Prediction and Prevention of Complications in General Surgery

John C Woodfield

A thesis submitted for the degree of
Doctor of Philosophy
of the University of Otago, Dunedin
New Zealand

2011
Abstract

This thesis examines two concepts that impact on surgical complications – prediction and prevention. The starting point was a randomised controlled trial (RCT) which compared the frequency and the cost of infection for two prophylactic antibiotics. A parallel study also examined the surgeons’ ability to predict wound infection and major complications. The results generated a number of important clinical questions, and it is the studies that were performed to investigate these questions which determined the direction and development of the thesis.

Although prediction is fundamental to clinical practice, there was minimal objective data on the ability of the surgeon to predict complications. How good is the surgeon at predicting major complications? Can prediction be improved? How important is it when the surgeon changed the prediction of complications at the end of the procedure? This thesis addresses these questions. In contrast the ability of prophylactic antibiotics to prevent infection has been extensively studied and the theory behind their use is well understood. We wanted to assess the ‘best prophylactic antibiotic’ to use in abdominal surgery, both in terms of frequency of infection and cost effectiveness. We were also interested in the ability of the surgeon to predict wound infection and on the impact of prophylactic antibiotics on chest infection and urinary infection.

The RCT comparing the use of ceftriaxone and cefotaxime showed that ceftriaxone had some advantages, including better prophylaxis for patients undergoing appendicectomy, and better prophylaxis against chest and/or urinary infection. This also demonstrated that the cost of infection was significantly different between the two antibiotics, and that this could be used to demonstrate significant differences in the severity of infection. For some endpoints there was a significant difference in the cost of infection even when the frequency of infection was the same. Differences in the frequency of chest infection and urinary infection were the impetus for performing a meta-analysis of all RCT assessing the prophylactic use of ceftriaxone. This showed that ceftriaxone was significantly better at preventing infection than other appropriately selected antibiotics. This difference was noted mainly in high risk cases. For the wound this was most noticeable after clean-contaminated and contaminated surgery. For chest infections this was after upper abdominal surgery and for urinary infections after pelvic surgery.
The study looking at the prediction of wound infection, using a 100mm visual analogue scale (VAS), was performed immediately before and after surgery. This showed that the surgeon was poor at predicting who would, or would not, go on to develop a wound infection. However an increase in the prediction of infection postoperatively identified a high risk group of patients who developed significantly more infections. The finding that microbiological and surgical factors were only weakly predictive lead to a study which aimed to assess the importance of patient factors, as indicated by the American Society of Anesthesiologists (ASA) classification of physical status, on the development of wound infection. This demonstrated that in the context of optimal prophylactic antibiotic use that the impact of wound contamination was minimized and that host factors, as represented by ASA score, become more important in preventing wound infection.

The prediction of major complications by the surgeon was then assessed in a three part prospective study. In part one prediction was made using a 100mm VAS immediately before and after surgery. In part two a preoperative multifactorial VAS, using six additional subscales was introduced. In part three predictions were completed following the presentation of detailed outcome feedback. Surgical prediction was initially ‘as good as’ the Physiology and Operative Severity Score for the enUmeration of Mortality and Morbidity (POSSUM) score. After the presentation of clinically relevant feedback there was an improvement in the prediction of major complications (from very good to excellent). This is the first time that feedback has been shown to improve the prediction of major complication.

The prediction of major complications study also supported the following observations. Increasing the prediction of major complications postoperatively was important, and confirmed the importance of intraoperative events in determining the outcome of surgery. Surgical prediction was able to improve both the discrimination and goodness of fit of a multifactorial model for predicting complications. One reason for this is the ability of the surgeon to identify infrequent, but clinically important, risk factors. The prediction of complications by the surgeon using a VAS was accurate, versatile (when compared to a number of risk models), broadly applicable (able to be used by a number of surgeons and for a wide range of procedures) and adaptable. A major limitation is the need for formal reliability testing.
When a VAS is intended to be used as a linear scale it is helpful to clearly label (or ‘anchor’) the VAS.

Further studies need to look at the reliability and potential applications of surgical prediction of major complications using a VAS score.
Acknowledgements

I would like to thank Rachel, who has provided both support and encouragement through the good and bad times. Without her support this thesis would not have been possible.

I would also like to thank my supervisor Professor Andre van Rij, who has given invaluable support and guidance. It was hid recommendations and vision that made it possible for this thesis to be presented.

This thesis builds on the work of others. The randomised trial involving ceftriaxone and cefotaxime, the study on the prediction of wound infection, and the first study on the prediction of major complications, was designed and run by Mr Ross Pettigrew and by Professor Andre van Rij. My contribution included clinical involvement in the study, collection and analysis of the data and writing the scientific papers. The opportunity provided to be involved in these studies, and the training received through this work is most appreciated. The design and running of the other projects in this thesis (studies three, four and five and the meta-analysis) are my own work.

