Information for intensive care evolution : methods to assess and improve data quality and data processing

Arts, D.G.T.

Citation for published version (APA):

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (http://dare.uva.nl)
Chapter 4

Quality of data collected for severity-of-illness scores in the Dutch National Intensive Care Evaluation (NICE) registry

Danielle G.T. Arts, Nicolette F. de Keizer, Gert-Jan Scheffer, Evert de Jonge

Intensive Care Medicine 2002;28:656-659
Abstract
To be useful, for example for assessment of quality of care, data in medical registries must be of good quality. The aim of this study is to analyse the quality of data used to measure severity of illness in the Dutch National Intensive Care Evaluation (NICE) registry, after implementation of quality improving procedures.
Data from a random sample of admission to nine Dutch ICUs were re-abstracted from the paper patient record or the Patient Data Management System and compared to the data contained in the registry. The re-abstracted data were considered to be the gold standard.
The mean percentages inaccurate and incomplete data, per hospital, over all variables, were 6.1% ± 4.4 (SD) and 2.7% ± 4.4 (SD) respectively. The mean difference in severity-of-illness scores between registry data and re-abstracted data was 0.4 points for APACHE II and 0.2 points for SAPS II. The mean difference in predicted mortality according to APACHE II and SAPS II between registry data and re-abstracted data was 0.36% and 0.04% respectively.
We concluded that the current data quality of the NICE registry is good and justifies evaluative research. These positive results might be explained by the implementation of several quality assurance procedures in the NICE registry, such as training and automatic data checks.
Introduction

Severity models are increasingly used to compare patient outcomes between intensive care units (ICU). Quality assessment of ICU’s can be performed by comparing observed mortality in the ICU population with the estimated overall risk of hospital mortality in that population. To ascribe differences in mortality ratios (observed mortality / predicted mortality) between ICU’s to true differences in the quality of intensive care practice, we must be certain that they cannot be explained by insufficient accuracy and consistency of the data used for calculation of the predicted mortality. Inaccurate registration and variability of the data can have a number of causes. Fery-Lemonnier et al. [1] discussed some of the problems causing inaccurate APACHE II data collection, including complex calculation methods and lack of unambiguous definitions. In addition Chen et al. [2] mentioned lack of clear instructions concerning the timing of APACHE II data collection as a source of error. To optimise the quality of medical registry data the centers that participate in a registry should follow certain procedures that aim to minimise the occurrence of erroneous and missing data [2-15]. Common errors in data collection should be avoided by providing unequivocal data definitions, clear procedures for data collection and training of the data collectors [1-8,15-17]. Furthermore, database quality can be improved by automatically verifying whether data are complete and within the defined range of values, and by applying consistency checks.

The Dutch National Intensive Care Evaluation (NICE) registry [18] was set up in 1996 to gain insight into and to improve the effectiveness of Dutch intensive care medicine. The NICE registry contains 96 data items for each patient that has been admitted to one of the participating ICU’s. Based on data from the first 24 hours of ICU admission a scoring system, APACHE II [19] APACHE III [20], SAPS II [21], MPM$_{0/24}$ II [22], or LODS [23] assigns a score to each patient to quantify his severity of illness. This score can be turned into a probability of hospital mortality by a logistic regression equation. NICE has adopted a framework of measures to obtain reliable data with clear data definitions, local and central procedures to collect data, automatic evaluation of completeness and quality of the data, and obligatory training of data-collectors of every participating center. This study was undertaken to determine the reproducibility of critical care data after institution of a framework of measures aimed at improving quality of the data.

45
Materials and methods

The NICE dataset contains 96 variables representing characteristics of the patients and the outcome of ICU treatment. It includes demographic patient data, admission and discharge data and all variables necessary to calculate mortality risks according to the prognostic models APACHE II, APACHE III, SAPS II, MPM_{0/24} II, and LODS.

Data collection in the NICE registry takes place manually or automatically. Manual data collection implies that the data are manually entered into a local registry database, possibly preceded by filling in a specially designed case record form. In case of automatic data collection the registry data is automatically extracted from the electronic patient record or Patient Data Management System (PDMS) [24] and placed in the local registry database. Each month the data from the local databases is transferred to the central registry database at the NICE co-ordinating center. To assure quality of the data in the NICE registry several quality assurance measures have been implemented. The NICE registry has developed a data dictionary, containing clear definitions of all NICE data items. Some of these definitions are mentioned on the case record forms or on the data entry screens. All participants of the NICE registry are obliged to attend a training in collecting the data accurate, according to the stated data definitions. In the NICE registry both the lowest and the highest values for physiological and laboratory data over the first 24 hours of ICU admission are registered. The worst values, required for prognostic scoring models, are selected automatically when the prognostic scores are calculated. Both locally and centrally the data are automatically checked for range and consistency. Detected data errors are reported back to the local centers, so they can correct the data. The NICE co-ordinating center continuously maintains intense communication with local participating centers.

