



ANZICS



ANZICS
Centre for Outcome
and Resource Evaluation

APD Data Audit Program Report

Cycle 1: 2007-2009

Contents

Acknowledgements	3
Abbreviations.....	3
Ethics	3
Executive Summary	4
Overview of the Data Audit Process	5
Limitations of the Data Audit Process	5
Summary of the 2007-2009 Data Audit Cycle.....	6
Audit Aims	6
Audit Protocol.....	6
Audit Cycle.....	6
How the audit data was analysed.....	7
Missing data.....	7
Inter-observer reliability.....	7
APACHE II score and predicted mortality	7
Exclusions	7
Analysis.....	8
Statistical analysis.....	8
The 2007-2009 Data Audit Cycle: Analysis	9
1.0 Participation	9
2.0 Audited unit demographics	11
3.0 Inter-observer reliability.....	12
4.0 APACHE II scores and predicted Risk of Death	14
4.1 Basic analysis of differences in APACHE II scores and predicted mortality	14
4.2 Variability of individual APACHE II scores and predicted risk of death.	15
4.3 Absolute and actual change in APACHE II scores and predicted mortality by unit.	16
4.4 Relationship between reliability and change in APACHE II score and predicted mortality.....	17
5.0 Changes in APACHE II scores and predicted mortality by variable.....	18
5.1 How often does an unreliable entry cause a score change?	18
5.2 Level of agreement between original and audit APACHE II scores	19
5.3 Absolute versus actual APACHE II score change by variable	20
5.4 Modeling the effect of score change in predicted Risk of Death	21
5.5 Absolute versus actual change in APACHE II predicted mortality by variable.....	22
6.0 Analysis based on hospital classification	23
6.1 Changes in APACHE II scores	23
6.2 Changes in predicted mortality	24
6.3 Impact of different variables in predicted mortality by hospital classification	25

7.0 Impact of ANZICS training on APACHE II scores and predicted mortality	26
7.1 Impact of ANZICS training on changes to APACHE II scores and predicted mortality	26
7.2 Impact of ANZICS training on variable-specific changes to predicted mortality	27
7.3 Impact of a dedicated data collector on variable-specific changes to predicted mortality	28
8.0 SMR Analysis	29
9.0 Missing data	30
10.0 Additional analysis	32
10.1 Comparing APD admission numbers with CCR admission numbers.....	32
10.2 Analysis of APD admission trends over a 4 year period.	33
10.3 Incidence of chronic cardiovascular and chronic respiratory coding in the APD in 2008.	34
General Summary	36
Summary Table of the most problematic variables.....	37
Recommendations for contributing sites	38
Recommendations for CORE	39
Recommendations for the APD Data Audit Program:	39
Appendix 1: List of audit variables.....	40
Appendix 2: Tolerance levels for individual variables	41
Appendix 3: Generation of an APACHE II score	42
Appendix 4: Generation of an APACHE II risk of death.....	43

Acknowledgements

The first cycle of the APD Data Audit Program would not have been completed without the time and effort of the following people:

Peter Hicks, who designed the database that was used to collect and analyse the audit data.

Those involved in the audit pilot study: Jacqueline Martin, Peter Hicks, Catherine Norrish, Shaila Chavan, Carol George, Peter Stow and Graeme Hart.

The auditors: Shaila Chavan, Jessica Andrews, Ruth Carless, Brigit Roberts, Tom Hunter, Christina Parkes, Mardhie Coleman, Doris Lam, and Ann Whitfield.

The data audit sub-committee: Steve Webb, Dan Mullany and Arthas Flabouris.

The co-opted members of the data audit sub-committee: David Pilcher and Graeme Hart.

Michael Bailey and Shaila Chavan, who provided additional statistical analysis.

Kelly Drennan, who provided CCR data.

The ANZICS CORE data audit project team: Ashley Fletcher, Jostein Saethern and Shaila Chavan.

All the ICU directors, NUMs, data managers and data collectors who facilitated the audits.

Abbreviations

ABG	Arterial blood gas
ACT	Australian Capital Territory
ANZICS	Australian and New Zealand Intensive Care Society
APACHE	Acute physiology and chronic health evaluation
APS	Acute physiology score
AP2	APACHE II
APD	Adult Patient Database
CCR	Critical Care Resources
CORE	Centre for Outcome and Resource Evaluation
GCS	Glasgow coma score
HDU	High Dependency Unit
HR	Heart rate
ICU	Intensive Care Unit
MAP	Mean arterial pressure
ROD	Risk of death
RR	Respiratory rate
SMR	Standardised Mortality Ratio

Ethics

Ethics Approval was obtained for the release of these de-identified results through the Alfred Hospital Ethics Committee.

Executive Summary

The ANZICS Centre for Outcome and Resource Evaluation (CORE) Adult Patient Database (APD) Data Audit Program was initiated in 2007 to determine the reliability of data submitted to the APD. The program was designed to run in 3 year cycles, during which 50 units across all hospital classifications and contributing jurisdictions would be visited. The first cycle of the audit program was completed in 2009 and the results from these audits have now been combined to produce this report. It is hoped this report will provide a picture as to the overall reliability of data held by the APD, the impact unreliable data collection may be having on reported APACHE II scores and predicted mortality, and highlight findings that will be important for improving APD data quality in the future.

Overall 41,000 variables, from 1102 ICU admissions to 44 hospitals were audited. While there was no difference in median APACHE II score between the original and audit data, the original data revealed a higher mean APACHE II score by 0.4 (17.2 in original compared to 16.8 in audit). Similarly median and mean predicted mortality were higher in the original data by 1.3% (17.2% versus 15.9%) and 1.8% (26.0% versus 24.2%) respectively. Although these were small differences, they represented a significant bias towards over-estimation of APACHE II scores and predicted risk of death in the original dataset.

This bias was predominantly caused by over coding of chronic health variables for APACHE II. These variables proved to be the most problematic in terms of both unreliable collection and subsequent impact on APACHE II scores and predicted mortality. CORE will be targeting the coding of chronic conditions as a priority to further improve the quality of the ANZICS APD, through training workshops and dedicated site-targeted campaigns.

Finally, of great interest was the finding that units with APD-trained data collection and entry staff showed significantly less variation in both APACHE II scores and predicted mortality when compared with units where staff hadn't attended training. These units also showed less significant bias in their collection of individual variables. Further to this, those unit with trained staff and a dedicated data collector were able to completely eliminate significant bias in their collection of individual variables. This highlights the importance of both APD training and the use of dedicated data collectors within units. We hope these findings will provide units with the evidence they need to gain the resources required to ensure reliable collection of APD data.

We would like to take this opportunity to thank all those involved in the Data Audit Program: the units who agreed to participate, the data collection/entry staff and data managers who facilitated the audits and the auditors who re-extracted the data for analysis. We believe this has been a highly successful program and hope that the findings presented in this report will help guide APD data collection practices in the future.

Allison van Lint, PhD
Senior Project Officer
ANZICS CORE

Dr David Pilcher
APD Director

Associate Professor Graeme Hart
Chair
CORE Management Committee

Overview of the Data Audit Process

ANZICS CORE currently runs a number of automated validity and completeness checks on all data submitted to the APD. While these checks can be used to identify missing and illogical data, they fail to ensure that the data being collected is reliable (i.e., that 2 people would collect the same set of data for any given patient). CORE initiated a data audit program in 2007 to fill this gap in data quality checks. The audit process was based on a pilot study conducted in 2006¹, and was structured to run in cycles of 3 years. It was planned that during every 3 year cycle 50 sites, spread over the 4 hospital classifications, would be audited and every jurisdiction with sites contributing to the APD would be visited.

The first cycle of the audit program was run on an "invitational" basis, ICUs were invited to participate and only those who agreed were audited. Once a unit had agreed to participate, a random selection of 25 admissions that had been previously submitted to the APD by that unit was extracted from the APD central database for audit. These admissions had to have an ICU length of stay of at least 24 hours. A list of these admissions was sent to the unit so that medical records and accompanying test results could be made available to the auditor. The CORE-trained auditor then used the medical records and accompanying results to re-extract those variables required to generate an APACHE II score and predicted risk of death (ROD) for the 25 admissions. Additional demographic variables were also collected in the audit (a full list of audit variables and their relevance to APACHE II can be found in Appendix 1). The re-extracted data was entered into an ACCESS database which had been pre-loaded with the originally submitted data. The database then identified where original data was missing, where original data did not match that entered by the auditor and also generated an APACHE II score and predicted ROD for each patient.

Each audited unit received a report detailing their level of missing and unreliable data. The impact of unreliable data collection on their APACHE II scores and predicted ROD was explained, and their most problematic variables were identified and methods for improving data collection were provided.

2009 saw the completion of the first 3 year cycle of APD data audits. The results of the audits performed during this cycle have been combined and are presented in this report to provide an idea of the impact that unreliable data collection is having on the APACHE II scores and predicted mortality rates reported by the APD, and to identify which variables are most affected by unreliable data collection. When interpreting the results presented in this report one should be aware of several limitations to the audit process as it has been run during this first cycle.

Limitations of the Data Audit Process:

- Only patients with an ICU length of stay of at least 24 hours were included in the audit, it is unclear whether this is producing a bias in the audit results.
- It has not been established whether the 25 randomly selected admissions were a representative sample of each unit's case mix.
- The audit is currently run on an invitational basis - sites are invited to participate and are allowed to decline. This may explain the low representation of private hospitals in this audit cycle. It may also bias participants, in that units who take their data collection seriously and are therefore more likely to have good data collection practices, may be more likely to participate.
- The audit is currently based on APACHE II rather than APACHE III-J. This means that more patients are excluded from the predicted mortality analysis due to non-weighted diagnoses than would be if the audit was based on APACHE III-J.
- For the majority of units (n=32), the audit did not measure the accuracy of recording patient outcomes, therefore standardised mortality ratios (SMRs) could not be analysed.
- Where patient outcomes were collected (12 units), 25 records was too small a sample size to allow meaningful interpretation of SMRs for individual units.

¹ Martin J, Hicks P, Norrish C, Chavan S, George C, Stow P, Hart GK. Designing and implementing an Australian and New Zealand intensive care data audit study. *Int J Health Care Qual Assur* (2009) 22:572

Summary of the 2007-2009 Data Audit Cycle

Audit Aims:

- Identify the level of inter-observer variability present in the data held by the APD.
- Determine the effect this inter-observer variability may be having on APD reported APACHE II scores and predicted mortality.
- Identify variables where inter-observer variability has a large impact on APACHE II scores and predicted mortality.
- Estimate the impact missing data is having on reported APACHE II scores and predicted mortality.
- Determine the impact that ANZICS APD training has on data quality.

Audit Protocol:

- Auditors included CORE staff and external auditors who attended a "Training the Auditor" session.
- Sites were invited to participate in the audit process.
- 25 randomly selected records were audited per site.
- Admissions audited had to have an ICU length of stay of at least 24 hours.
- 38 variables were audited: APACHE II variables plus demographic and admission/discharge variables (see Appendix 1 for a complete list).
- The reliability of the raw data collected was determined.
- An APACHE II score and predicted mortality was generated for both datasets, and compared.

Audit Cycle:

- 44 sites were audited across all regions that contribute to the APD.
- A total of 1102 admissions were audited.
- 9 auditors were used, including 3 CORE staff who performed 84% of the audits.
- ICU admissions from 2006 through to 2008 were audited.
- An average of 25 admissions were audited at each site (min = 21, max = 33).

How the audit data was analysed

This report compares 2 sets of data, that originally submitted to the APD by the unit and that collected by the auditor. From here on, the data originally submitted to the APD is referred to as the "original dataset", while the data collected by the auditors is referred to as the "audit dataset".

A total of 1102 ICU admissions were audited. The section below outlines how the data was treated for various different analyses throughout this report:

Missing data:

- Missing data in the APD is treated as normal and, if a scoring variable is missing, it generates an APACHE II score of zero for that variable. Throughout the analysis, when analysing individual variables, where a difference between the 2 datasets was caused by a missing value in the original dataset, the difference was included in the analysis. However, where a difference was caused by a missing value in the audit dataset, the difference was excluded from the analysis. For example, if the original heart rate was missing and the audit heart rate had a score of 2, then this difference of -2 was included in the analysis. However, if the original heart rate had a score of 2 and the audit heart rate was missing, then this difference of 2 was excluded from the analysis.

Inter-observer reliability:

- To determine inter-observer reliability, a tolerance level was set for each variable such that if the difference between the original entry and the audit entry was greater than the allowed tolerance then the entry was deemed "unreliable" (see Appendix 2 for a list of tolerance levels).
- If an original value was missing and the corresponding audit value was present, the original value was deemed unreliable. If the original value was present and the audit value was missing then this entry was excluded from the reliability analysis.
- Percentage agreement was determined by taking the number of reliable entries/total entries x 100. The number of total entries for each variable excluded entries where the audit value alone was missing (as described above).

