted exponentially weighted moving average
RAMR	risk adjusted mortality rate
RMSPE	root mean squared prediction error
SAPS	Simplified Acute Physiology Score
SLOSR	standardized ICU length of stay ratio
SMR	standardized in-hospital mortality ratio
SOFA	Sequential Organ Failure Assessment
SPC	statistical process control
SRR	standardized readmission ratio
VLAD	variable life adjusted display