This project involved extensive statistical analysis. Statistical advice and supervision were provided by Dr Lindsay Plank (Department of Surgery, Auckland). His advice and help was invaluable. This was provided voluntarily, and his friendship (and the opportunity provided to move into his office to perform logistic regression analysis using SAS) is appreciated.

The work on the prediction of complications at Leeds was made possible through the supervision and efforts of Mr Peter Sagar. The supportive environment and the enthusiasm of the surgeons at the Leeds General Infirmary were essential to making this study a success. The generation of POSSUM scores was helped by Mr Dermot Burke who provided a software package designed for this purpose. The running of the project was also helped by research fellows at the LGI, led by Dr Dinesh Thekkinkattil.

The meta-analysis was helped greatly by Dr Nagy Beshay, and by staff at the Royal Australasian College of Surgeons library, who helped us to track down many ‘difficult to find’ articles.
Table of Contents
Abstract .......................................................................................................................... i
Acknowledgements ..................................................................................................... iv
Table of Contents ......................................................................................................... v
Table of Tables ........................................................................................................... xv
Table of Figures .......................................................................................................... xvii
Abbreviations .............................................................................................................. xix
CHAPTER 1 ...................................................................................................................... 1

1. Introduction .............................................................................................................. 1
   1.1 Prediction and prevention ..................................................................................... 1
   1.2 The impact of nosocomial infection and wound infection ................................. 1
   1.3 A brief overview of use of prophylactic antibiotics to prevent wound infection ... 2
   1.4 Use of cephalosporin antibiotics for prophylaxis against wound infection ....... 3
   1.5 Infections at a distance from the site of the surgical incision ......................... 4
   1.6 Cost and cost effectiveness of prophylactic antibiotics ..................................... 6
   1.7 Prediction of wound infection ........................................................................... 7
   1.8 Reasons for studying the use of prophylactic antibiotics to prevent infection after abdominal surgery ................................................................. 7
   1.9 Links between prediction and prevention ........................................................... 8
   1.10 Prediction of risk is fundamental ...................................................................... 8
   1.11 Use of risk factors in assessing the risk of major complications after surgery .... 9
   1.12 Scoring systems ............................................................................................... 10
   1.13 Prediction of complications by the clinician .................................................... 11
   1.14 Use of a visual analogue scale ........................................................................ 12
   1.15 Differences between preoperative and postoperative prediction ................... 13
   1.16 Questions about surgical prediction ................................................................ 13
   1.17 Overview of thesis ............................................................................................ 13

CHAPTER 2 ...................................................................................................................... 17

2. A prospective randomised double blinded controlled trial comparing antibiotic prophylaxis with Ceftriaxone and Cefotaxime in abdominal surgery. A comparison of the frequency of infection ................................................................. 17
   2.1 Rationale for study .............................................................................................. 17
      2.1.1 Choice of two third generation cephalosporins ........................................... 17
      2.1.2 In-vitro activity of study antibiotics .............................................................. 17
      2.1.3 Pharmacokinetic considerations ................................................................. 18
2.1.4 Cost ...........................................................................................................18
2.1.5 Safety profile and single dose prophylaxis ..............................................19
2.1.6 Need for a well-designed randomised controlled trial comparing these two antibiotics ...........................................................................................................18

2.2 Methods of study ........................................................................................19
  2.2.1 Inclusion, randomisation, blinding ............................................................19
  2.2.2 Antibiotic administration ........................................................................20
  2.2.3 Definition of endpoints ............................................................................20
  2.2.4 Assessment of endpoints ........................................................................21
  2.2.5 Microbiology ...........................................................................................21
  2.2.6 Statistical analysis ...................................................................................22
  2.2.7 Ethical approval and consent ....................................................................22
  2.2.8 Pharmaceutical involvement in study ......................................................22

2.3 Results ..........................................................................................................23
  2.3.1 Inclusion and exclusion ..........................................................................23
  2.3.2 Patient demographic and operative data ...............................................23
  2.3.3 Wound infections (superficial and deep surgical site infection) ..............23
  2.3.4 Deep peritoneal infection (space surgical site infection) .........................25
  2.3.5. Distant infection ....................................................................................25
  2.3.6 Combined infection ................................................................................25
  2.3.7 Microbiology ...........................................................................................27

