An evolution of trauma care evaluation: A thesis on trauma registry and outcome prediction models
Joosse, Pieter

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: http://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.
Chapter 7

IMPROVING PERFORMANCE AND AGREEMENT IN INJURY CODING USING THE ABBREVIATED INJURY SCALE; A TRAINING COURSE HELPS.

PIETER JOOSSE
MARISKA AC DE JONGH
CCHM (KATINKA) VAN DELFT-SCHREURS
MICHEL HJ VERHOFSTAD
J CAREL GOSLINGS

Submitted.
ABSTRACT

Background
Reliability of injury coding underlying trauma databases is crucial for valid outcome comparison. Reliability in injury coding can possibly improve by training and education.

Objective
To assess the influence of a dedicated training course on the ability of participants to assign correct codes and on inter-observer agreement using the Abbreviated Injury Scale (AIS).

Methods
Twelve participants followed a one-day training course in injury coding. The AIS code for 47 injuries was recorded before (pre-test), during (test) and after (post-test) the course. The ability to assign the correct code and severity score was calculated, as well as the inter-observer agreement for all three tests.

Results
The number of correct assigned codes overall improved significantly during and after the course. At rater level, the median percentage of correct unique codes was 71.3% in the pre-test, 86.2% in the test, and 83.0% in the post-test. Improvement was significant for test compared to pre-test. The median percentage of correct severity scores was 85.9% in the pre-test, 90.5% in the test, and 91.5% in the post-test. Improvement was significant for test compared to pre-test. Fleiss kappa's fell within the substantial to almost perfect range and improved significantly during and after the course compared to the pre-test.

Conclusion
The results of this study emphasized the use of training in AIS coding to improve the ability to assign correct codes and to reduce inter-observer variability. It is advisable that all who are involved in injury coding using the AIS are trained by a dedicated course.
INTRODUCTION

Trauma databases enable analysis of performance data of individual institutions compared to a standard. They also serve to monitor the effectiveness of efforts taken to improve the quality of trauma care. Injury coding underlying trauma databases needs to be standardized and reproducible in order to draw valid conclusions from trauma databases.

The classification of injuries by type and severity according to the Abbreviated Injury Scale (AIS) forms the basis of several commonly used measures of overall injury severity such as the Injury Severity Score (ISS)\(^1\) and the New Injury Severity Score (NISS).\(^2\) The AIS is an anatomically based, consensus derived, internationally accepted global severity scoring system that classifies an individual injury by body region according to its relative severity. The first version of the AIS was published in 1971 with the purpose to categorize types and severity of injuries sustained in motor vehicle crashes.\(^3\) The AIS has since been revised several times by the Association for the Advancement of Automotive Medicine (AAAM) Committee on Injury Scaling, most recently in 2005,\(^4\) followed by an update in 2008.\(^5\)

Accurate and consistent application of the AIS is fundamental to sound injury data collection and the validity of methods that incorporate the AIS. Injury coding using the AIS is subject to variation between observers.\(^6-10\) Variation in injury coding can possibly be reduced through training and education.\(^11\) However, in none of these studies inter-observer reliability was assessed prior and after the course. Neither has the inter-observer variation of the AIS98 been assessed corrected for chance.

The main objective of this study is to assess prospectively the influence of a dedicated training course on the ability of participants to assign correct injury codes and on inter-observer agreement using the AIS98.

MATERIALS AND METHODS

In the Netherlands a national trauma registry holds data of all acute, hospitalized trauma patients. Eleven designated level-1 trauma centres are responsible for the collection, submission and scrutiny of trauma survival data within regional networks rendering national coverage. The core dataset contains patient demographics, physiological parameters and anatomical injury descriptions. To date, the Dutch national trauma database uses the AIS 1990 revision 1998 update (AIS98).\(^12\)