To measure current data quality we visited ICUs from all 9 hospitals that had been collecting data for the NICE registry for at least one year. For each of these ICUs we randomly selected 20 patients that had been admitted in August or September 1999. The data were re-abstracted from the paper patient record or the PDMS by one of the authors (D.A.). To evaluate the accuracy (= the extent to which registry data are consistent with the original source data) and the completeness (= the extent to which all necessary data that have been registered in the original source are available in the registry) of all NICE data items (N=96) the re-abstracted data were compared to the data contained in the registry. The re-abstracted data were considered to be the gold standard. Percentages inaccurate and missing data items were calculated for all data items clustered into four groups: admission & discharge data, diagnostic data, physiologic and laboratory data and Glasgow Coma Scale scores. Registered data items were found to be inaccurate if (1) a categorical value was not equal to the gold standard value, (2) a numerical value deviated from the gold
standard value more than acceptable, for example a deviation in systolic blood pressure of more than 10 mmHg below or over the gold standard systolic blood pressure was considered inaccurate (See the appendix for a complete list of the criteria) or (3) a value caused a deviation from the gold standard APACHE II score or SAPS II score. A data item was found to be missing when it was not registered, while it was available according to the paper-based patient record or the PDMS. We specifically analysed the quality of the variables used to calculate APACHE II and SAPS II scores. Here APACHE II and SAPS II data items were considered inaccurate, only if they caused a deviation from the gold standard score. In addition we calculated kappa coefficients to determine the for chance adjusted agreement in assignment of APACHE II and SAPS II scores. We also investigated the influence of the observed data errors on prognostic scoring by calculating the differences between mean APACHE II and SAPS II scores and mortality risks based on the registry data and based on the re-abstracted data. A paired non-parametric Wilcoxon signed ranks test was performed to assess whether or not the mean differences were different from zero. A p-value < 0.05 was considered statistically significant.

Results

A total of 146 patient records from all ICUs were reviewed in this evaluation. Percentages inaccurate and missing data items per ICU and per group of data items are displayed in figure 4.1. Of all 9 ICUs the mean percentages inaccurate and missing data items were 6.1% ± 4.4 (SD) and 2.7% ± 4.4 (SD) respectively. The highest percentages inaccurate data items concerned physiologic data, such as systolic blood pressure, and laboratory data, such as white blood cell count. Percentages inaccurate admission and discharge data, diagnostic data and Glasgow Coma Scores were low.

ICU A2 had a relatively high percentage inaccurate data items (15.8%). In the ICUs where data was collected automatically (ICUs A1 and A2) inaccurate data were mostly due to programming errors in the extraction software. Both ICUs A1 and A2 had relatively high percentages of missing data items (6.0% and 13.3% respectively). This was caused by the fact that the extraction software lacked extraction queries for some data items or that specific variables were not registered in the PDMS, which caused the missing of these specific data items for all patients admitted to that ICU. Comparison of the data in the electronic patient record of hospital A2 to the data in the paper based patient record showed that the data in the electronic record, the extraction source for the NICE data, was incomplete. As a result of the incompleteness of the extraction source, many values in the NICE registry appeared not to be the real extreme values over the first 24 hours of ICU admission. Glasgow Coma Scores at admission
and after 24 hours were not registered in the electronic patient record for all patients admitted at ICU A2. Whereas the lowest GCS scores over the first 24 hours of ICU admission, used for calculation of APACHE II and SAPS II scores, were registered.

![Graph showing percentages of inaccurate and incomplete data per ICU](image)

**Figure 4.1 - Percentages inaccurate and incomplete data per ICU, over all data items.** The bars represent percentages inaccurate or incomplete data per group of data items, for each ICU. ICUs are divided into ICUs that automatically collect the data (A1 and A2) and ICUs that manually collect the data (M1-M7).

In ICUs where data were collected manually (ICUs M1 till M7) most errors were made during collection of the data on the case record forms. Only a very small fraction of the data errors was made during the manual entering of the data into the database. Most of the erroneous physiologic and laboratory data concerned data acquired before or after the first 24 hours of ICU admission. Another relatively large source of error was the selection of the blood gas sample for registration of blood gas values (PaO2, PaCO2, FIO2, pH) and calculation of the alveolar-arterial oxygen difference (A-aDO2). The sample used should be the one that results in the highest A-aDO2. An incorrectly selected blood gas sample may result in multiple inaccurately registered values. Finally calculation of minimal urine production over 8 hours and total urine production over 24 hours resulted in a relatively large number of errors.
Table 4.1 displays for all APACHE II and SAPS II data items the agreement between registry data and re-abstracted data. In this case, data were considered inaccurate, if they caused a deviation from the gold standard score and mortality risk. The agreement was very high (>90%) in most cases. The data item ‘Mechanical ventilation during the first 24 hours of ICU admission’ showed relatively low agreement, probably due to frequent misinterpretation of the data definition. Table 4.2 also displays a kappa (K) coefficient for each data item. Most variables show very good agreement (0.81 < K < 1.00). Kappa for white blood cell (WBC) count was good (K = 0.77) for APACHE II and moderate (K = 0.43) for SAPS II scoring. These relatively low kappa coefficients were to a large extent caused by a programming error in the extraction software of hospital A2 (see figure 4.1). This error caused the value of zero to be recorded in the registry when the WBC was not measured. In the assignment of APACHE II and SAPS II scores this value was considered to be extremely low and therefore unjustly generated the maximum number of score points.