2.4 Discussion .....................................................................................................27
  2.4.1 Wound infection (superficial and deep incisional surgical site infection) .27
    2.4.1.1 Both antibiotics provided effective prophylaxis for wound infection ...27
    2.4.1.2 Ceftriaxone provided better prophylaxis for appendicectomy ........27
    2.4.1.3 Why would ceftriaxone provide better prophylaxis following appendicectomy? 27
    2.4.1.4 Differences in wound microbiology ..................................................27
  2.4.2 Deep peritoneal infection (surgical space infection) ...............................30
  2.4.3 Distant infection ......................................................................................31
  2.4.4 Differences between the two antibiotics when cases with anastomotic morbidity are excluded .................................................................32

2.5 Limitations of this study ..............................................................................32

2.6 Conclusions ..................................................................................................32
  2.6.1 Similarities between the antibiotics ........................................................33
  2.6.2 Differences between the antibiotics ........................................................33
  2.6.3 The importance of a therapeutic concentration of antibiotic during the time of bacterial contamination .............................................................33

2.7 Further questions for study .........................................................................33

2.8 Publication ....................................................................................................34

CHAPTER 3 ........................................................................................................34

3. Using cost of infection as a tool to demonstrate a difference in prophylactic antibiotic efficacy. A prospective randomised double blinded comparison of the
Accuracy of the surgeon’s clinical prediction of wound infection using a visual analogue scale. ................................. 51

4.1 Introduction .................................................................................................. 51

4.1.1 Why predicting wound infection may be useful ................................. 51
4.1.2 Overview of risk factors for wound infection ................................. 51
5. ASA classification of physical status as a predictor of wound infection

Bookmark not defined.

5.1 Introduction

5.1.1 How important is the global health of the patient in the development of a wound infection?

5.1.2 The American Society of Anesthesiologists classification of physical status as a measurement of the global health of the patient

5.1.3 Impact of ASA classification of physical status in the context of optimal prophylaxis
5.3 Results.................................................................................................................................................. 70
  5.3.1 Number of patients .......................................................................................................................... 70
  5.3.2 Association between risk factors and the rate of wound infection ............................................. 70
  5.3.3 Interaction between the ASA score and the classification of the wound ..................................... 70
  5.3.4 Logistic regression analysis .......................................................................................................... 71
5.4 Discussion ........................................................................................................................................... 72
  5.4.1 ASA classification of physical status as a predictor or wound infection in other studies ............ 72
  5.4.2 The impact of optimal prophylactic antibiotic use on the predictive ability of the ASA classification of physical status ................................................................. 72
  5.4.3 Implication of these results ........................................................................................................... 73
    5.4.3.1 Improving the ASA classification of physical status .... Error! Bookmark not defined.
    5.4.3.2 Is there a role for giving additional prophylactic antibiotics to patients with a high ASA score? ................................................................................................................ 73
    5.4.3.3 ASA classification of physical status and improving the host’s defence against infection ........ 74
5.5 The relationship between wound infection and different components of the ASA score ... 74
  5.5.1 Functional status and wound infection ............................................................................................ 74
  5.5.2 Medical comorbidities, nutritional intervention and wound infection ......................................... 75
  5.5.3 Acute physiology and wound infection .......................................................................................... 75
  5.5.4 Merits of a combined approach ...................................................................................................... 77
5.6 Limitations of this study ...................................................................................................................... 78
5.7 Conclusion ......................................................................................................................................... 78
5.8 Further questions for study ................................................................................................................ 78
5.9 Publication ......................................................................................................................................... 79

CHAPTER 6 ............................................................................................................................................... 80
6.  A meta-analysis of randomised controlled trials assessing the prophylactic use of Ceftriaxone. A study of wound, chest and urinary infections. ........................................... 80
6.1 Introduction ......................................................................................................................................... 80
  6.1.1 Should ceftriaxone be considered as a first-line prophylactic antibiotic for the prevention of wound infection? ...................................................................................... 80
  6.1.2 A study of wound, chest and urinary infections. ......................................................................... 80
  6.1.3 Why a meta-analysis now? ............................................................................................................ 81
  6.1.4 Questions to be answered by the meta-analysis .......................................................................... 81
6.2 Methods ............................................................................................................................................. 81
  6.2.1 Search strategy .............................................................................................................................. 81
  6.2.2 Inclusion and exclusion criteria ................................................................................................... 82
  6.2.3 Analysis of papers ........................................................................................................................ 82
  6.2.4 Analysis of the data ...................................................................................................................... 83
6.3 Results .............................................................................................................................................. 84
7. A prospective study of the ability of the surgeon to predict major complications 108

7.1 Introduction ................................................................. 108

7.2 Background information for Study Two: Summary of previous studies looking at clinical prediction of complications ......................................................... 108

7.2.1 Study One: The Dunedin Study ........................................ 108
7.2.2 Other studies looking at clinical prediction ................................. 111