APACHE II score and predicted mortality:

Exclusions:

- 17 patient admissions in the original dataset were missing all their physiology, and therefore did not generate an APACHE II acute physiology score (APS). Patients without an APS do not generate an APACHE II score or predicted ROD in the APD. These patients were excluded from all analysis, therefore there were 1085 admissions included in the inter-observer reliability analysis, the APACHE II score analysis and the analysis of missing data.
- The APD currently asks that patients be given an APACHE III-J diagnosis which is then mapped to the relevant APACHE II diagnosis. The APACHE II diagnosis code 306 (haematologic undefined) is used when patients are given an APACHE III-J haematological diagnosis. The APACHE II diagnosis code 308 (no APACHE II equivalent) is used when patients are given an APACHE III-J diagnosis code that has no APACHE II equivalent. Both 306 and 308 have no diagnostic weighting and do not generate an APACHE II predicted ROD. There were 112 patients coded as either 306 or 308 in either the original or audit datasets, these patients did not generate a predicted ROD and were excluded from all predicted mortality analysis. In addition, 4 patient admissions in the original dataset were missing a diagnosis and therefore did not generate a predicted ROD. These patients were also excluded from the predicted mortality analysis. Therefore there were 969 admissions included in all APACHE II predicted mortality analysis.

Analysis:

- The method used for generating an individual APACHE II score is described in Appendix 3.
- The method for calculating an individual predicted ROD is described in Appendix 4.
- In all analysis, when looking at differences between the 2 datasets, the audit value for an admission was subtracted from the original value for that admission. In this way, a positive difference indicates the original value was higher, while a negative difference indicates that the audit value was higher.
- When presenting predicted ROD/mortality results, for individuals the results are presented as predicted ROD (here a value of 0.2 represents a 20% risk of mortality), for group analysis the results are presented as predicted mortality (total predicted ROD/number of admissions x 100) so that results presented represent the predicted mortality per 100 admissions.

Statistical analysis:

- Box plots (or box and whisker plots) have been used to present a range of different data in this report. A box plot shows the minimum, the lower quartile, the median, the upper quartile and the maximum of a dataset. Each segment on a box plot represents 25% of the data.
- Kappa analysis was performed on the individual variable scores from the original and audit datasets. Kappa analysis determines the level of agreement between 2 datasets that is not due to chance. Kappa should lie between 0 and 1 where 1 is perfect agreement and 0 is the level of agreement expected by chance (a negative kappa indicates a level of agreement less than chance). The following scale is generally accepted:

Kappa <0 = Less than chance agreement

0.01 - 0.20 = Slight agreement

0.21 - 0.40 = Fair agreement

0.41 - 0.60 = Moderate agreement

0.61 - 0.80 = Substantial agreement

0.81 - 0.99 = Almost perfect agreement

- Where data showed a normal distribution, the t-test was used to determine whether the differences between original and audit data were significantly different to zero. A *p* value less than 0.05 was considered significant.
- Where data did not have a normal distribution, the Wilcoxon signed rank test was performed on differences between the 2 datasets to determine if the differences were significantly different from zero. A *p* value less than 0.05 was considered significant.
- The F-test (for equality of 2 variances) was used to determine whether 2 datasets, both with normal distribution, showed homogeneity of variance. A *p* value less than 0.05 was considered significant.

The 2007-2009 Data Audit Cycle: Analysis

1.0 Participation

The 2007-2009 data audit cycle audited ICU admissions from 2006 through to 2008. In most cases the audited admissions at each site were taken from a 6 month time period, however in some cases this extended to a 12 month period. In all, a total of 1102 admissions were audited from 44 different units across all contributing jurisdictions. The target of 50 units was not reached, most likely as a result of staff changeovers at CORE which resulted in the audit program being suspended on 2 occasions.

As shown in Table 1.1, in the smaller jurisdictions such as the ACT, Tasmania and Northern Territory, 50-100% of contributing sites were audited in this first audit cycle, while in the larger jurisdictions of Victoria, New South Wales and Queensland between 22-29% of sites were audited. It is anticipated that it will take a further 2 audit cycles (an additional 6 years) for all sites within these larger jurisdictions to be audited if the audit process continues as currently structured.

Table 1.1 A breakdown of the units audited during the 2007-2009 audit cycle by jurisdiction.

Region	Number of units audited	% APD contributing units audited	Time period audited
Australian Capital Territory	2	100%	2008
New South Wales	10	23%	2006
Northern Territory	2	100%	2008
South Australia	4	40%	2007/8
Tasmania	1	50%	2008
Queensland	7	22%	2008
Western Australia	4	100%	2007/8
Victoria	9	29%	2006
New Zealand	5	56%	2008

An initial aim of the audit program was to audit units from all 4 hospital classifications; metropolitan, tertiary, rural/regional and private. The spread of audited units across these classifications is shown in Table 1.2. As is immediately obvious, a very low number of private hospitals (n=4) participated in the audit process. This may be a bi-product of the "invitational" nature of the audit project. As a result of their case mix private hospitals tend to have lower APACHE II scores and predicted ROD as compared to the other 3 hospital classifications (shown later in Sections 6.1 and 6.2). As will be shown in Figure 5.3, changes in low APACHE II scores have less impact on predicted ROD than changes to intermediate scores do. Therefore if one of the aims of this audit is to identify variables which are causing changes to our reported predicted ROD then the low representation of private hospitals may not be a huge issue. However, as an additional aim is to identify the reliability of data across all contributing units, future audit cycles should encourage all hospital types to participate at an equal rate.

Table 1.2 A breakdown of the units audited during the 2007-2009 audit cycle by hospital classification.

Hospital Classification	Number of units audited	% Contributing units audited
Metropolitan	15	55%
Tertiary	15	48%
Rural/Regional	10	32%
Private	4	11%

As outlined at the start of this report, 25 ICU admissions were audited at each site. The yearly admissions received by the audited units varied greatly and as a result, the percentage of admissions audited at each site also showed a high level of variation, as shown in Figure 1.1.

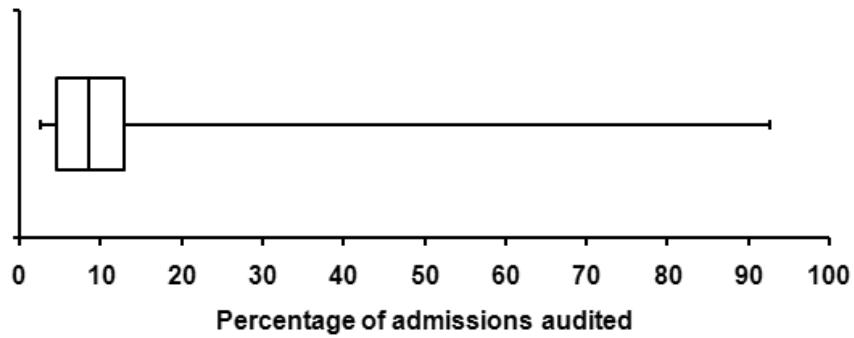


Figure 1.1 The percentage of admissions audited at each site.

The percentage of admissions that had been audited at each site was determined based on the number of admissions each site had submitted to the APD for the audit period. The box plot shows the minimum (2.5), lower quartile (4.6), median (8.6), upper quartile (12.9) and maximum (92.6) percentage of admissions audited across the 44 sites.

2.0 Audited unit demographics

Audited units were asked to fill out a survey regarding their data collection and entry personnel, their training and their access to APD data collection tools. 80% of audited units completed the survey, the results are presented below.

Table 2.1 Results from the Data Audit Survey.

How many data collectors does your unit have?		
1 dedicated collector	2 or more collectors	No assigned collected
28.5%	68.5%	3%

Who collects and enters the data in your unit?*		
	Data Collection	Data Entry
Clinician	16	2
Nurse	13	13
Data Manager	7	5
Administrative personnel	6	13
Ward Clerk	5	7
Registrar	1	0

How do your data collectors/data entry personnel get trained?*		
	Data Collector	Data Entry
ANZICS	4	6
In house	12	8
Trained by previous	10	10
Combination of ANZICS/in house/previous	9	9
No training	3	4
8 units had both data collection and entry staff who had attended ANZICS APD training		

Access to data collection tools	
Units with access to the APD data dictionary	85%
Units using AORTIC (ANZICS-supplied data collection software)	84%
<i>Units using AORTIC for most recent APD submission (2009 Q4) = 88%</i>	

* These results are presented as whole numbers rather than percentages because some units selected more than 1 option.

3.0 Inter-observer reliability

One of the initial aims of the data audit project was to investigate the level of inter-observer reliability of data held by the APD. Inter-observer reliability refers to the rate with which 2 different data collectors would collect the same data for any given patient - in this case the level of reliability between the original data and the audit data was analysed.

The data audit measured inter-observer reliability by measuring the percentage agreement between the original entries and the audit entries across 38 different variables (see Appendix 1). For each variable a tolerance level was set such that if the difference between the original entry and the audit entry was greater than the allowed tolerance then the entry would be deemed "unreliable" (see Appendix 2 for a list of tolerance levels). The percentage agreement across the entire original and audit datasets was determined (Table 3.1), as was the percentage agreement between original and audit entries for each hospital classification (Table 3.1) and each individual unit (Figure 3.1). The 17 admissions missing all their physiological data in the original dataset were excluded from this analysis. Individual variable entries where the original value was available but the audit value was missing were also excluded, as there was no way of determining whether the original value was reliable in these cases.

Table 3.1 Percentage Agreement of all admissions and by hospital classification.

Percentage Agreement across all admissions (n=1085)	
81.9%	
Hospital Classification	Percentage Agreement
Metropolitan (n=367)	82.0%
Tertiary (n=365)	81.9%
Rural/Regional (n=252)	81.7%
Private (n=101)	82.1%

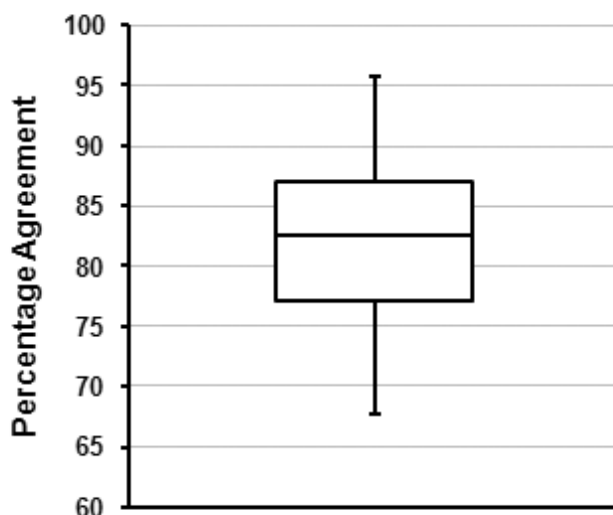


Figure 3.1 The percentage agreement of entries by unit.

The percentage agreement across all 38 variables was determined for each individual audited unit. The box plot shows the minimum (67.7%), lower quartile (77.1%), median (82.5%), upper quartile (87.0%) and maximum (95.8%) percentage agreement for the 44 audited units.

The percentage agreement for each individual variable was also determined (Table 3.2). Of interest among the scoring variables, chronic cardiovascular and chronic respiratory had lower agreement than the other chronic health variables, while the physiology variables as a group (respiratory rate, temperature, sodium etc) showed the lowest levels of agreement, indicating that these are the variables most affected by inter-observer variability.

Table 3.2 The percentage agreement of each audit variable divided into scoring and non-scoring variables and listed in order of agreement.

Variable	Percentage Agreement
<i>Demographic and other non-scoring variables</i>	
Sex	98.71
ICU Admission Date	96.22
Hospital Admission Date	94.38
Intubated	93.52
ICU Discharge Date	93.36
Ventilated	90.65
ICU Admission Time	84.50
Hospital Admission Time	83.55
ICU Discharge Time	68.92
<i>APACHE II scoring variables</i>	
Age	98.99
Chronic Liver	98.99
Chronic Renal	98.15
Immune Disease	98.06
Immune Suppressed	97.88
Elective	93.98
Acute Renal Failure	93.78
ICU Source	92.44
Chronic Respiratory	90.96
Chronic Cardiovascular	88.48
GCS Motor*	86.46
GCE Eye*	84.63
GCS Verb*	83.01
Diagnosis	82.95
White Cell Count	77.36
Heart Rate	76.18
Potassium	74.35
Mean Arterial Pressure	73.49
Haematocrit	71.06
Glasgow Coma Score	70.43
Temperature	66.17
Sodium	66.17
Creatinine	66.11
FiO ₂	65.83
pH	65.48
paO ₂	64.88
Respiratory Rate	64.55
paCO ₂	61.39
Oxygenation (FiO ₂ /paO ₂ /paCO ₂)	57.42
Bicarbonate*	53.32

* The analysis of these variables was treated differently to the rest. Any entry where an original OR audit value was missing was excluded from the analysis. This is because these variables are not always required for calculating an APACHE II score and some units leave these variables blank when it is assumed they are not required.

4.0 APACHE II scores and predicted Risk of Death

4.1 Basic analysis of differences in APACHE II scores and predicted mortality

Of the 1102 admissions audited, 1085 qualified for an APACHE II score in both the original and audit datasets. An analysis of the differences in these scores is shown in Table 4.1. A total of 969 audited admissions qualified for a predicted Risk of Death (ROD) in both datasets and this analysis is shown in Table 4.2.

Table 4.1 APACHE II scores.

	Number of admissions	Total APACHE II score	Median APACHE II score	Mean APACHE II score	Difference in mean Score
Original	1085	18611	16	17.15	0.39
Audit		18185	16	16.76	

Exclusions:

17 admissions missing all physiology in the original dataset

Table 4.2 Predicted Mortality.