Table 4.1 – Percentage agreement between registry data and re-abstracted data in the assignment of APACHE II and SAPS II scores. Kappa coefficients per variable.

<table>
<thead>
<tr>
<th>APACHE II variables</th>
<th>Agreement %</th>
<th>Kappa</th>
<th>SAPS II variables</th>
<th>Agreement %</th>
<th>Kappa</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chron. Diagnoses c</td>
<td>91.1</td>
<td>0.74</td>
<td>Mechanical ventilation</td>
<td>84.9</td>
<td>0.68</td>
</tr>
<tr>
<td>WBC count a</td>
<td>91.1</td>
<td>0.77</td>
<td>PaO2/FIO2</td>
<td>90.4</td>
<td>0.84</td>
</tr>
<tr>
<td>Mean blood pressure</td>
<td>91.1</td>
<td>0.84</td>
<td>WBC count a</td>
<td>92.5</td>
<td>0.43</td>
</tr>
<tr>
<td>A-aDO2 / pao2</td>
<td>93.8</td>
<td>0.85</td>
<td>Bicarbonate</td>
<td>95.2</td>
<td>0.88</td>
</tr>
<tr>
<td>Heart rate</td>
<td>94.5</td>
<td>0.91</td>
<td>Systolic blood pressure</td>
<td>95.9</td>
<td>0.93</td>
</tr>
<tr>
<td>pH</td>
<td>95.2</td>
<td>0.85</td>
<td>Potassium</td>
<td>97.3</td>
<td>0.82</td>
</tr>
<tr>
<td>Hematocrit</td>
<td>95.2</td>
<td>0.91</td>
<td>Heart rate</td>
<td>97.3</td>
<td>0.97</td>
</tr>
<tr>
<td>Respiratory rate</td>
<td>95.2</td>
<td>0.94</td>
<td>GCS scores b</td>
<td>98.0</td>
<td>0.85</td>
</tr>
<tr>
<td>Temperature</td>
<td>95.9</td>
<td>0.94</td>
<td>Urea</td>
<td>98.6</td>
<td>0.96</td>
</tr>
<tr>
<td>Potassium</td>
<td>97.3</td>
<td>0.94</td>
<td>Temperature</td>
<td>98.6</td>
<td>0.89</td>
</tr>
<tr>
<td>GCS scores b</td>
<td>97.3</td>
<td>0.80</td>
<td>Age</td>
<td>99.3</td>
<td>0.99</td>
</tr>
<tr>
<td>FIO2</td>
<td>97.3</td>
<td>0.95</td>
<td>Admission type</td>
<td>99.3</td>
<td>0.99</td>
</tr>
<tr>
<td>Sodium</td>
<td>98.6</td>
<td>0.86</td>
<td>Sodium</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>Acute renal failure</td>
<td>98.6</td>
<td>0.83</td>
<td>Bilirubin</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>Creatinine</td>
<td>99.3</td>
<td>0.98</td>
<td>Urinary output</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>Age</td>
<td>99.3</td>
<td>0.99</td>
<td>Chron. Diagnoses d</td>
<td>100</td>
<td>1</td>
</tr>
<tr>
<td>Admission type</td>
<td>99.3</td>
<td>0.98</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a WBC = white blood cell, b GCS = Glasgow Coma Scale, c Cirrhosis, cardiovascular insufficiency, respiratory insufficiency, chronic dialysis, immunological insufficiency, d Haematological malignancy, neoplasm, AIDS
Table 4.2 displays the differences in mean APACHE II and SAPS II scores and predicted mortality based on the registry data and on the re-abstracted data. Overall agreement regarding the APACHE II and the SAPS II scores and mortality probabilities was very good. The detected data errors had no impact on the scores and the calculation of case-mix corrected mortality risks. Figures 4.2 and 4.3 show for the APACHE II model and for the SAPS II model that the in general individual predicted probabilities of in-hospital death based on the registry data do not differ significantly from those based on the re-abstracted (gold standard) data. Figure 4.3 displays a few outliers for the SAPS II model.