7.3 Rationale for Study Two: The Leeds Study .................................... 112

7.3.1 Outstanding Questions .............................................. 112
7.3.2 Approach to answering these questions .................................. 113
7.3.2.1 Improving prediction ........................................... 113
7.3.2.2 Technical proficiency and a change in prediction .................. 113
7.3.2.3 Surgical prediction and models assessing the risk of complications 114

7.4 Design ........................................................................ 114

7.4.1 Overview .................................................................... 114
7.4.2 Inclusion criteria .......................................................... 114
7.4.3 Definition of complications .............................................. 114
7.4.4 Definition of complications .............................................. 114
7.4.4.1 Documentation of preoperative risk factors .................. 117
7.4.4.2 Documentation of intraoperative risk factors ..................................................118
7.4.5 Documentation of complications .......................................................................119
7.4.6 Documenting the prediction of a major complications ........................................120
  7.4.6.1 VAS form used in Phase I .............................................................................120
  7.4.6.2 Prediction using multiple VAS scales in Phases II and III ..............................120
  7.4.6.3 Prediction using POSSUM ...........................................................................124
7.4.7 Clinical feedback ................................................................................................124
7.4.8 Statistics ..............................................................................................................125
  7.4.8.1 Statistics for assessing sample size ...............................................................125
  7.4.8.2 Statistical analysis of the results ..................................................................126
7.4.9 Ethics committee approval ...............................................................................126

7.5 Results ..................................................................................................................126
  7.5.1 Number of patients and predictions .................................................................126
  7.5.2 Number of predictors ....................................................................................126
  7.5.3 Demographics and pathology .........................................................................127
  7.5.4 Details of surgery ..........................................................................................127
  7.5.5 Complications ................................................................................................129

7.6 Feedback ..............................................................................................................129

7.7 Results: Accuracy of prediction of major complications by the surgeon ............131
  7.7.1 VAS scores with and without complications ..................................................131
  7.7.2 Discrimination and goodness of fit of prediction of major complications ........132
  7.7.3 Calibration of prediction of complications against observed complications ....134
  7.7.4 Mortality ........................................................................................................135
  7.7.5 Statistical issues pertaining to the comparison between surgical prediction and
      POSSUM ..............................................................................................................136
  7.7.6 Subanalysis looking at surgical speciality and the experience of the predictor....137
      7.7.6.1 The impact of different surgical specialities ...........................................126
      7.7.6.2 The impact of different levels of experience ...........................................126

7.8 Discussion: Accuracy of prediction of major complications by the surgeon ..........138
  7.8.1 Phase I: Initial assessment. Initial prediction ‘as good as’ POSSUM ................138
  7.8.2 Phase II: Structured assessment. No improvement in prediction. .................139
  7.8.3 Phase III: Impact of feedback .........................................................................139
  7.8.4 Literature comparing POSSUM and surgical prediction ...............................141

7.9 Implications of the change in prediction before and after surgery .....................141
  7.9.1 Previous work assessing a change in VAS score ..............................................142
  7.9.2 Results .............................................................................................................142
  7.9.3 Discussion .......................................................................................................146
      7.9.3.1 Comparison of results with previous study ...........................................146
      7.9.3.2 Potential applications ...........................................................................147

7.10 Inclusion of VAS in a model for predicting complications ...............................147
  7.10.1 Introduction ....................................................................................................147
  7.10.2 Methods ......................................................................................................148
      7.10.2.1 Model design .......................................................................................148
      7.10.2.2 Statistical analysis .............................................................................148
  7.10.3 Results ..........................................................................................................148
8. Factors to consider in the design and implementation of a visual analogue scale...162

8.1 Introduction...........................................................................................................162

8.2 Variable use of the ‘global’ visual analogue scale..............................................162

8.2.1 Introduction....................................................................................................162

8.2.2 Methods.........................................................................................................162

8.2.3 Results...........................................................................................................164

8.2.3.1 Results of questionnaire.............................................................................164

8.2.3.2 Impact of using adjusted percentage scores on the prediction of complications .164

8.2.4 Discussion.....................................................................................................167

8.2.4.1 Implications of the variable use of the VAS..............................................167

8.2.4.1.1 Impact on the results of Study Two (Leeds study)...............................167

8.2.4.1.2 Implication on the results of Study One (Dunedin study)...................167

8.2.4.2 Different ways of using an unlabelled VAS scale ....................................168

8.2.4.3 Using visual analogue scales to document subjective and objective measurements.................................................................168