	Number of admissions	Total APACHE II predicted ROD	Median APACHE II predicted mortality	Mean APACHE II predicted mortality	Difference in mean predicted mortality
Original	969	251.58	17.15%	25.96%	1.79%
Audit		234.25	15.90%	24.17%	

Exclusions:

17 admissions missing all physiology in the original dataset

4 admissions missing a diagnosis in the original dataset

112 admissions that had a non-weighted diagnosis in either/both the original and audit datasets (APACHE II diagnosis codes 306 and 308)

While there was no difference in median APACHE II score between the original and audit data, the original data revealed a higher mean APACHE II score by 0.4 (17.2 in the original compared to 16.8 in the audit). Similarly median and mean predicted mortality were higher in the original data by 1.3% (17.2% versus 15.9%) and 1.8% (26.0% versus 24.2%) respectively. Although these were small differences, they represented a significant bias in the original dataset towards over-estimation of APACHE II scores ($p = 0.01$) and predicted risk of death ($p = 0.0001$).

4.2 Variability of individual APACHE II scores and predicted risk of death.

Correlation analysis was performed on both the individual APACHE II scores and predicted ROD from each dataset (data not shown). Both score and predicted ROD analysis showed a strong positive correlation ($r = 0.834$ and 0.857 respectively), but also demonstrated that variation existed between the 2 datasets. To ascertain where the variation between the datasets was most pronounced, Bland-Altman analysis was performed on both the APACHE II scores and predicted ROD. Bland-Altman analysis graphs the difference between an individual patient's score or predicted ROD in the original and audit dataset against the mean of the 2 values, and identifies where the greatest variability between the datasets is present. In Figure 4.1, the greatest variability in APACHE II scores appears to occur in admissions with high APACHE II scores (mean of 25-40), and the majority of these admissions had a higher score in the original dataset than in the audit dataset (points above the x axis). Figure 4.2 indicates that the greatest variability in predicted ROD occurred in patients with an intermediate predicted ROD (mean of 0.5).

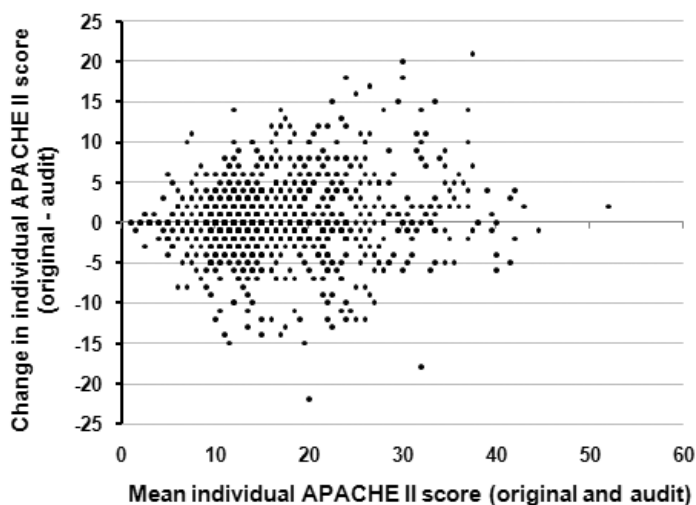


Figure 4.1 Bland-Altman analysis of individual APACHE II scores.

The difference between individual APACHE II scores in the original and audit datasets (y axis) was plotted against the mean of the 2 scores (x axis), $n=1085$.

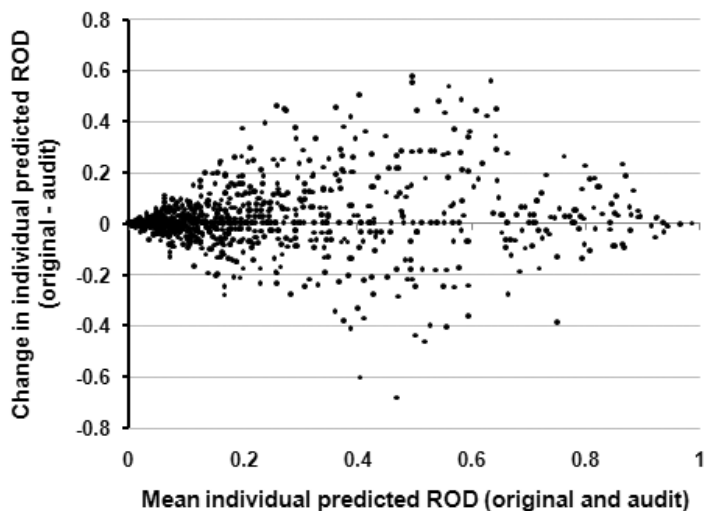


Figure 4.2 Bland-Altman analysis of individual predicted ROD.

The difference between individual predicted ROD in the original and audit datasets (y axis) was plotted against the mean of the 2 RODs (x axis), $n=969$.

4.3 Absolute and actual change in APACHE II scores and predicted mortality by unit.

Given that variability clearly existed within the datasets, the absolute and actual differences in APACHE II score and predicted mortality were determined for each individual unit, to identify the extent to which different units were contributing to this variability (Figure 4.3). The change in APACHE II score (A) is shown as the average change per patient for each unit, while the change in predicted mortality (B) is shown as the change in predicted mortality across all patients within each unit.

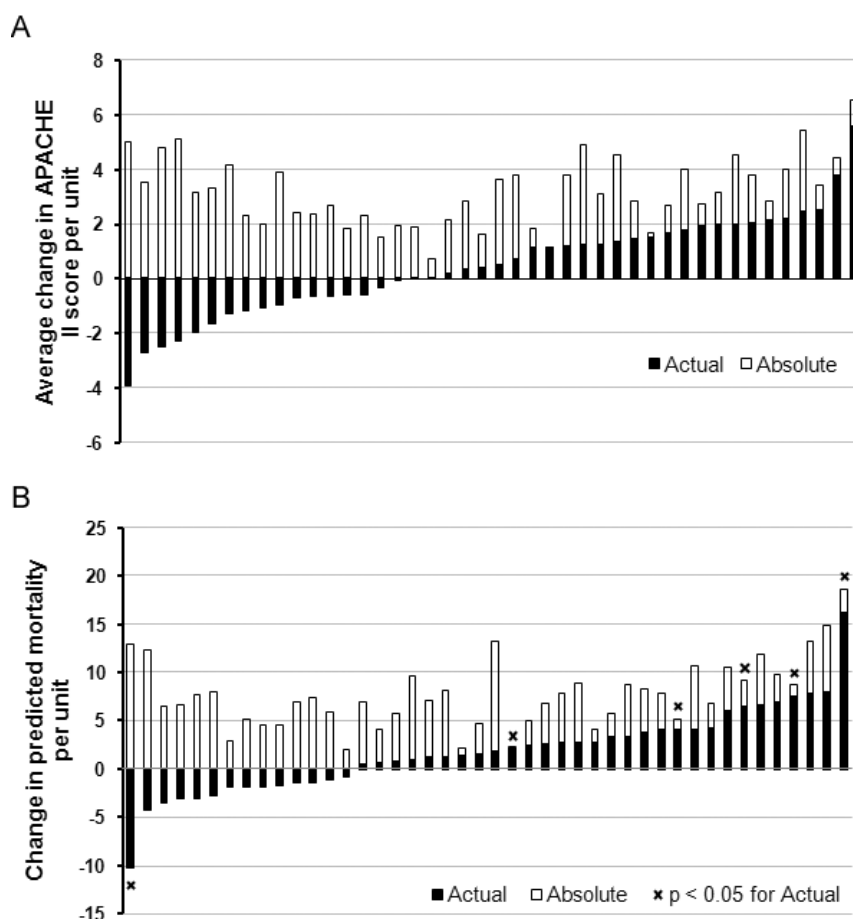


Figure 4.3 Actual and Absolute change in APACHE II score and predicted mortality by unit.

(A) The average actual and absolute difference in APACHE II score per patient was determined for each audited unit. The 1085 admissions with an APACHE II score in both the original and audit datasets were included in this analysis. (B) The actual and absolute change in predicted mortality was determined for each audited unit. The 969 admissions with predicted ROD in both the original and audit datasets were included in this analysis. In each figure, each bar represents a different unit. The actual difference for a unit was determined by leaving individual differences as positive and negative and adding them together (where the original value was higher the difference would be positive, where the audit value was higher the difference would be negative). The absolute difference for a unit was determined by treating all individual differences as positive and adding them together. The Wilcoxon signed rank test was used to determine if the actual change in ROD was significantly different to zero. Units with a p value less than 0.05 (considered significant) are indicated.

Of interest:

- 61% of units (n=27) over-estimated their APACHE II scores when compared with the audit dataset.
- 68% of units (n=30) over-estimated their predicted mortality when compared with the audit dataset.
- 9% of units (n=4) had a change in predicted mortality of less than 1%.
- 80% of units (n=35) had a change in predicted mortality of less than 5%.
- 4.5% of units (n=2) had a change in predicted mortality of greater than 10%.
- 87% of units with an average actual score change of at least +2 per patient (7 out of 8 units), also had a predicted mortality change greater than 5%.
- Only 25% of units with an average actual score change of at least -2 per patient (1 out of 4 units), also had a predicted mortality change of greater than 5%.

4.4 Relationship between reliability and change in APACHE II score and predicted mortality

The level of correlation between reliability (percentage agreement) and change in either APACHE II score or predicted mortality was determined (Figure 4.4). As shown, the correlation coefficients (r) were quite small for each analysis, indicating that the level of correlation in both cases was low. However, even though the correlation between increasing percentage agreement and decreasing change in APACHE II score and ROD was small, Figure 4.4 does show that as percentage agreement increases, the variability in APACHE II score and ROD decreases.

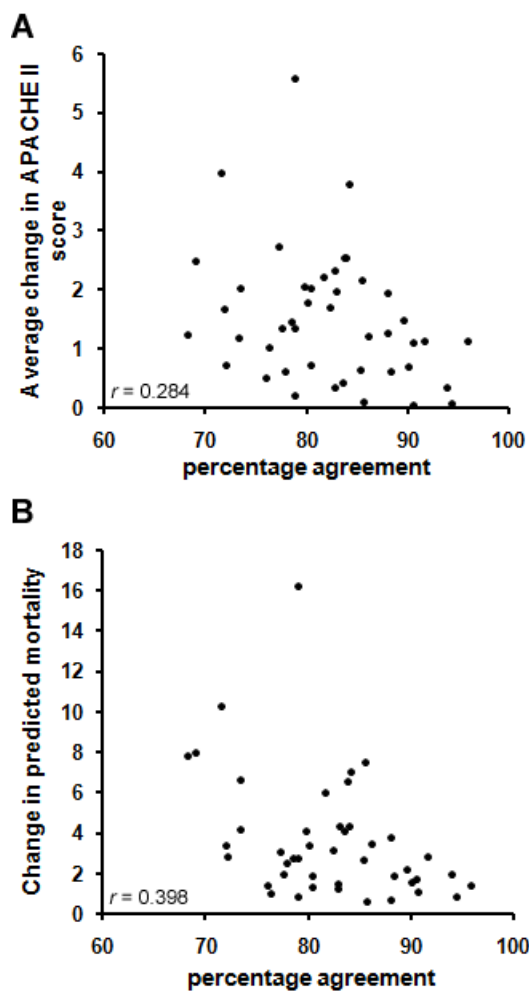


Figure 4.4 Scatter plot of percentage agreement against average change in APACHE II score and change in predicted mortality.

(A) Percentage agreement versus average actual change in APACHE II score for individual units, (B) percentage agreement versus actual change in predicted mortality for individual units. For this analysis all actual changes were treated as positive to allow any trends to become evident.

5.0 Changes in APACHE II scores and predicted mortality by variable

5.1 How often does an unreliable entry cause a score change?

The number of times an original variable entry was found to be unreliable was compared to the number of times a score difference occurred (Figure 5.1). Of interest, variables such as chronic health status, heart rate, mean arterial pressure and Glasgow coma score almost always produced a score change when the entries in the original and audit datasets were different. Alternatively, sodium almost never produced a score change despite the fact that it was often unreliable.

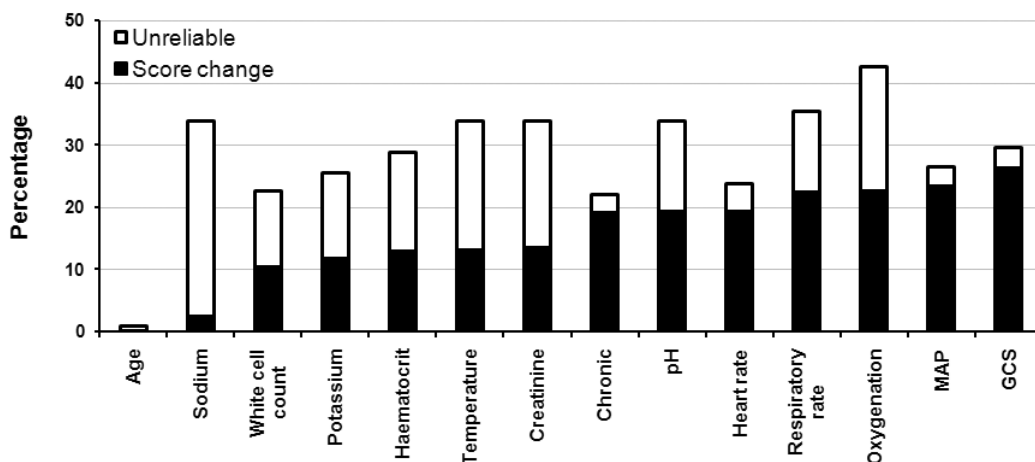


Figure 5.1 The percentage of unreliable entries compared to the percentage of entries that produced a score change between the original and audit datasets.