Data were collected during a course in injury coding of the Sint Elisabeth Hospital, Tilburg, the Netherlands. This is an annual one-day course open to all health care workers (e.g. physicians, nurses, research assistants or medical record technicians) who are in charge of injury coding for trauma registration. A trauma surgeon, a neurosurgeon and a trauma registry coordinator are responsible for the content of the course programme. The course director is trained by an AAAM certified instructor. The course was initiated in 2004 and is currently the only ‘public’ course on injury coding in the Netherlands. During the course the participants learn how to use the AIS98 dictionary and how to apply coding rules. They practice in injury coding of single and multiple injuries, extracting injuries from medical records and calculation of the ISS. Previous experience of the participants in AIS coding was assessed by a questionnaire.
Participants coded a sample of 47 injuries before, during and after the course. A pre-test was sent to their home address together with the course instructions and AIS dictionary and collected at the beginning of the course. The exercises the participants made during the course served as the test. The test was filled out on carbon paper and collected before the correct answers to the test were discussed. Finally, a post-test was sent and returned by email six weeks after completion of the course. The injuries for the pre-test and test were identical. For the post-test a different sample of 47 injuries was made in order to avoid that participants would be able to look up the correct answers from their course manual. The level of difficulty for the post-test was comparable to the pre-test and test. Also the distribution of AIS severity scores was comparable. Injuries were either modified from patient charts or contrived by the directors of the course based on their prevalence and educational value. None of the participants had prior knowledge to the exercises.

The injuries that had to be coded were divided over 6 exercises; general, extremities, thoracic and pelvic contents, head and face, external and spine. All injuries had to be considered as isolated injuries. Two injuries could not be assigned an AIS code. These injuries were ‘ingestion of a safety-pin’ (i.e. a foreign body) and ‘painful neck’ (i.e. a symptom or rather a consequence of an unknown anatomic injury). The correct answer was to leave these injuries uncoded. For the purpose of analysis these injuries were assigned an AIS severity score of 0. Injuries that were left uncoded deliberately by the participants but should have been coded were also assigned an AIS severity of 0. In case of missing values for the test due to time restraints the answers were adopted from the pre-test, if present. An example of the (pre-)test and post-test for thoracic and pelvic injuries is given in Table 1.

<table>
<thead>
<tr>
<th>(Pre-)test</th>
<th>Injury</th>
<th>AIS code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flail chest with lung contusion</td>
<td>450204.4</td>
<td></td>
</tr>
<tr>
<td>Rib fractures; 4 right, 3 left with hemothorax and bilateral lung laceration</td>
<td>450230.3, 441450.4</td>
<td></td>
</tr>
<tr>
<td>Acetabular fracture</td>
<td>852800.2</td>
<td></td>
</tr>
<tr>
<td>Myocardial contusion</td>
<td>441002.1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Post-Test</th>
<th>Injury</th>
<th>AIS code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type B pelvic fracture (‘open book’)</td>
<td>852806.4</td>
<td></td>
</tr>
<tr>
<td>Rib fractures; 4 right, 3 left with hemothorax</td>
<td>450232.4</td>
<td></td>
</tr>
<tr>
<td>Sternal fracture</td>
<td>450804.2</td>
<td></td>
</tr>
<tr>
<td>Superficial stab wound to the thorax</td>
<td>416002.1</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 | An example of (pre-)test and post-test with model answers

Actual performances score of the participants was measured for pre-test, test and post-test. The median percentage was determined for correct unique 7-digit AIS codes as well as for the correct AIS severity score. For example, a bimalleolar fracture (851612.2) that was coded as a lateral malleolar fracture (851608.2) was judged false for the 7-digit AIS code and judged...
A training course for AIS coding correct for the AIS severity score. The difference in valid percentage of correct answers before, during and after the course was tested with the Wilcoxon signed-ranks test.

Inter-observer agreement on the AIS severity score was measured using Fleiss (generalized) kappa for pre-test, test and post-test with each participant serving as unit of observation. Eight categories were offered for the AIS severity; these categories were 0, 1, 2, 3, 4, 5, 6 and 9. Participants with incomplete tests were excluded from inter-observer analysis. The kappa statistic, first introduced by Cohen\textsuperscript{13} and later generalised by Fleiss,\textsuperscript{14} is the percentage agreement corrected for chance agreement under the baseline assumption that the ratings are made independently. Kappa can range from -1 to +1. It is negative when expected agreement is greater than observed, zero when expected agreement is equal to observed, positive when observed agreement is greater than expected agreement, and +1 when agreement is perfect. No objective criteria to interpret kappa exist. We used the guidelines as presented by Landis and Koch\textsuperscript{15} to judge the strength of agreement. Kappa statistics for separate tests were judged as independent measures and tested for significance with the unpaired student t-test.

Data analysis was undertaken using the Statistical Package for the Social Sciences (SPSS) version 18.0 (SPSS, Chicago, Ill., USA). Fleiss (generalized) kappa was calculated in syntax for SPSS (courtesy to J.E. King). Significance was attributed to p values of less than 0.05.