8.2.4.4 Use of an anchored VAS in predicting major complications .................169
CHAPTER 9

9. Conclusions

9.1 Part A

9.2 Wound infection

9.2.1 The mechanism of action of ceftriaxone

9.2.2 Ceftriaxone and stratification of the risk of wound infection

9.3 Chest infection and urinary infection

9.3.1 Mechanism of prophylactic antibiotics in preventing chest infection and urinary infection

9.3.2 Ceftriaxone and stratification of risk to prevent urinary and chest infection

9.4 Cost effectiveness of prophylactic antibiotics

9.5 Lessons learned from the surgeon’s prediction of wound infection

9.5.1 Protocols may be better than prediction

9.5.2 The importance of the surgeon changing the prediction of wound infection during the operation

9.6 The importance of the ASA score

9.7 Recommendations arising from Part A

9.7.1 Recommendations for the use of ceftriaxone as a prophylactic antibiotic

9.7.2 Pathophysiology of postoperative infections

9.7.3 Clinical aspects of postoperative infections

9.8 Limitations of this study

9.9 Part B

9.10 The context of performing a study assessing the prediction of major complications

9.11 Accuracy of surgical prediction using a VAS

9.11.1 How accurate is surgical prediction?
9.11.2 Can surgical prediction be improved? .......................................................... 183
9.12 What lessons can be learned about the optimal use of the VAS for scoring the prediction of major complications? .......................................................... 184
9.13 Clinical feedback ............................................................................................. 184
9.14 What is the importance of the surgeon changing the risk of complications postoperatively? .......................................................... 184
9.15 Can surgical prediction be used to improve a model assessing the risk of complications? .......................................................... 185
9.16 Can surgical prediction be used to impact on patient management? .............. 185
9.17 Conclusions to Part B .................................................................................... 185
9.18 Closing summary ............................................................................................ 186

References ............................................................................................................ 188

APPENDIX A: ........................................................................................................ 219
APPENDIX B: ........................................................................................................ 222
APPENDIX C: ........................................................................................................ 2300
APPENDIX D: ........................................................................................................ 233
Table of Tables

Table 2.1 Exclusion from study ................................................................. 23
Table 2.2 Patient demographic and operative data .................................... 24
Table 2.3 Wound infection ..................................................................... 25
Table 2.4 Peritoneal and distant infections ............................................. 26
Table 2.5 Combined infections ............................................................... 26
Table 2.6 Wound microbiology ............................................................... 28
Table 3.1 Cost of superficial and deep incisional surgical site infections .... 41
Table 3.2 Cost of chest and urinary infection .......................................... 43
Table 3.3 Cost of all infection ................................................................. 43
Table 4.1 VAS scores for patients with and without a wound infection .... 54
Table 4.2 Reasons for changing the risk of wound infection postoperatively 56
Table 4.3 Risk factors for predicting a wound infection ............................. 58
Table 4.4 Logistic regression analysis ..................................................... 60
Table 5.1 Associations between risk factors for wound infection and the incidence of wound infection ......................................................... 69
Table 5.2 The interaction between ASA classification of physical status and the wound classification system on the incidence of wound infection .......................................................... 71
Table 5.3 Logistic regression analysis ..................................................... 71
Table 6.1 Summary of meta-analysis of studies assessing wound infection 89
Table 6.2 Summary of meta-analysis of studies assessing urinary tract infection 94
Table 6.3 Summary of meta-analysis of studies assessing pneumonia .......... 99
Table 7.1 Surgeons’ VAS scores for patients with and without complications 109
Table 7.2 Reasons given by the surgeon for changing the VAS score for major complications ................................................................. 110
Table 7.3 Summary of logistic regression analysis for the prediction of operative risk using the Oto model ................................................................. 111
Table 7.4 Demographics and pathology ................................................. 127
Table 7.5 Type of surgery ..................................................................... 128
Table 7.6 Number of patients with a complication .................................... 129
Table 7.7 Breakdown of major complications .......................................... 130
Table 7.8 Differences in VAS scores in patients with and without complications 131
Table 7.9 Accuracy of prediction of major complications using the surgeon’s VAS prediction and the POSSUM morbidity score ................................................................. 132
Table 7.10 Discrimination of prediction for surgeons of different levels of experience 138
Table 7.11 Reasons given for changing the VAS score for major complications 139
Table 7.12 Relationship between categorical classification and the written reason for a change in risk assessment ................................................................. 145
Table 7.13 Summary of univariate odds ratios and adjusted odds ratios for major complications using logistic regression analysis ................................................................. 149
Table 7.14 Model results for different options .......................................... 151
Table 7.15 Likelihood ratios for surgical prediction of major complications 159
Table 8.1 Percentage prediction of major complications for a surgeon redefining the VAS scale according to the ‘complexity of surgery’ ................................................................. 166
Table 8.2 Comparison of the discrimination of prediction before and after adjustment to a percentage scale .................................................................166
Table 8.3 Standardised beta coefficients for predicting major complications ..................173