The percentage of unreliable entries was compared to the percentage of entries that caused a score change. Unreliable oxygenation entries are a combination of FiO_2 , paO_2 and $paCO_2$, unreliable pH entries are a combination of pH and bicarbonate (bicarbonate was used where pH was not available).

5.2 Level of agreement between original and audit APACHE II scores

Kappa analysis was performed on the individual variable scores to determine the level of agreement between the original and audit scores (Table 5.1). This gave an indication not only of how many score differences were present but the magnitude of these differences as well. Chronic conditions, for example, did not have the most score differences (Figure 5.1), but since the majority of these score differences resulted in a score change from 0 to 5, this variable had the lowest kappa coefficient, with moderate agreement (Table 5.1).

Table 5.1 Level of agreement between APACHE II scores for individual variables

Variable	Kappa coefficient	Level of Agreement
Age	0.997	Almost perfect
White cell count	0.804	Substantial
Temperature	0.773	Substantial
Sodium	0.764	Substantial
Creatinine	0.753	Substantial
Haematocrit	0.751	Substantial
Potassium	0.710	Substantial
pH	0.691	Substantial
Heart rate	0.680	Substantial
Respiratory rate	0.669	Substantial
Oxygenation	0.633	Substantial
MAP	0.559	Moderate
GCS	0.549	Moderate
Chronic	0.507	Moderate

5.3 Absolute versus actual APACHE II score change by variable

The actual difference in APACHE II score between the original and audit datasets was 426 (see Table 4.1), while the absolute difference was 3422. These differences were then broken down across the APACHE II scoring variables, to determine the actual and absolute score change for each variable (Figure 5.2). Where the actual change is above the x axis it indicates that the original dataset had a higher score and where the actual change is below the x axis it indicates that the audit dataset had a higher score. Of interest, chronic health status had the largest actual score change, with patients consistently having higher chronic health scores in the original dataset when compared to the audit dataset. GCS had the highest absolute change in APACHE II score, however most of these differences balanced themselves out over the 1085 patients to produce an actual change of just -15. Seven variables had an actual score change that was significantly different to zero, suggesting that there was a bias in the original dataset towards over-estimating (creatinine, pH, and chronic health status), or under-estimating (respiratory rate, mean arterial pressure, heart rate, and oxygenation) these scores when compared with the audit dataset.

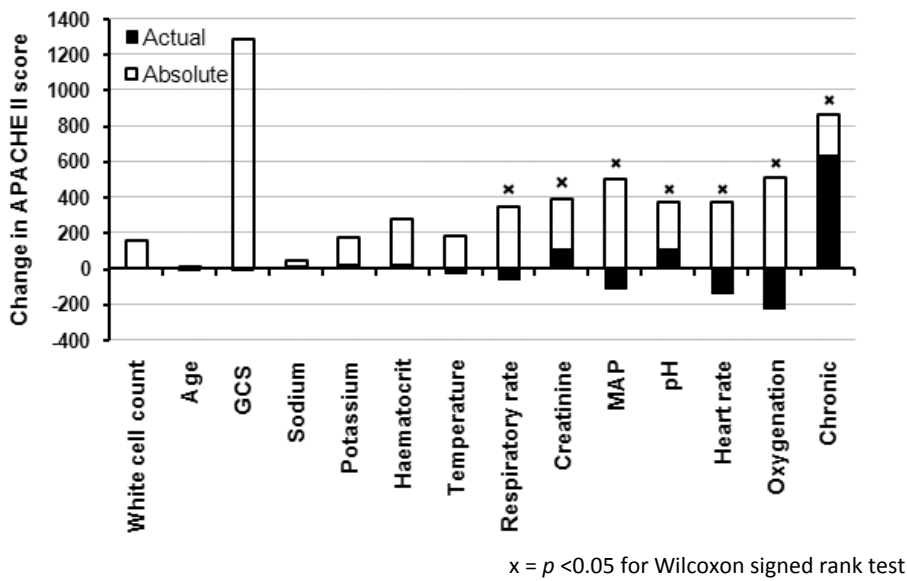


Figure 5.2 Absolute and actual change in APACHE II score by variable.

The difference in APACHE II score for each admission in the 2 datasets was determined for each variable. The absolute change was determined by treating all differences as positive and adding them together. The actual difference was determined by leaving the individual differences as positive or negative and adding them together. The variables are graphed in increasing order of actual change in APACHE II score.

5.4 Modeling the effect of score change in predicted Risk of Death

To get an idea as to the impact a change in APACHE II score has on predicted mortality, the change in predicted ROD was modeled as the most problematic variables changed score (the variables modeled were those with the largest absolute change in APACHE II score from Figure 5.2). From this modeling (Figure 5.3) a few conclusions can be drawn:

- Changes in APACHE II scores have the biggest impact on ROD when patients have an intermediate score, as seen by the increased gradient of the lines for medium scores. This possibly explains why the biggest variability in predicted ROD between the 2 datasets was seen in patients with an intermediate ROD (Figure 4.2).
- Changes in diagnosis have the biggest impact on patients with intermediate APACHE II scores and patients with high scores who change between a medium and low diagnosis weighting.
- Changes to APACHE II scores and diagnosis have little impact on predicted ROD when patients have a low APACHE II score, or have a high APACHE II score and a medium to high diagnosis weighting.

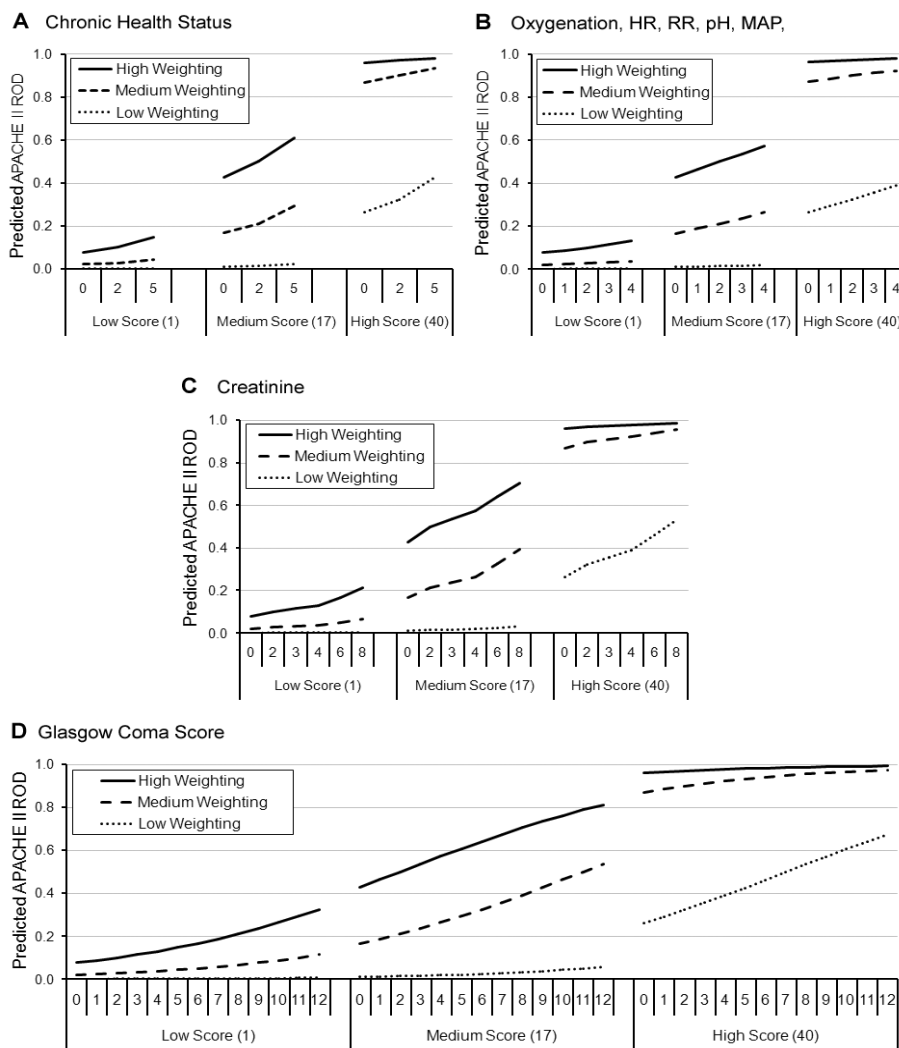


Figure 5.3 Modeling of change in predicted ROD as different variables change in score.

The change in predicted ROD was modeled based on changes in score to the most problematic variables; (A) Chronic health status which has scoring options of 0, 2 and 5, (B) heart rate (HR), respiratory rate (RR), pH, oxygenation and MAP which range in score from 0-4, (C) Creatinine which has scoring options of 0, 2, 3, 4, 6 and 8, and (D) Glasgow Coma Score which has a scoring range of 0-13. The modeling was based on a high diagnostic weighting (code 107, weight 0.891, the highest diagnostic weighting found in the original and audit datasets), a medium weighting (code 110, weight -0.424, the median weighting found in the original and audit datasets) and a low weighting (code 122, weight -3.535, the lowest weighting found in the original and audit datasets). This was combined with a low score (1, the lowest score in the original and audit datasets), a medium score (17, the median score found in the original and audit datasets), and a high score (40, a score of 40 with a GCS of 13 reaches the highest score found in the original and audit datasets, 53).

5.5 Absolute versus actual change in APACHE II predicted mortality by variable

The effect each variable was having on predicted mortality was determined. The variables APACHE II diagnosis, elective and ICU source were included in this analysis as they impact on predicted ROD. The absolute and actual change in predicted mortality is shown for each variable (Figure 5.4). Chronic health status appeared to be having the largest impact on predicted mortality, increasing it by 1.2%. GCS had the greatest potential for impacting on predicted mortality, with an absolute difference of almost 3%, however these differences seem to balance out over the entire dataset and resulted in an actual change in predicted mortality of just 0.2%. Six variables had an actual change in predicted mortality that was significantly different to zero, suggesting that there was a bias in the original dataset towards over-estimating (elective, creatinine, pH, and chronic health status), or under-estimating (heart rate and oxygenation) the predicted mortality associated with these variables when compared with the audit dataset.

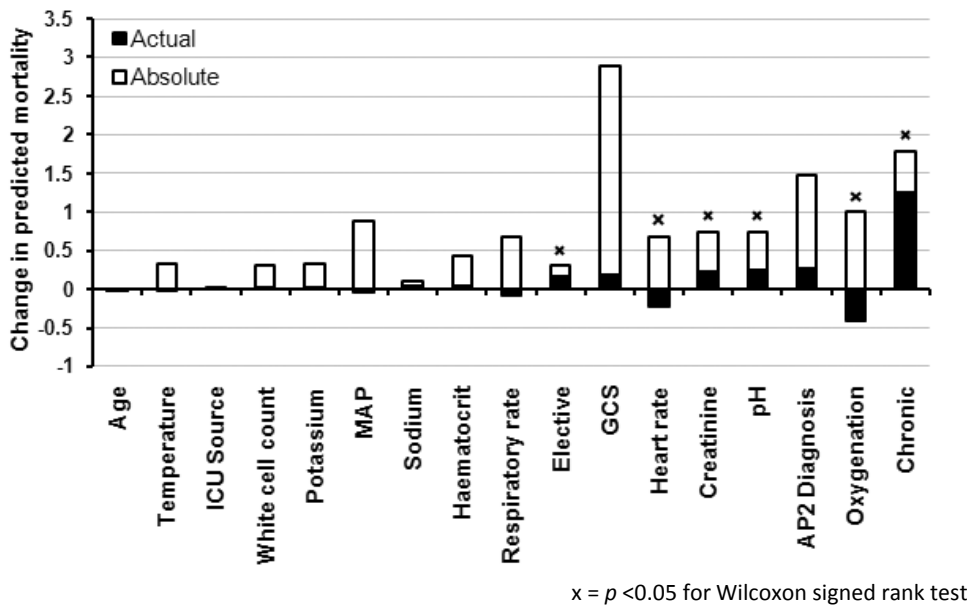


Figure 5.4 Absolute and actual change in predicted mortality based on variable.

The variable scores in the original dataset were replaced one variable at a time with the scores from the audit dataset. As each set of original scores were replaced by audit scores the predicted ROD for each patient was recalculated. The actual and absolute differences between the original predicted ROD and the recalculated predicted ROD produced by each variable was then determined. The predicted mortality was determined based on the 969 patients who had a predicted ROD in both the original and audit datasets. The variables are graphed in increasing order of actual change in predicted mortality.

6.0 Analysis based on hospital classification

6.1 Changes in APACHE II scores

The mean APACHE II score for admissions in the original and audit dataset were analysed based on hospital classification (Table 6.1). The mean change in APACHE II score between admissions in the 2 datasets is also shown. It appeared that metropolitan, rural/regional and private units were tending to over-estimate their scores, while tertiary units tended to under-estimate their scores, when compared with the audit dataset.

Table 6.1 APACHE II score differences by Hospital classification.

Hospital Classification	Original mean APACHE II score	Audit mean APACHE II score	Mean change in individual APACHE II score
Metropolitan (n=367)	17.4	16.5	0.85
Tertiary (n=365)	17.6	18.1	-0.57
Rural/Regional (n=252)	17.7	16.6	1.09
Private (n=101)	13.6	13.1	0.47

The spread of change in APACHE II scores was analysed based on hospital classification (Figure 6.1). The median change in individual APACHE II scores across all hospital classifications was zero. As demonstrated by the boxes in Figure 6.1, 50% of admissions within each hospital classification clustered tightly around the median showing very little change in APACHE II scores between the original and audit datasets. Indeed, 80% of admissions across all hospital classifications had an APACHE II score change of 5 or less (data not shown).