RESULTS

Twelve persons participated in the injury coding course. Occupation and experience in injury coding of the participants varied. They included five data managers (of whom two were medical students), four physicians, one research coordinator with extensive experience in medical research but no medical background, one research nurse and one emergency department clerk (former nurse). Half of the participants had at least 3 months experience in injury coding.

Overall Analysis

A total of 564 (12×47) injuries had to be coded for pre-test, test and post-test each. Coded injuries from incomplete test were included in the overall and participant’s level analysis. A total of 495 injuries were coded in the pre-test, while three pre-tests were incomplete. A total of 545 injuries were coded in the test, while one test was incomplete. The post-test was completed by ten participants (response rate 83%) who coded 470 injuries altogether. Two participants did not return the post-test. The percentages of correct unique 7-digit codes are presented in Figure 1 and correct AIS severity scores in Figure 2.

Analysis at Participant’s Level

Ten participants improved their percentage of correct answers in the test compared to the pre-test ($p = 0.007$, Wilcoxon signed-ranks test). Six participants improved their percentage of correct answers in the post-test compared to the pre-test, while two participants improved compared to the test ($p = 0.203$ and $p = 0.207$ respectively; Wilcoxon signed-rank test). The median valid percentage of correct unique 7-digit codes was 71.3% in the pre-test, 86.2% in the test, and 83% in the post-test.
Chapter 7

**Figure 1** | Percentages correct unique 7-digit codes
*test vs pre-test **post-test vs pre-test

**Figure 2** | Percentages correct AIS severity codes  
*test vs pre-test **post-test vs pre-test
Ten participants improved their percentage of correct AIS severity score in the test compared to the pre-test ($p = 0.031$; Wilcoxon signed-ranks test). Seven participants improved their percentage of correct AIS severity score in the post-test compared to the pre-test, while four participants improved compared to the test ($p = 0.138$ and $p = 0.594$ respectively; Wilcoxon signed-rank test). The median valid percentage of correct AIS severity scores was 85.9% in the pre-test, 90.5% in the test, and 91.5% in the post-test.

Inter-observer Agreement
Kappa statistics fell within the substantial and almost perfect range. Kappa statistic was 0.74 (0.72 – 0.77) for 9 complete pre-tests, 0.81 (0.79 – 0.83) for 11 complete tests, and 0.79 (0.77 – 0.81) for 10 complete post-tests (Table 2). Inter-observer agreement on the AIS severity score improved significantly for the test compared to the pre-test and post-test compared to pre-test ($p = 0.0006$ and $p = 0.013$; unpaired $t$-test).

<table>
<thead>
<tr>
<th>Number of participants</th>
<th>Kappa (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-test</td>
<td>9</td>
</tr>
<tr>
<td>Test</td>
<td>11</td>
</tr>
<tr>
<td>Post-test</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 2 | Inter-observer variation using the abbreviated injury scale 1990 revision 1998 update. CI; confidence interval. Guidelines for the interpretation of kappa statistics (15).

Kappa Degree of Strength
0.00 – 0.20 Negligible
0.21 – 0.40 Fair
0.41 – 0.60 Moderate
0.61 – 0.80 Substantial
0.81 – 1.00 Almost perfect

Analysis at Case Level
Injuries that were coded incorrectly for more than 50% involved primarily compound injuries, e.g. ‘gun shot wound to cerebrum and cerebellum’, ‘penetrating injury to liver with more than 20% blood loss and perforation of the portal vein’. For the test and post-test, injuries with a higher injury severity score were more often coded incorrectly ($p = 0.045$ and $p = 0.008$; Linear-by-Linear Association).

DISCUSSION
In the present study, a dedicated course on injury coding using the AIS98 significantly improves the number of correct coded injuries and assigned injury severity. The learning effect of a course seems to last, as the results of this study show improved performance scores until six weeks after the course. However, at participant’s level the benefits of the course were no longer significant at the post-test. Besides the ability to assign the correct code, the
inter-observer agreement among raters with varying background improves significantly by following a course. Compound injuries and more severe injuries tend to be more often coded incorrectly.