Table D.1 Summary of operative procedures .......................... 236
Table D.2 Complications after surgery in patients contacted by telephone (388 patients) 237
Table D.3 Severity of complications diagnosed after discharge from hospital ............... 239
Table of Figures

Figure 1.1 Influence of risk prediction on clinical practice ......................................................... 9
Figure 1.2 Overview of Thesis ................................................................................................. 15
Figure 2.1 Serum concentration of Ceftriaxone and Cefotaxime over 24 hours following intravenous administration of 1g of antibiotic ................................................................. 18
Figure 2.2 Serum concentration of ceftriaxone compared to the MIC 90 of most pathogenic bacteria ................................................................................................................................. 30
Figure 3.1 Breakdown of the total cost of infection for both antibiotics ........................................ 39
Figure 3.2 Comparison of the mean cost of different infections following the use of prophylactic antibiotics in abdominal surgery ............................................................................. 42
Figure 4.1 Flow diagram demonstrating the incidence of wound infection in low and high risk preoperative categories and the influence of changing the prediction of wound infection postoperatively ................................................................. 55
Figure 4.2 ROC curves comparing wound categories and the surgeon’s preoperative and postoperative prediction of complications ......................................................................................... 59
Figure 6.1 Flow diagram for identification and inclusion of papers ............................................. 83
Figure 6.2 Funnel plot for all wound infection before exclusion of three studies on or outside the 95% confidence interval line .......................................................................................... 86
Figure 6.3 Funnel plot after removing three studies on or outside the 95% percentile line .... 86
Figure 6.4 Forest plot with odds ratio and 95% confidence intervals for all studies assessing wound infection .......................................................................................................................... 87
Figure 6.5 Forest plot for studies with double blinding and CDC definition of infection assessing incisional SSI ................................................................. 90
Figure 6.6 Funnel plot for 35 studies assessing UTI ................................................................. 92
Figure 6.7 Forest plot with odds ratio and 95% confidence intervals for all studies assessing urinary tract infection (excluding endoscopic urology) ................................................................. 93
Figure 6.8 Forest plot for studies with a CDC definition of symptomatic urinary tract infection .95
Figure 6.9 Funnel Plot for 37 studies analysed for pneumonia .................................................. 96
Figure 6.10 Forest plot with odds ratio and 95% confidence intervals for all studies assessing pneumonia .......................................................................................................................... 97
Figure 6.11 Forest plot for studies with a Jadad score of 4 or 5 assessing pneumonia .......... 98
Figure 7.1 Form used to describe the definition of a major complication ............................ 116
Figure 7.2 Form for preadmission and preoperative risk factors .............................................. 117
Figure 7.3 Form for intraoperative risk factors .................................................................... 118
Figure 7.4 Form for documentation of complications ............................................................ 119
Figure 7.5 (a) Form for surgeon’s assessment of major complications in Phase I (b) Instructions about completion of the global VAS ................................................................................. 121
Figure 7.6 Form for the surgeon’s assessment of major complications in Phases II & III .... 123
Figure 7.7 Number of predictions by different surgeons ....................................................... 127
Figure 7.8 ROC curves for Phase I of the study ................................................................. 133
Figure 7.9 ROC curves for Phase III of the study ................................................................. 133
Figure 7.10 Calibration graph of surgeon’s postoperative VAS prediction of major complications ................................................................................................................................. 134
Figure 7.11 Calibration graph of POSSUM morbidity score against major complications . 135
Figure 7.12 Calibration graph of P-POSSUM mortality score against mortality....136
Figure 7.13 Flow chart for major complications, illustrating changes in the VAS score....143
Figure 8.1 Questionnaire (page 161)...........................................Error! Bookmark not defined.
Figure 8.2 Example of logarithmic use of the VAS........................................165

Figure C.1 The Lens Model.......................................................................................230

Figure D.1 Flow diagram for inclusion and exclusion of patients...............................234
Abbreviations

ACS: American College of Surgeons
AIDS: Acquired Immunodeficiency Syndrome
APACHE: Acute Physiology and Chronic Health Evaluation
ASA: American Society of Anesthesiologists
ASB: Asymptomatic bacteriuria
AUC: Area under the curve
CDC: Centre for Disease Control
GDP: Gross Domestic Product
LGI: Leeds General Infirmary
MBC: Minimal Bactericidal Concentration
MIC: Minimal Inhibitory Concentration
mmol/l: millimoles per litre
MRSA: Methicillin-Resistant Staphylococcus aureus
NNIS: National Nosocomial Infection Surveillance
NNT: Number Needed to Treat
NSQIP: National Surgery Quality Improvement Program
O:E: Ratio of observed to expected
O&G: Obstetrics and Gynaecology
POSSUM: Physiology and Operative Severity Score for the enUmeration of Mortality and Morbidity
RCT: Randomised controlled trial
ROC curve: Receiver Operating Characteristic curve
SENIC: Study on the Efficacy of Nosocomial Infection Control
RUS: Resource Utilisation System
SAS: Statistical Analysis Software
SPSS: Statistical Package for the Social Sciences
SUTI: Symptomatic Urinary Tract Infection
TPI: Tissue Penetration Index
UTI: Urinary Tract Infection
VAS: Visual Analogue Scale
Part A