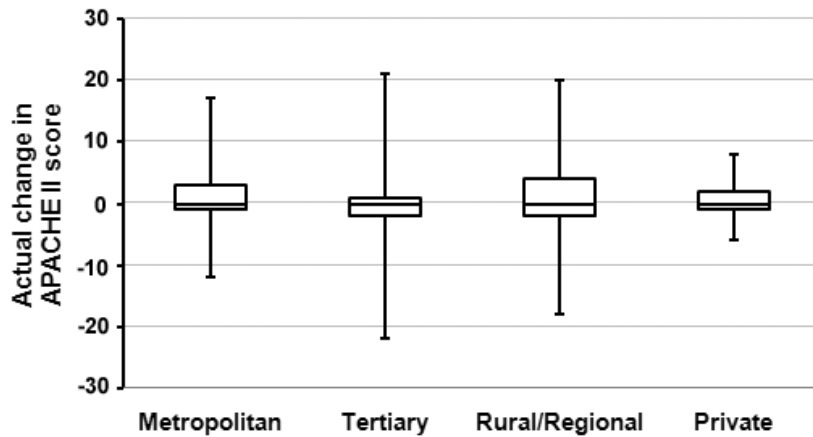


Figure 6.1 Change in individual APACHE II scores by hospital classification.

The change in APACHE II score for each individual admission was determined for all admissions within each hospital classification. The box plot shows the minimum, lower quartile, median, upper quartile and maximum actual change in APACHE II score for each hospital classification. Each segment of the box plot represents 25% of admissions. Metropolitan n=367, Tertiary n=365, Rural/Regional n=252, Private n=101.

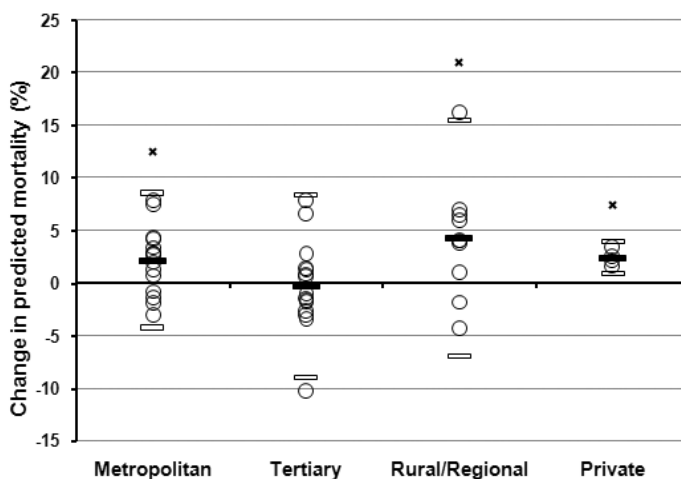
6.2 Changes in predicted mortality

Table 6.2 shows the original and audit predicted mortality for all admissions within each hospital classification and the difference between the two. As with the APACHE II score analysis, the audit suggested that metropolitan, rural/regional and private units were tending to over-estimate their predicted mortality, while tertiary units were tending to under-estimate.

Table 6.2 Change in predicted mortality by hospital classification.

Hospital classification	Original predicted mortality (%)	Audit predicted mortality (%)	Difference
Metropolitan (n=344)	25.36	23.24	2.12
Tertiary (n=306)	26.67	27.05	-0.38
Rural/Regional (n=239)	29.68	25.65	4.03
Private (n=80)	14.55	12.14	2.41

Figure 6.2 shows the change in predicted mortality for all units based on hospital classification. Metropolitan, rural/regional and private units all had mean changes in predicted mortality that were significantly different to zero (t-test, $p < 0.05$), suggesting a bias among these units towards over-estimation of predicted mortality. There was no significant bias in the tertiary units, with a balance between units over-estimating and under-estimating their predicted mortality as compared to the audit results. Private units showed less variation in predicted mortality than the other 3 classifications, however this may be due to the small number of private units audited. Both the rural/regional and tertiary classifications had a unit that sat outside 2 standard deviations from the mean of that group.



x = $p < 0.05$ for t-test

Figure 6.2 Change in predicted mortality based on hospital classification.

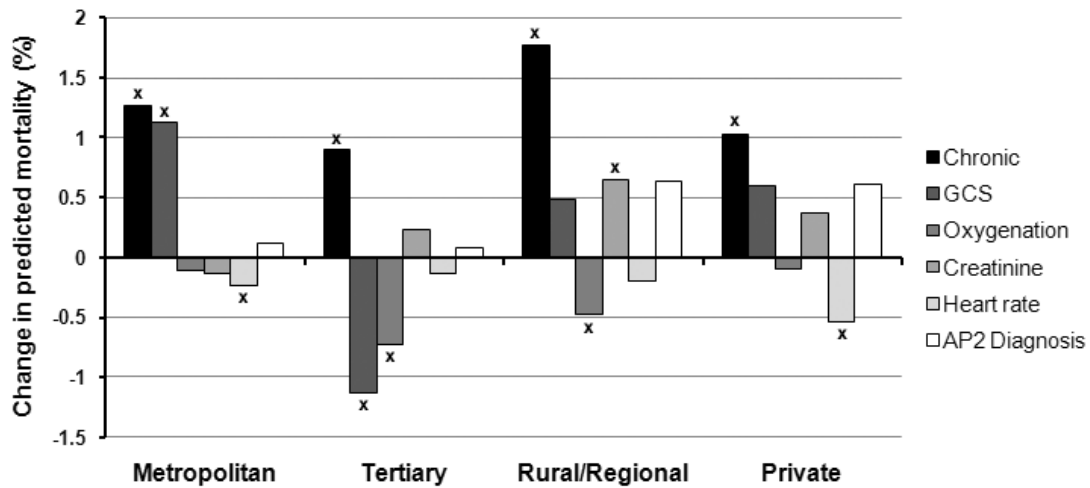
The change in predicted mortality was determined for each unit and is presented based on hospital classification. Each open circle represents a unit, the solid line represents the mean and the open line represents 2 standard deviations from the mean. Metropolitan n=15 units, Tertiary n=15, Rural/regional n=10, Private n=4.

6.3 Impact of different variables in predicted mortality by hospital classification

The impact of each variable on predicted mortality was determined for each hospital classification. Those that produced a change in predicted mortality of at least 0.5% in any one hospital classification are shown in Figure 6.3.

Of interest:

- Chronic health status appeared to be increasing the predicted mortality of units within all 4 classifications (as compared to audit results), most evident in rural/regional units. The difference in predicted mortality caused by chronic health status was significantly different to zero for all 4 hospital classifications, suggesting a consistent bias across units in all hospital classifications to over-code chronic health co-morbidities.
- Tertiary units were the only units where GCS was found to be decreasing predicted mortality (as compared to audit results). The change produced by GCS was significantly different to zero for both the metropolitan and tertiary units, suggesting a bias in metropolitan units to over-estimate GCS scores and a bias in tertiary units to under-estimated GCS scores.
- APACHE II diagnosis appeared to be increasing the predicted mortality in both rural/regional and private units (as compared to audit results), however the difference was not significantly different to zero for either classification.



x = $p < 0.05$ for Wilcoxon signed rank test

Figure 6.3 The change in predicted mortality produced by the different variables for the different hospital classifications.

The change in predicted mortality produced by the different APACHE II variables was determined for each hospital classification. The variables that produced a change in predicted mortality of at least 0.5% in any one hospital classification are shown. Metropolitan n=331, Tertiary n=319, Rural/Regional n=239, Private n=80.

7.0 Impact of ANZICS training on APACHE II scores and predicted mortality

7.1 Impact of ANZICS training on changes to APACHE II scores and predicted mortality

The data audit survey results (Table 2.1) allowed the audited units to be grouped according to whether or not their data collection and entry personnel had attended ANZICS APD training. These groupings were then used to ascertain whether attending ANZICS APD training impacted on changes to APACHE II scores and predicted mortality (Figure 7.1). As shown, data from units where collection and entry staff had attended ANZICS APD training had a significantly smaller spread of change in both APACHE II scores and predicted mortality, compared to units where staff had not attended training.

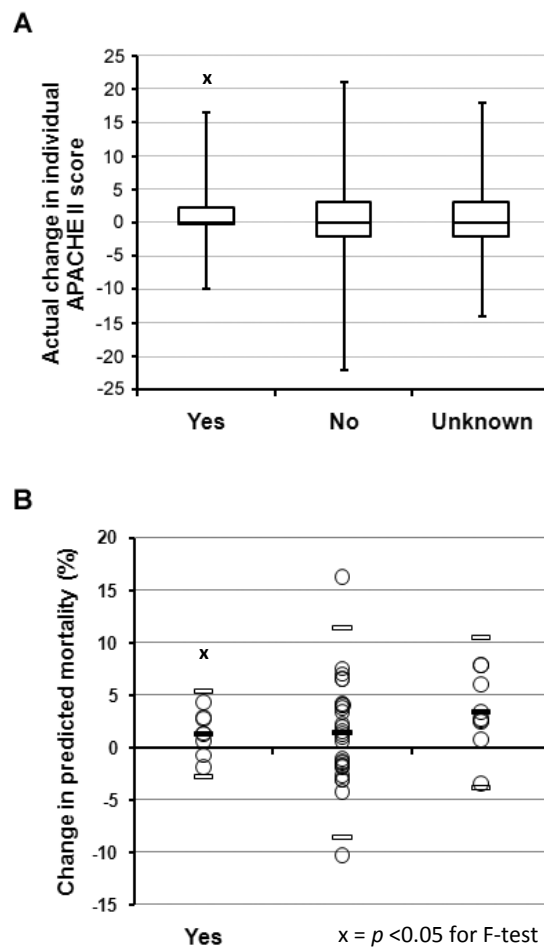
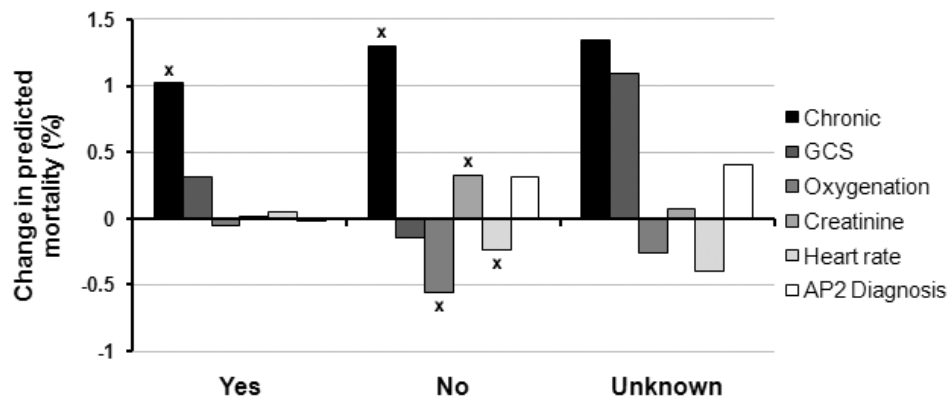


Figure 7.1 Impact of ANZICS APD training on the change in APACHE II score and predicted mortality.

Audited units were grouped according to whether their data collection and entry personnel had attended ANZICS APD training (yes, n=193), had not (no, n=676), or whether this information had not been collected (unknown, n=216). (A) The actual change in individual APACHE II scores was determined for each group. The box plot shows the minimum, lower quartile, median, upper quartile and maximum actual change in APACHE II score for each group. (B) The actual change in predicted mortality for each unit was determined. Open circles represent individual units, the solid line represents the mean and the open lines represent 2 standard deviations from the mean. Yes n=8 units, No n=27, Unknown n=9. Statistical analysis was performed on the "yes" and "no" groups only.

7.2 Impact of ANZICS training on variable-specific changes to predicted mortality

The bias in the collection of individual variables was determined based on whether or not staff from units had attended ANZICS APD training, the results are presented in Figure 7.2. The results show a bias towards over-estimating the chronic health contribution to predicted mortality regardless of whether staff had attended ANZICS APD training or not. Interestingly, chronic health was the only variable that reached significance for units with staff that had attended training. Conversely, in the group where staff had not attended training 3 additional variables had changes in predicted mortality that were significantly different to zero: oxygenation, creatinine and heart rate.



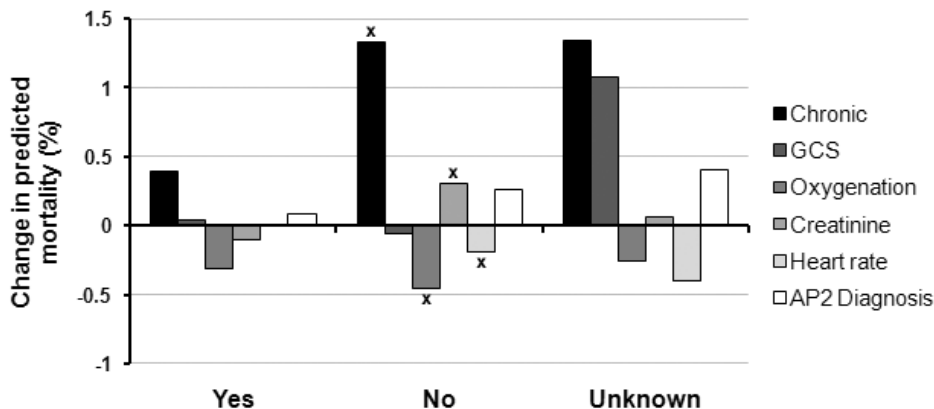
x = $p < 0.05$ for Wilcoxon signed rank test

Figure 7.2 Impact of ANZICS training on the change in predicted mortality by variable.