To our knowledge, no earlier study documented the influence of a course on the quality and reliability of the AIS in a prospective design. Mackenzie and colleagues examined the reliability of the AIS 1980 revision when applied by observers with varying backgrounds. The kappa’s reported in this study fell within the substantial range, as did the kappa’s in our study. Required for participation in the study was attendance at a 5-hour training workshop, but reliability was not assessed prior to the study. Read-Allsopp conducted a coding exercise amongst 27, mostly experienced injury coders four months after a dedicated training program in AIS88 coding. An overall accuracy rate of 79.8% was reported, almost identical to the overall percentage of correct unique 7-digit codes in the present study. Also in this study, reliability was not assessed prior to the training and consequently the influence of the training on the results could not be established. Secondly, inter-rater reliability was reported as overall accuracy rate or percentage correct codes at participant’s level. No attempt for correction for chance was made.

Coding variability of the AIS has flow-on effects for the overall utility of AIS-coded trauma databases. The ISS is extensively used to describe overall injury severity or as a component in outcome prediction models for trauma patients. In a German study a wide variation in ISS for individual patients depending on the observer was reported. The study dates from 1992, however it is not explicitly stated which AIS version was used. Zoltie and De Dombal reported a wide variation in recorded injury severity scores. Fifteen observers were asked to identify every anatomical injury from 16 case notes and code the injuries according to the AIS 1990 revision. The probability that any two observers would agree on the exact ISS was 28% and on the correct ISS range was 50%. In contrast with our study, observer agreement did not depend whether observers attended a coding course or were coding actively. Neale and colleagues investigated the quality of routine coding in the Queensland Trauma Registry. Despite a relatively low agreement between coders for unique 7-digit AIS98 codes, they report an intraclass correlation coefficient for the ISS of 0.90. The present study does not report the influence of a course on ISS agreement, but shows improved agreement on AIS severity codes. Nowadays, at most places ISS calculation has been computerized, and therefore improving agreement on AIS severity codes will result in better agreement on ISS.

Limitations
This study has several limitations that need to be considered. The participants were tested for their ability to assign a code from the AIS code book to a clearly described, isolated injury. In practice, the reliability and accurateness of injury coding, and finally the results of comparative outcome analysis of trauma data, will also depend on the ability of coders to extract relevant information from patient records. This part of the injury coding process is also part of the training course in injury coding. However, availability and format of patient data will differ between institutions and therefore this issue was not included in the study. The present study assesses a part of the problem of reliable injury coding and therefore comparison of our results with earlier mentioned studies is hampered.

Participants in this study varied in background and experience, some of them had no (para) medical training and lacked anatomical knowledge. Physicians and nurses are known to
A training course for AIS coding

be more reliable in their coding. Therefore, in participants with a medical background or experience in coding the learning effect will possibly be not as strong as we found in our study. However, the number of participants was too small to test this hypothesis in a subgroup analysis.

Despite the fact that the results of this study show improved performance and reliability in injury coding, yet approximately one on five of the unique AIS codes and one on ten of the AIS severity scores was wrongly coded. Further improvement is possibly achieved by case discussions between coders for more severe and compound injuries.

Different conditions of the pre-test, test and post-test may have biased the results. Pre-test and post-test were made at a self-chosen moment without time restraint, while the test was made in a supervised setting with time restraint. Theoretically, yet very unlikely, contact could have taken place between participants to discuss pre–or post-test. The slight drop in correct unique 7-digit codes for the post-test can be explained as a decline in motivation amongst the participants. The post-test consisted of different injuries than the pre-test and test. This was inevitable in order to avoid a strong ‘recall’ effect. We minimized the bias this might have introduced by developing a post-test that was comparable in level of difficulty and with the same AIS severity distribution as the pre-test and test.

The AIS codeset has expanded in size to over 2000 injury descriptors in the 2005 revision. The Dutch trauma registry has not yet adopted the AIS 2005 or 2008 update and therefore we were obliged to use the AIS98 for this study. Various efforts have been made to facilitate mapping of the 1998 to 2005 AIS version. However, this would have obscured the learning effect of the course on the inter-observer agreement.

With respect to the statistical analysis a weighted kappa analysis would have been favourable since the AIS severity code is an ordinal scale. Scrutiny of our data only revealed 7 cases in which the severity code differed more than one rank from the correct code. Hence, we found use of the non-weighted kappa justifiable.

CONCLUSION

Reliability of injury coding is required to improve trauma care and to compare quality of trauma care between different regions and hospitals. The results of this study emphasized the use of training in AIS coding to improve ability to assign correct codes and to reduce variability. It is advisable that all who are involved in injury coding using the Abbreviated Injury Scale are trained by a dedicated course.
Chapter 7

REFERENCES