Prevention of Complications
CHAPTER 1
1. Introduction

1.1 Prediction and prevention
This thesis examines two concepts that impact on surgical complications – prediction and prevention.
Prediction: The ability of the surgeon to predict major complications and wound infection is assessed, problems with prediction are identified and the possibility of improving prediction is examined.
Prevention: This section concentrates on optimising the effectiveness of prophylactic antibiotics. Our starting point was to assess optimal prophylactic antibiotic use in abdominal surgery, both in terms of the frequency of infection and cost effectiveness. This study raised a number of questions about risk factors for wound infection and about the potential role of prophylactic antibiotics to prevent chest infection and urinary infection.

The order of the thesis is determined by how events unfolded. This started with a randomised controlled trial on prophylactic antibiotic use. A number of important questions arose from this study. It is the studies which were designed to answer these questions which become the unifying theme which links the different components of this thesis together.

Part A: Prevention of infection
1.2 The impact of nosocomial infection and wound infection
The prevention of nosocomial infection continues to be a major challenge for health workers. The extent, and increasing magnitude, of the problem is illustrated in the United States of America (USA). In 1995 the estimated cost of nosocomial infection was $4.5 billion, with this contributing to 88,000 deaths (Weinstein 1998). By 2002 nosocomial infections contributed to 98,000 deaths and in 2009 the cost was $28.4 billion (Payson 2009).

Nosocomial infection is particularly important following surgery, as the breach of normal barriers to infection, and the catabolic metabolism induced by surgery, predispose the patient to infection. The most frequent nosocomial infections occurring following surgery are wound infection, urinary infection, pneumonia, space surgical site infections and blood stream infections (bacteraemia and septicaemia).
Wound infection [throughout the thesis wound infection is the same as the Centres for Disease Control (CDC) definition of superficial and deep incisional surgical site infection (Horan et al. 1992)] accounts for approximately a quarter of nosocomial infection (Nichols 2001) and can cause significant morbidity, increasing both the length of hospital stay and the overall cost of the operative procedure (Wenzel 1992; Kirkland et al. 1999; Woodfield et al. 2005).

1.3 A brief overview of use of prophylactic antibiotics to prevent wound infection

The classification of surgical wounds into four categories of contamination (Cruse and Foord 1980) has been widely used to document the rates of wound infection (Culver et al. 1991; Hayley et al. 1985; Garibaldi et al. 1991). The initial reported rates of infection, with optimal hospital care, but predating the introduction of prophylactic antibiotics, were 1.5% for clean, 7.7% for clean-contaminated, 15.2% for contaminated and 40.0% for dirty wounds (Cruse and Foord 1980). This finding was a major catalyst to the introduction of prophylactic antibiotics, with a subsequent reduction in wound infection rates to less than 10% for clean-contaminated and contaminated surgery (Wittmann and Condon 1991; Sandusky 1980; Jones and Lang 1993; Castillo and Stempel 1992).

When prophylactic antibiotics were initially introduced there were wide variations in clinical practice. This included differences in who should be given prophylactic antibiotics, the choice of antibiotic used, the timing of antibiotic administration and the duration of antibiotic use. It has been the adoption of evidence based medicine, through laboratory based studies, pharmacokinetic studies and randomised controlled trials, which has led to a more rational approach to prophylactic antibiotic use. Another problem associated with prophylactic antibiotic use has been noncompliance with optimal proven prophylactic antibiotic regimens in clinical practice. A number of studies have now demonstrated that greater compliance with evidence based regimens is associated with a decrease in infection (Dellinger, Hausmann et al. 2005).

In terms of the science and microbiology of wound infection an important initial step was the demonstration that antibiotics must be present in therapeutic concentrations in the wound at the time of contamination to assist the host’s defence against infection (Burke 1961; Polk et al. 1980). This means that antibiotics need to be given shortly before surgery, or at least intraoperatively (Bates et al. 1989), and should be at a therapeutic concentration when the
surgical incision is closed (Burke 1961; Polk et al. 1980; Classen et al. 1992). Antibiotics started after the wound is closed do not prevent postoperative wound infection (Burke 1961; Classen et al. 1992). This finding has been confirmed by studies measuring the concentration of antibiotics in the plasma, site of surgical incision, blister fluid and in other sites of the body. These have also demonstrated that an effective prophylactic antibiotic should be present at the site of potential contamination at a concentration equal to, or greater than, the MIC 90 for the majority of expected bacterial pathogens. The MIC 90 is the concentration of antibiotic required to inhibit 90% of the isolates of bacteria. Differences in the in-vitro MIC 90 of antibiotics can usually be used to explain differences in antibiotic performance. However antibiotics with a similar in-vitro MIC 90 may have a different in-vivo MIC 90 because of differences in pharmacokinetics, such as differences in protein binding and antibiotic distribution.