Audited units were divided into 3 groups; those where data collection and entry personnel had attended ANZICS training (yes, n=175), those where they had not (no, n=599), and those where this information was not collected (unknown, n=195). The actual change in predicted mortality produced by different variables was determined for each group, the 6 variables with the largest changes are shown. Statistical analysis was performed on the "yes" and "no" groups only.

7.3 Impact of a dedicated data collector on variable-specific changes to predicted mortality

Units were grouped based on whether they had a dedicated data collector in combination with ANZICS APD-trained data collection and entry personnel (Figure 7.3), there were only 4 units that satisfied this criteria. Statistical analysis of this "yes" group indicated that no variables produced a change in predicted mortality that was significantly different to zero, suggesting that no data collection bias was consistent across these units. When the "no" group was analysed it was found that chronic health status, oxygenation, creatinine and heart rate all produced a change in predicted mortality that was significantly different to zero. This suggested a consistent bias among these units to over-estimate chronic health and creatinine, and to under-estimate oxygenation and heart rate.



x = $p < 0.005$ for Wilcoxon signed rank test

Figure 7.3 Impact of a dedicated data collector and ANZICS training on the change in predicted mortality produced by different variables.

Audited units were divided into 3 groups; those with a dedicated data collector and collection/entry personnel who had attended ANZICS training (yes, n=82), those who did not meet the 'yes' criteria (no, n=692), and those where this information was not collected (unknown, n=195). The actual change in predicted mortality produced by different variables was determined for each group, the 6 variables with the largest changes are shown. Statistical analysis was performed on the "yes" and "no" groups only.

8.0 SMR Analysis

During the audit of 2 jurisdictions, the hospital outcome of each audited patient was also collected. This was then compared with the hospital outcomes held in the APD central database. A total of 295 admissions were included in this analysis, Table 8.1 shows the percentage of patients in each category, the percentage agreement for each category and the differences found between the original and audit hospital outcomes.

Table 8.1 Reliability of hospital outcomes.

Hospital outcome (n=295)	Percentage of admissions	Percentage agreement	Issues
Died in hospital (n=35)	11.9	100	
Discharged home alive (n=197)	66.8	98	4 differences, audit found all 4 were transferred to a rehab facility
Transferred to rehabilitation facility (n=28)	9.5	96	1 difference, audit found patient was discharged home
Transferred to other ICU (n=4)	1.3	50	2 differences, audit found both were transferred to another acute hospital
Transferred to other acute hospital (n=30)	10.2	90	3 differences, audit found 2 were discharged home and 1 was transferred to a rehab facility
Missing (n=1)	0.3	NA	

The Standardised Mortality Ratios (SMRs) were then determined for these units from the original and audit data, based on those patients who qualified for inclusion in an APACHE II SMR analysis (patients with a predicted ROD and a hospital outcome other than "transfer to another ICU"). The results (not shown) demonstrated that the sample sizes were so small that the confidence intervals made any interpretation meaningless. Therefore, the data from all 12 units was combined and an original and audit SMR determined (Figure 8.1). As shown, the audit SMR was higher than the original SMR, however the 95% confidence intervals showed a high degree of overlap.

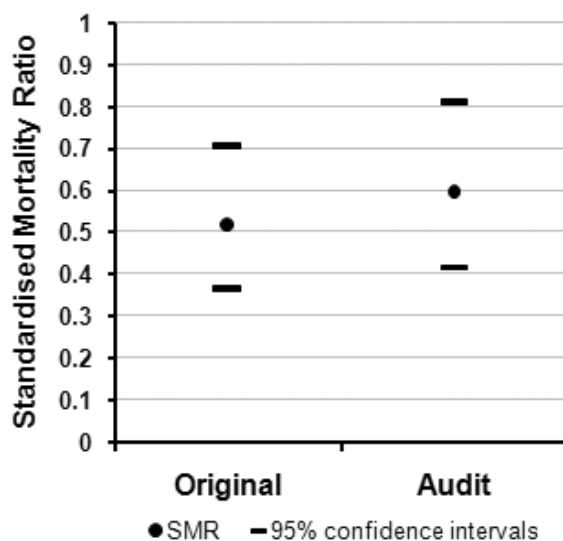


Figure 8.1 Original and audit SMR for 12 audited units combined.

The data from 12 audits was combined to determine an overall original SMR (n=260) and audit SMR (n=262). The 95% confidence intervals are also shown.

9.0 Missing data

The amount of data missing from the original and audit datasets was determined and is shown in Table 9.1. The amount of missing data common to both datasets is also shown, along with the level of missing data in the APD central database for 2006 through to 2008 (the years covered by this audit cycle). The 17 ICU admissions missing all physiology data in the original dataset were included in this "missing" analysis, therefore 1102 admissions were analysed. In all, across the 27 variables, 3.0% of data was missing from the original dataset. For the most part, the spread of missing data from the original dataset was comparable with the spread of missing data from the central database, suggesting that the audited records were representative of the central database in terms of missing data.

Table 9.1 Percentage of data missing from the original and audit datasets and the APD central database.

Variable	Missing from original dataset	Missing from audit dataset	Missing data common to both datasets	Missing from APD central database (2006-2008)*
Demographic and other non-scoring variables				
Hospital Admission Time	4.6	0.9	0.1	0.8
Ventilated	3.2	0.6	0.1	2.0
Intubated	4.6	0.6	0.1	2.3
Hospital Admission Date	2.3	0	0	0.3
ICU Discharge Time	0	0.4	0	0
ICU Admission Date	0	0.1	0	0
ICU Admission Time	0	0.1	0	0
ICU Discharge Date	0	0	0	0
Sex	0	0	0	0.04
APACHE II scoring variables				
Oxygenation	19.5	15.3	13.4	20.8
Glasgow Coma Score	6.9	0.4	0.1	6.6
Haematocrit	5.6	1.5	0.8	8.1
pH (+ bicarbonate)	5.3	2.9	1.9	5.5
White cell count	4.1	2.9	0.6	4.1
Potassium	3.9	0.5	0.4	3.7
Sodium	3.9	0.6	0.4	3.7
Creatinine	3.8	1.3	0.6	4.2
Respiratory Rate	3.4	1.2	0	3.7
Mean Arterial Pressure	3.4	0.6	0	3.1
Heart rate	3.3	0.6	0	4.1
Temperature	3.3	0.6	0	3.0
Acute Renal Failure	0.9	0.7	0	0.2
ICU Source	0.7	0.1	0.1	0.2
Elective	0.1	0.5	0	0.8
Diagnosis	0.4	0	0	1.0
Age	0	0.1	0	0
Chronic Health Status	0	0.1	0	0
Number of admissions analysed	1102	1102	1102	190,979
Total % missing data	3.0	1.2	0.7	2.9
% admissions missing ALL physiology	1.5	0	-	1.7

* The data analysed from the APD central database excluded ICU Length of stay <24 hrs and Age < 16 to match the audit data.

Of the variables listed in Table 9.1, 85% in the original dataset and 96% in the audit dataset were at least 95% complete. The variables with the highest levels of incompleteness in the original dataset were oxygenation and Glasgow Coma Scores. Of the missing oxygenation variables in the original dataset (19.5%), 13.4% were also found to be missing by the auditors. This suggests that in the majority of cases, when oxygenation was missing in the original dataset it was due to results being unavailable to the data collector (either because tests were not performed or results were not available). In contrast, of the 6.9% missing Glasgow coma scores in the original dataset only 0.1% were also found to be missing by the auditors. This suggests that the majority of missing Glasgow coma scores in the original dataset were actually available to data collectors but failed to be collected or entered (this could be explained by sites leaving the GCS fields blank when a patient has a normal GCS, as this results in the APD treating the GCS as normal).

When admissions with missing data are submitted to the APD the missing variables are treated as normal and, if a scoring variable is missing, it generates an APACHE II score of zero for that variable. Given that data missing from the original dataset was not always "found" by the auditors, it is not possible to use the audit data to accurately determine the true value that should be assigned to each missing variable in the original dataset. However, it has been determined that where an entry was missing in the original dataset but present in the audit dataset, this missing data had a total APACHE II score of 397. If this was added to the total APACHE II score for the original dataset it would increase the mean APACHE II score by 0.36, from 17.15 to 17.51, thus increasing the difference between the original and audit datasets. The impact on predicted mortality has not been determined at this time.

10.0 Additional analysis

Several additional avenues of analysis were explored in conjunction with the audit analysis. These focused on two different areas: firstly, to ascertain whether the APD was receiving all admissions from submitting units and secondly, to determine the extent of over-coding of chronic conditions across all submitting units.

10.1 Comparing APD admission numbers with CCR admission numbers.

The number of admissions submitted to the APD by each unit for the 2007/2008 financial year was compared to the number submitted to the Clinical Care Resources (CCR) Registry (the CCR collects data based on financial year). A total of 124 units submitted data to both the APD and the CCR during the 2007/2008 financial year, 65% of these units showed very consistent admission numbers between the 2 databases (Figure 10.1A). The remaining 35% showed differing levels of inconsistency between APD and CCR admission numbers (Figure 10.1B). An initial analysis was performed to try and explain the differences seen (as shown in Figure 10.1B), however further analysis is required.

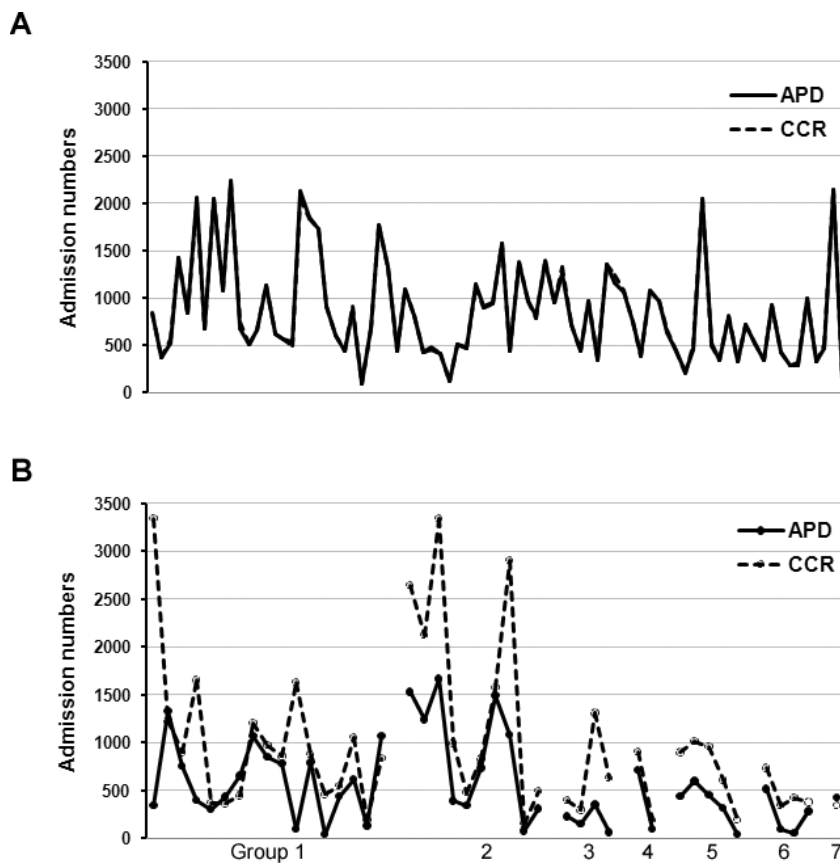


Figure 10.1 Comparison of APD and CCR admission numbers for 2007/2008.

(A) The 80 units that showed consistent admission numbers between the APD and CCR. (B) The 43 units that showed inconsistent admission numbers between the APD and CCR. These units have been grouped according to possible explanations for the inconsistencies (x axis); (1) differences could not be explained at this time, (2) HDU admissions were not submitted to the APD, (3) data is missing from the APD, (4) units started submitting to the APD after July 2007, (5) units stopped submitting to the APD prior to June 2008, (6) data quality issues prevented the APD from accepting data, (7) unit may not be submitting HDU admissions to the CCR.

10.2 Analysis of APD admission trends over a 4 year period.

The number of quarterly admissions submitted to the APD by individual units was determined for the previous 4 years (2006 through to 2009). These numbers were then graphed to identify any unusual trends in quarterly admission numbers. There were 126 different units included in the analysis. Most units showed consistent admission numbers (as demonstrated in Figure 10.2A), however almost 15% of units showed some level of inconsistency in their admission numbers (as demonstrated in Figure 10.2B). This suggests that inconsistencies do exist in APD submission numbers and this type of basic analysis may be a useful tool for identifying those units that require follow-up.