Table 4.2 - Comparison of mean scores and probabilities based on the registry data and based on the re-abstracted (gold standard) data, results of the Wilcoxon signed ranks test.

<table>
<thead>
<tr>
<th></th>
<th>Registry</th>
<th>Gold standard</th>
<th>Difference</th>
<th>Wilcox</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean ±SD</td>
<td>Mean ±SD</td>
<td>Mean (min-max)</td>
<td>p-value</td>
</tr>
<tr>
<td>SAPS II score</td>
<td>29.6 ±15.1</td>
<td>29.9 ±14.4</td>
<td>-0.4 (0-16)</td>
<td>0.331</td>
</tr>
<tr>
<td>SAPS II probability</td>
<td>16.6% ±20.0</td>
<td>16.6% ±19.3</td>
<td>0.04% (0-35.3)</td>
<td>0.907</td>
</tr>
<tr>
<td>APACHE II score</td>
<td>14.2 ±5.6</td>
<td>14.4 ±5.6</td>
<td>-0.17 (0-8)</td>
<td>0.275</td>
</tr>
<tr>
<td>APACHE II probability</td>
<td>13.8% ±16.5</td>
<td>14.1% ±16.3</td>
<td>-0.4% (0-20.9)</td>
<td>0.197</td>
</tr>
</tbody>
</table>

Figure 4.2 - Predicted mortality according to APACHE II based on registered data plotted against predicted mortality based on re-abstracted data, for individual patients. The line represents perfect agreement.
Data quality in the NICE registry

Discussion

The value of medical registries, for example in the assessment of quality of care, strongly depends on the quality of the data included in the registry. To be able to make reliable inferences, these data need to be accurate and complete. The NICE registry has implemented many procedures aiming to improve the quality of data. The results of the present study show that the data contained in the NICE database are indeed of good quality and suggest that current data quality assurance procedures are effective.

The primary goal of the NICE registry is the evaluation of quality of intensive care by determining case-mix-adjusted mortality. For this, prognostic models such as APACHE II and SAPS II are used to calculate predicted mortality. Very high agreement was found between registry data and re-abstracted data for APACHE II and SAPS II scores and for predicted mortality based on these models. High agreement was also found for the individual items that are used in these models with the exception of the SAPS II item “mechanical ventilation” which had a relatively low agreement rate of 84.9%.

The criteria used for evaluation of the quality of APACHE II and SAPS II data implied that a data value was considered inaccurate, only if it would result in a different score than the gold standard score. Since APACHE II and SAPS II use
different threshold values in the assignment of scores, the number of inaccurate values for data items used by both models, such as GCS scores, WBC count and heart rate, can be different.

Our results differ somewhat from earlier reports in the literature. GCS scoring is a well-known source of erroneous data, especially in the case of sedated patients [2, 16, 17, 25]. The very low number of erroneous GCS scores in our study can be explained by the explicit definitions for this item, including what to do in case of sedated patients, that are given on every case record form as well as on the electronic data entry module. Also, during the obligatory training of every participating center, much attention is given to this subject. Definitions on the case record form are also given for other “difficult” items, such as diagnoses of chronic illness, oliguria, highest A-aDO$_2$ and definitions of surgical vs. medical admissions. Another source of errors reported in the literature is the use of physiologic data from before or after the first 24 hours in the ICU [2, 15, 16]. To avoid this type of error, specific attention is given to this point during training sessions. Holt et al. [16] reported that 38% of errors occurred in choosing the most abnormal from highest and lowest values. In the NICE registry, both the highest and lowest values are recorded, the worst value is automatically selected by a computer algorithm. Lastly, errors can be avoided by letting physicians register all data necessary for calculations, rather than the calculated values such as A-aDO$_2$ themselves.

We did evaluate the accuracy of all database items, not only the data used for calculating the APACHE II and SAPS II models. Whereas in most ICU’s the number of errors was very low, some important outliers could be identified. Both ICU’s that collect their data automatically from a PDMS had a relatively high percentage missing data items due to the fact that the extraction software lacked extraction queries for some items and because some items were not registered in the PDMS. For example one ICU had 66% missing GCS items, due to the fact that the software only contained extraction queries for the most abnormal GCS scores but not for the GCS at admission and at 24 hours after admission. Both ICUs have resolved the errors in their extraction software, which caused the missing data items, and are now able to provide the data accurate and complete. This clearly shows the importance of data quality evaluation during site visits to detect systematic errors in local data collection procedures.

In conclusion, this study shows that overall data quality of medical registries can be very good. The NICE registry data are of very good quality in view of the intended use, severity-of-illness scoring. Very high agreement was found between registry data and re-abstracted data for APACHE II and SAPS II scores and for predicted mortality based on these models.
References