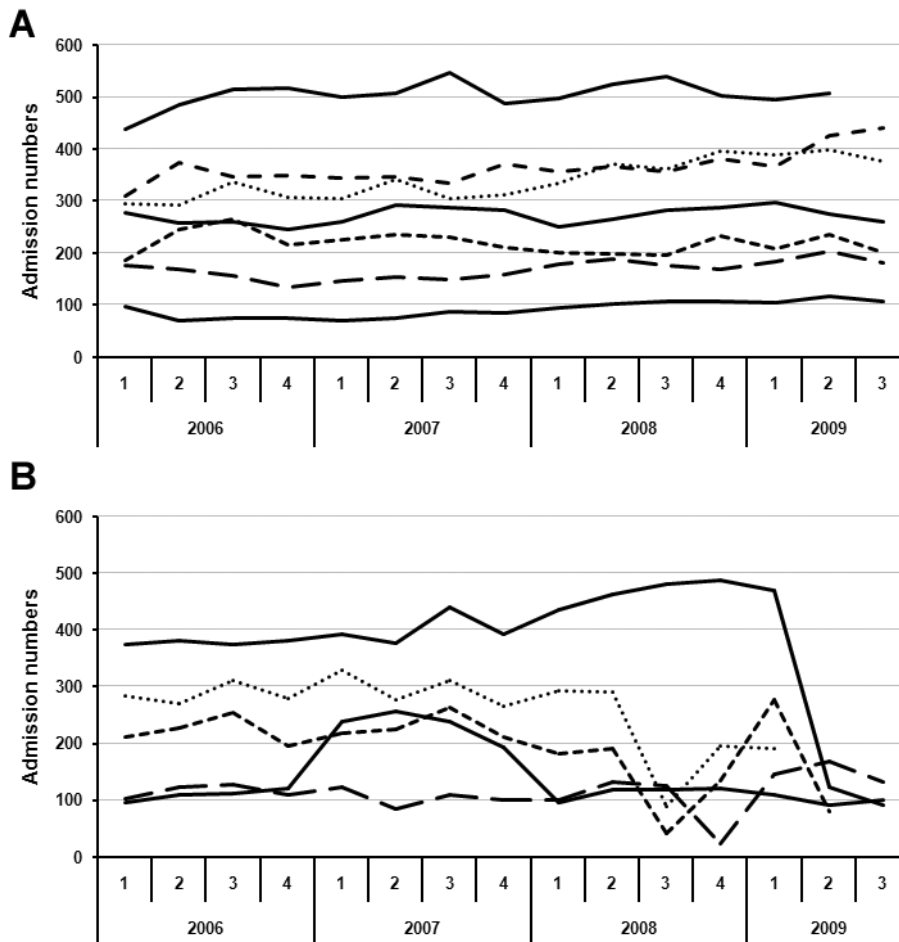


Figure 10.2 Quarterly admission numbers submitted to the APD for 2006 to 2009.

(A) Example of individual units where the quarterly admission numbers to the APD remain consistent over time. (B) Example of individual units where the quarterly admission numbers to the APD show inconsistencies. In each graph each line represents an individual unit.

10.3 Incidence of chronic cardiovascular and chronic respiratory coding in the APD in 2008.

Analysis of the audit results indicated that chronic health status had a large impact on APACHE II scores (Figure 5.2) and predicted mortality (Figure 5.4). The analysis also indicated that, compared to the audit dataset, most audited units were over-estimating the incidence of APACHE II co-morbidities. Given that chronic cardiovascular and chronic respiratory were found to be the most unreliable of the APACHE II co-morbidities (Table 3.2), the reported incidence of these conditions was determined across all units that submitted data to the APD for 2008, the most current data included in this audit cycle. The units were then divided into hospital classification and the proportion of patients coded as chronic cardiovascular (Figure 10.3) and chronic respiratory (Figure 10.4) per unit was compared to the average proportion per unit for each classification as determined by the audit data. The audit average for chronic coding per unit was determined using audits performed by the four main auditors (those who performed more than 1 audit each). It is clear from Figures 10.3 and 10.4 that many units are over-coding both chronic cardiovascular and chronic respiratory co-morbidities.

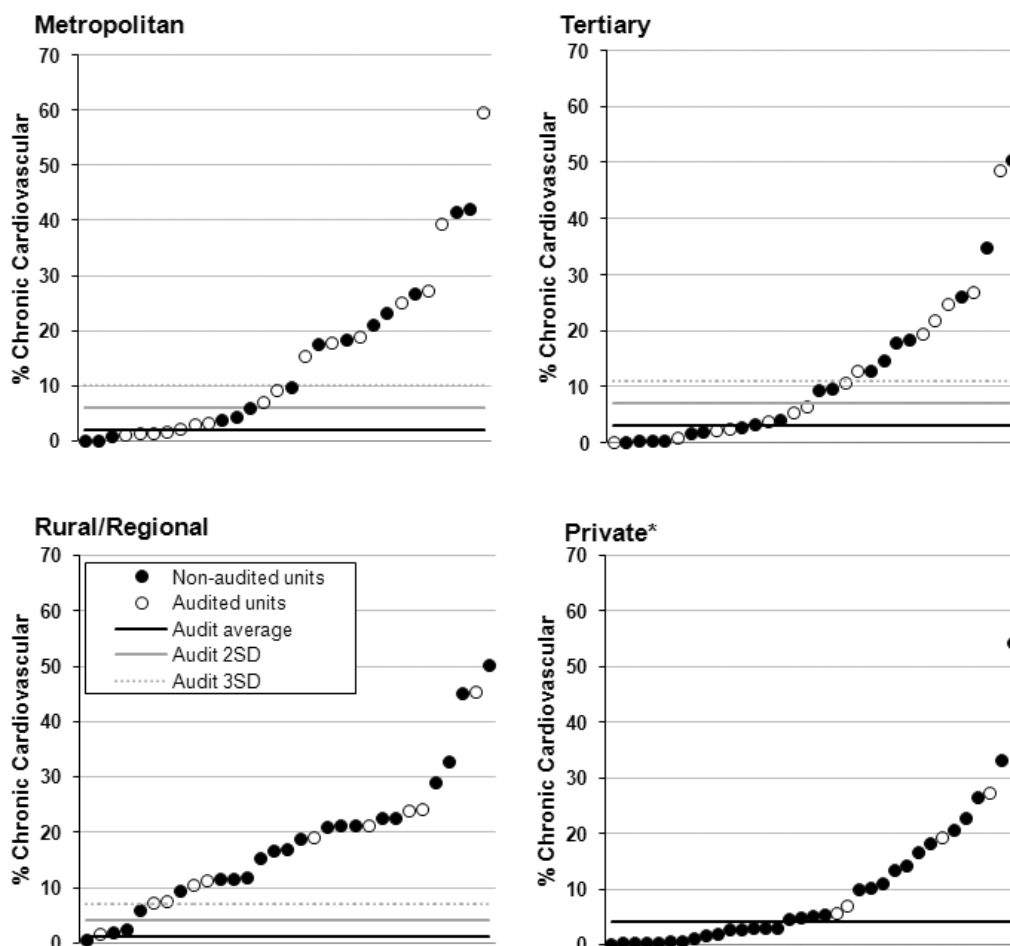


Figure 10.3 Level of chronic cardiovascular coding by hospital classification.

The proportion of patients coded as chronic cardiovascular per unit was determined for data submitted to the APD during 2008. Units were grouped by hospital classification. Closed circles represent units that did not participate in this first cycle of the data audit program, open circles represent units that did participate, the solid black line represents the average proportion of patients per unit coded as chronic cardiovascular in the audit dataset, the solid grey line represents 2 standard deviations from the average, the broken line represents 3 standard deviations from the average. *The 4 audited private units all had the same proportion of chronic cardiovascular coding, therefore there was no standard deviation.

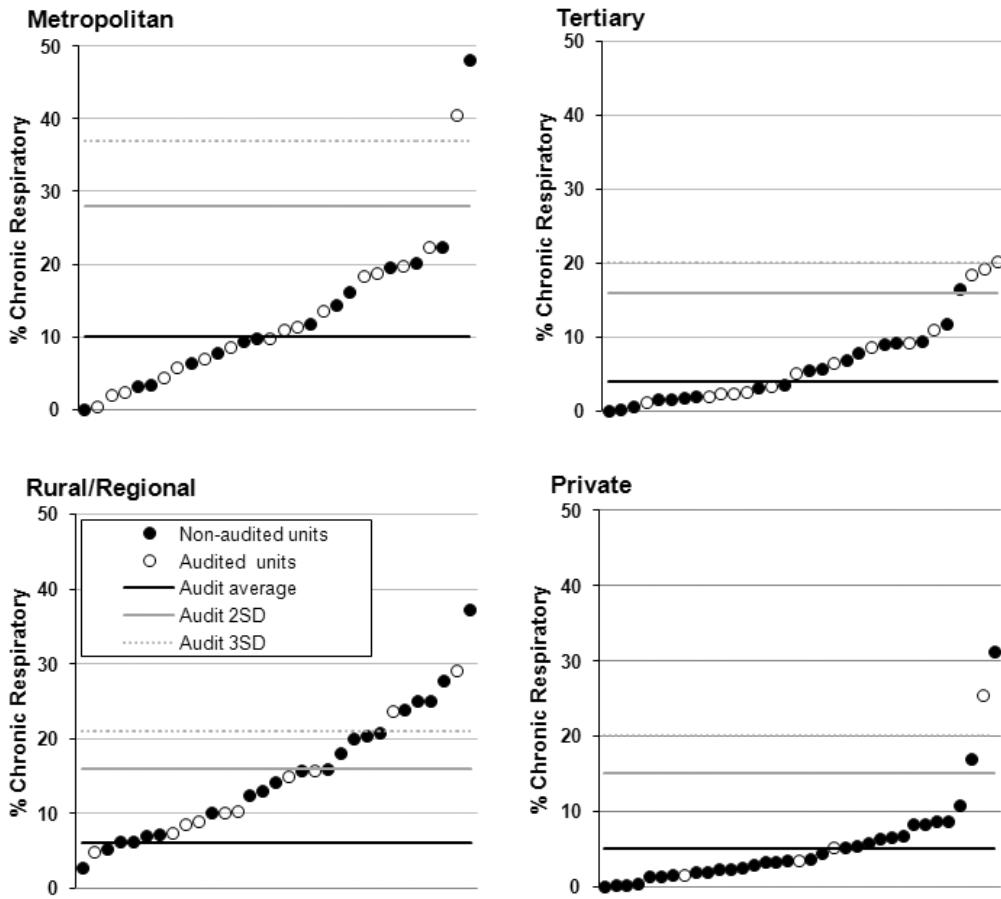


Figure 10.4 Level of chronic respiratory coding by hospital classification.

The proportion of patients coded as chronic respiratory per unit was determined for data submitted to the APD during 2008. Units were grouped by hospital classification. Closed circles represent units that did not participate in this first cycle of the data audit program, open circles represent units that did participate, the solid black line represents the average proportion of patients per unit coded as chronic respiratory in the audit dataset, the solid grey line represents 2 standard deviations from the average, the broken line represents 3 standard deviations from the average.

General Summary:

- 1102 admissions were audited.
- 17 admissions (1.5%) were excluded from the reliability, score and predicted mortality analysis because all physiology was missing from the original dataset - auditors were able to find all physiology data for these admissions.
- 116 additional admissions (10.5%) were excluded from all predicted mortality analysis because these patients had a non-weighted APACHE II diagnosis or were missing their diagnosis in either the original or audit dataset, and did not generate a predicted ROD.
- The percentage agreement across all variables was 82%.
- Highest percentage agreement was Age (98.9%), lowest percentage agreement was Oxygenation (57.4%).
- Variable with the highest Kappa statistic was Age (0.997), with the lowest was Chronic health status (0.507).
- No difference in median APACHE II score, change in mean APACHE II score of 0.4.
Change in median predicted mortality of 1.3%, change in mean predicted mortality of 1.8%.
These differences were small but represented a significant bias in the original dataset as a whole towards over-estimation of APACHE II scores ($p = 0.01$) and predicted mortality ($p=0.0001$), as compared to the audit dataset.
- 20% of units had a difference in predicted mortality between the original and audit data that was greater than 5%, all but 1 of these units was found to be over-estimating their predicted mortality. Given that the predicted mortality of the dataset as a whole was approximately 25%, a change of 5% to an individual unit's predicted mortality could have a large impact on that unit's SMR.
- 13% of units ($n = 6$) had a significant bias in their predicted mortality, all but 1 of these units was found to be over-estimating their predicted mortality when compared to the audit results.
- Metropolitan, rural/regional and private units had mean changes in predicted mortality that were significantly different to zero, suggesting a bias within these hospital classifications to over-estimate predicted mortality.
- Units that had data collection and entry personnel who had attended ANZICS APD training showed significantly less variability between original and audit APACHE II scores and predicted mortality (when compared to units where staff had not attended ANZICS training).
- Units with ANZICS-trained data collection and entry personnel were also less likely to show a significant bias in their collection of individual variables.
- Units with a dedicated data collector (combined with ANZICS-trained staff) showed no significant bias in the collection of individual variables - this was the only analysis performed where no such bias as seen.
- The over-reporting of chronic health co-morbidities, specifically chronic cardiovascular and chronic respiratory conditions, had the biggest impact on disease severity scores and predicted mortality.
- The variables most often associated with an over-estimation of predicted mortality: chronic health, creatinine, diagnosis, pH, and elective admission.
- The variables most often associated with an under-estimation of predicted mortality: oxygenation and heart rate.
- GCS was equally associated with over-estimation (metropolitan, rural/regional and private units) and under-estimation (tertiary units) of predicted mortality.
- Arterial blood gas results were the variables most often missing from both the original and audit datasets, suggesting these variables are the ones most often unavailable to data collectors.
- GCS was the variable most often missing from the original dataset but found by the auditor, suggesting this variable is the one most often missed (or not entered) by data collectors.

Summary Table of the most problematic variables

		Chronic health status	Glasgow coma score	Oxygenation	Heart rate	Creatinine	pH	Mean arterial pressure	Respiratory rate	Elective
Across entire dataset	Most often unreliable			✓						
	Most often produced a score change		✓							
	Largest actual score change	✓								
	Significant change in APACHE II score	✓		✓	✓	✓	✓	✓	✓	
	Score over-estimated	✓				✓	✓		✓	
	Score under-estimated			✓	✓			✓		
	Significant change in predicted mortality	✓		✓	✓	✓	✓			✓
	Predicted mortality over-estimated	✓				✓	✓			✓
	Predicted mortality under-estimated			✓	✓					
Significant change in Predicted Mortality (hospital classification)	Metropolitan	✓	✓		✓					
	Tertiary	✓	✓	✓						
	Rural/regional	✓		✓		✓				
	Private	✓			✓					
Significant change in Predicted Mortality (ANZICS training)	Training	✓								
	No training	✓		✓	✓	✓				
Significant change in Predicted Mortality (dedicated collector + training)	Dedicated collector									
	No dedicated collector	✓		✓	✓	✓				

Recommendations for contributing sites

- The results of the audit analysis indicate the importance of trained, dedicated data collectors. It is recommended that all sites have dedicated data collectors and highlight this need at a jurisdictional level.
- Sites should ensure that staff involved in APD data collection and entry attend APD training.
- All sites should participate in the APD data audit program.
- Sites should focus on improving their collection of problematic variables:
 - chronic health co-morbidities
 - Chronic cardiovascular and chronic respiratory coding should be closely monitored*
 - All data collectors should be made aware of the relevant definitions*
 - Cardiovascular = NYHA class IV (angina or symptoms at rest or on minimal exertion)*
 - Respiratory = Chronic restrictive, obstructive or vascular respiratory disease resulting in severe exercise restriction (unable to climb stairs or perform household duties) OR documented chronic hypoxia, hypercapnia, secondary polycythaemia, severe pulmonary hypertension or respiratory dependence.*
 - Oxygenation and pH
 - Where possible, all Arterial Blood Gas results during the first 24 hours in ICU should be collected*
 - Glasgow Coma Score
 - The GCS should not be obtained when a patient is sedated*
 - For sedated patients, the pre-sedation GCS needs to be determined (e.g., pre-theatre/during retrieval)*
 - Creatinine, Heart rate and MAP
 - The highest and lowest values during the first 24 hours of ICU admission need to be collected*
 - If results are not available, results from 1 hour prior to ICU admission can be used*

Recommendations for CORE

- The results of the audit analysis indicate the importance of trained, dedicated data collectors. These findings should be highlighted at a jurisdictional and unit level, and training should be made available to all jurisdictions.
- Emphasis should be placed on improving the collection of problematic variables: chronic health co-morbidities, oxygenation, creatinine, pH, GCS, heart rate, MAP and elective status. This should include:
 - Re-developing the definitions of these variables to ensure they are easily understood by data collectors. This is particularly important for the chronic health, oxygenation, GCS and elective variables.
 - The APD data dictionary should include updated definitions and the APACHE II and APACHE III-J scoring tables for all variables.
 - Units should be encouraged to have a simple, short list of problematic definitions available to data collectors at all times so definitions can be checked as data is collected rather than entered. This could be developed by CORE and distributed to units.
 - Data entry systems (such as AORTIC) should include definitions and additional logic checks at the point of data entry (i.e., admission following elective surgery with GCS of 3).
 - APD training should focus on the collection of these variables - those involved in APD training should be competent to discuss coding/data collection.
- Chronic health and oxygenation produced the largest changes in predicted mortality. Chronic health is problematic due to its "interpretive" nature, leading to consistent over-coding. Oxygenation is problematic because its complex definition makes it hard to quickly determine the worst oxygenation for a patient and the alternative, entering all arterial blood gas results for a patient (sometimes upwards of 15) can lead to data entry fatigue. While the APACHE III-J chronic conditions may not be as problematic as the APACHE II conditions, oxygenation will likely remain a problem. If the APD considers developing an Australian/New Zealand version of APACHE in the future it may be worthwhile considering whether a system could be developed that didn't include these variables in their current state. A simplified version of each variable may be an easy way to improve data quality.
- While not significant, it was clear from the audit that there was variability in the choice of diagnosis between the original and audit datasets. It may be worthwhile analysing problematic diagnoses to see if a consensus can be found on diagnoses that are commonly coded differently at different sites. This could then be incorporated into APD training, along with clarification of the APD position on patients coming from the operating theatre receiving a post-operative diagnosis.

Recommendations for the APD Data Audit Program:

- The audit should move to APACHE III-J variables, this would enable more patients to be included in the predicted mortality analysis, would cover more variables and would establish whether the APACHE III-J chronic conditions are more accurately coded than the APACHE II chronic conditions.
- The audit tool should remain capable of auditing APACHE II variables so that sites already audited can be re-audited to determine whether the audit process is improving data quality at individual sites. A balance between auditing new sites and re-auditing sites to assess improvement will need to be struck.
- Hospital outcome should be collected as part of the audit.
- If possible, participating in the audit program, should be made a compulsory aspect of contributing to the APD.
- The survey given to audited sites should be redesigned to include:
 - The FTE each site assigns to data collection/data entry.
 - The source of data for various variables.
 - The APD should recognise groups of dedicated data collectors as dedicated collectors.
- Note should be taken of the size of the unit being audited, if a unit receives a large number of admissions per year then a shorter audit period should be used to ensure a base level of admissions (e.g., 5%) are audited per site.
- The records to be audited should represent the case mix at each individual unit. This could be done by selecting for either diagnostic code or predicted risk of death when selecting records for the audit.

Appendix 1: List of audit variables

Variable	APACHE II relevance
Age	Used to calculate Age score
Sex	
Hospital admission date	
Hospital admission time	
ICU admission date	
ICU admission time	
ICU discharge date	
ICU discharge time	
Source of ICU admission	Used in calculation of Chronic score and predicted risk of death
Elective surgery admission	Used in calculation of Chronic score and predicted risk of death
Diagnosis	Used in calculation of predicted risk of death
Chronic conditions: Immune suppressed Immune disease Respiratory Cardiovascular Liver Renal	Used to calculate Chronic score (with ICU Source and Elective Surgery)
Glasgow Coma Score - Eye	
Glasgow Coma Score - Verb	
Glasgow Comma Score - Motor	
Total Glasgow Coma Score	Used to calculate GCS score
<i>Physiological variables from first 24 hours in ICU:</i>	
Temperature (high/low)	Used to calculate Temperature score
Heart rate (high/low)	Used to calculate Heart rate score
Respiratory rate (high/low)	Used to calculate Respiratory rate score
Systolic blood pressure (high/low)	Used to calculate Mean Arterial Pressure (MAP)
Diastolic blood pressure (high/low)	which is then used to calculate MAP score
Sodium (high/low)	Used to calculate Sodium score
Potassium (high/low)	Used to calculate Potassium score
Creatinine (high/low)	Used to calculate Creatinine score
Acute renal failure	Used in calculation of creatinine score
Haematocrit (high/low)	Used to calculate Haematocrit score
White cell count (high/low)	Used to calculate White cell count score
FiO ₂ (associated with worst ABG)	Used to calculate Oxygenation score
PaO ₂ (associated with worst ABG)	Used to calculate Oxygenation score
PaCO ₂ (associated with worst ABG)	Used to calculate Oxygenation score
pH (worst)	Used to calculate pH score
Bicarbonate (lowest)	Used to calculate pH score when pH is not available
Ventilated	
Intubated	

Appendix 2: Tolerance levels for individual variables

When determining percentage agreement of individual variables a tolerance level was set for each variable. When the difference between the original value and the audit value fell within the allowed tolerance, the entry was deemed "reliable". Tolerance levels were based on a combination of the unit of measurement for each individual variable and the method of collection. For example, respiratory rate is collected as a whole number often from a hand-written chart with graduations of 2 between each line, therefore the tolerance was set at 1.

Variable	Allowed tolerance
Age	< 1 year
Sex	0
Hospital admission date	0
Hospital admission time	< 1 hour
ICU admission date	0
ICU admission time	< 1 hour
ICU discharge date	0
ICU discharge time	< 1 hour
Source of ICU admission	0
Elective surgery admission	0
Diagnosis	0
Chronic conditions: Immune suppressed Immune disease Respiratory Cardiovascular Liver Renal	0
Acute renal failure	0
Intubated	0
Ventilated	0
Glasgow Coma Score - Eye	0
Glasgow Coma Score - Verb	0
Glasgow Comma Score - Motor	0
Total Glasgow Coma Score	0
Heart rate	9
Respiratory rate	1
Temperature	0.1
Mean arterial pressure	10
Sodium	0
Potassium	0.2
Creatinine	0
Haematocrit	0.01
White cell count	0.9
FiO ₂	0.05
PaO ₂	10
PaCO ₂	5
pH	0.19
Bicarbonate	0.5

Appendix 3: Generation of an APACHE II score

THE APACHE II score is generated by adding together the following:

Age Score + Acute Physiology Score* + Chronic Health Score**

* The Acute Physiology Score: the addition of the 12 physiology variable scores; heart rate score, temperature score, respiratory rate score, sodium score, potassium score, creatinine score, haematocrit score, white cell count score, mean arterial pressure score, pH score, oxygenation score and GCS score.

** Chronic Health Score: if a patient is coded as positive for one or more of the 5 chronic co-morbidities, the patient receives a chronic score of 5. If the patient is also coded as elective, with an ICU source of operating theatre/recovery, then the chronic score drops to 2.

Appendix 4: Generation of an APACHE II risk of death

The APACHE II predicted risk of death is determined as follows:

Factor = -3.517 + (APACHE II score x 0.146) + Diagnosis weighting + 0.603 (only if post emergency surgery)

Predicted risk of death = EXP(Factor)/(1+EXP(Factor))

AP2 Diagnosis	Weight	Non-op/Post-op	Description
101	-2.108	Nonoperative	Asthma/Allergy
102	-0.367	Nonoperative	COPD
103	-0.251	Nonoperative	Pulmonary Oedema (non cardiac)
104	-0.142	Nonoperative	Aspiration/Poisoning/Toxic
105	-0.128	Nonoperative	Pulmonary Embolism
106	0	Nonoperative	Respiratory Infection
107	0.891	Nonoperative	Respiratory neoplasm
108	-0.168	Nonoperative	Post respiratory arrest
109	-1.798	Nonoperative	Hypertension
110	-0.424	Nonoperative	Congestive Cardiac Failure
111	0.493	Nonoperative	CVS failure from Haem shock/hypovolaemia
112	-0.191	Nonoperative	Coronary Artery Disease
113	0.113	Nonoperative	Sepsis (any aetiology)
114	0.393	Nonoperative	Post Cardiac Arrest
115	-0.259	Nonoperative	Cardiogenic shock
116	0.731	Nonoperative	Dissecting Thoracic /abdominal Aneurysm
117	-1.368	Nonoperative	Rhythm disturbance
118	-1.228	Nonoperative	Multiple trauma
119	-0.517	Nonoperative	Head Trauma only
120	-0.584	Nonoperative	Seizure disorder
121	0.723	Nonoperative	ICH/SDH/SAH
122	-3.353	Nonoperative	Self-poisoning/Overdose
123	-1.507	Nonoperative	Diabetic Ketoacidosis
124	0.334	Nonoperative	Gastrointestinal bleeding
201	-1.376	Operative	chronic cardiovascular disease
202	-1.315	Operative	Peripheral Cardiac Disease
203	-1.261	Operative	Heart Valve Surgery
204	0.113	Operative	Sepsis (any aetiology after surgery)
205	-0.682	Operative	Haemorrhagic shock (after surgery)
206	0.393	Operative	Post cardiac arrest (after or during surgery)
207	-1.684	Operative	Multiple trauma
207	-1.684	Operative	Multiple trauma
208	-0.955	Operative	Head Trauma
209	-0.802	Operative	Thoracic surgery for neoplasm
211	-0.14	Operative	Respiratory insufficiency after surgery
212	-0.617	Operative	Gastrointestinal Bleeding
213	-0.248	Operative	GI surgery for neoplasm
214	0.06	Operative	GI Perforation/Obstruction
215	-1.204	Operative	Renal Surgery Neoplasm
216	-1.042	Operative	Renal Transplant surgery
217	-0.788	Operative	Craniotomy for ICH/SDH/SAH
218	-1.245	Operative	Craniotomy for neoplasm
219	-0.699	Operative	Laminectomy /Spinal Surgery

Continued:

AP2 Diagnosis	Weight	Non-op/Post-op	Description
301	-0.759	Nonoperative	Neurological Undefined
1301	-1.15	Operative	Neurological Undefined
302	0.47	Nonoperative	Cardiovascular -Undefined
1302	-0.797	Operative	Cardiovascular -Undefined
303	-0.89	Nonoperative	Respiratory Undefined
1303	-0.61	Operative	Respiratory Undefined
304	0.501	NonOperative	Gastrointestinal Undefined
1304	-0.613	Operative	Gastrointestinal Undefined
305	-0.885	Nonoperative	Renal Undefined
1305	-0.196	Operative	Renal Undefined
306	.	Nonoperative	No APACHE II equivalent
1306	.	Operative	No APACHE II equivalent
307	-0.885	Nonoperative	Metabolic Undefined
1307	-0.196	Operative	Metabolic Undefined
308	.	Nonoperative	No APACHE II equivalent
1308	.	Operative	No APACHE II equivalent