The importance of the half-life, which impacts on the duration of activity of an antibiotic, is also well recognized. For wound infection, as long as the antibiotic is present at a sufficient concentration when the wound in being closed, it will provide effective prophylaxis. The weight of evidence now confirms that a routine single dose of prophylactic antibiotics, usually at the time of induction of anaesthesia, provides adequate cover in cardiothoracic and abdominal surgery (Song and Glenny 1998; Rowe-Jones et al. 1990; McDonald et al. 1998). Exceptions to this would include:

a) When the agent used has a short half-life and the surgical procedure takes longer than two hours (Wittmann and Condon 1991);
b) When there has been significant loss of blood after administration of the antibiotic;
c) Vascular surgery for lower limb ischemia (Hall et al. 1988a).

Factors important in appropriately selecting prophylactic antibiotics therefore include the timing of antibiotic administration, the microbiological cover of the antibiotic, its penetration into the surgical incision, its duration of action and the safety profile of the antibiotic.

1.4 Use of cephalosporin antibiotics for prophylaxis against wound infection

Many cephalosporin antibiotics have been used for prophylaxis because of their activity against the majority of likely wound pathogens. A number of studies have now shown either equivalence or favourable protection when compared to other antibiotics (Wittmann and
Condon 1991; Kreter and Woods 1992; Song and Glenny 1998) and an excellent safety profile.

First generation cephalosporins have good cover against gram positive organisms and have been widely used in clean procedures where the most likely source of infection is from the skin. Antibiotic use is important in clean surgery when the consequences of an infection may be serious, such as in orthopaedic surgery when a prosthesis is inserted and in cardiothoracic surgery. In intra-abdominal surgery where the incision passes through a potentially contaminated structure (stomach, bladder, biliary tract) second generation cephalosporins have been widely used. Most of these agents have retained reasonable activity against gram positive aerobes and also have additional activity against gram negative organisms, therefore providing effective wound protection against both intra-abdominal organisms and skin organisms. In contaminated procedures, for example when the colon is entered, additional anaerobic cover, usually with metronidazole, is also given.

Third generation cephalosporins have also been extensively used. These antibiotics have an extended range of activity against gram negative organisms, and a variable activity against gram positive organisms (Chambers 1992). A number of authorities have stated that there is no advantage in using a third generation cephalosporin (Song and Glenny 1998; Kreter and Woods 1992) for prophylaxis. However this observation needs to be qualified by considering the type of wound being studied, and the third generation cephalosporin being assessed. Although there appears to be no additional advantage using third generation cephalosporins for clean wounds, this may not be the case in clean-contaminated or contaminated wounds. A number of studies have shown that third generation cephalosporins, especially ceftriaxone or cefotaxime, provide very effective prophylaxis. For example, randomised controlled studies comparing ceftriaxone with a first or second generation cephalosporin showed either a similar performance (Hjortrup et al. 1991; Lumley et al. 1992) or an improved performance in favour of ceftriaxone (de la Hunt et al. 1985; Matikainen and Hiltunen 1993; Morris 1994; Rotmann et al. 1991) in upper gastrointestinal, biliary and colorectal surgery.

1.5 Infections at a distance from the site of the surgical incision
Although there is overwhelming evidence for the use of prophylactic antibiotics to decrease wound infection (Wittmann and Condon 1991), the literature is less clear on whether prophylactic antibiotics can be used to decrease infections at other sites.
Many randomised studies assessing wound infection have also measured the frequency of other infections as ‘secondary’ endpoints. The most common of these have been chest infection (or pneumonia) and urinary tract infection. There is significant variation in these studies. Differences in the definition of secondary endpoints, duration of follow-up, the effort made to identify ‘secondary’ infections and the rate of diagnosed infections makes a comparison of these studies difficult. Although some studies have shown significant differences between prophylactic antibiotics in distant infections, including chest (Lumley et al. 1992; Morris 1994; de la Hunt et al. 1985) and urinary infections (Lumley et al. 1992; Morris 1993), this has not been a consistent observation.

The characteristics required for an antibiotic to prevent ‘other infections’ may be different from those needed to prevent a wound infection. This will be partly determined by when the contamination that leads to infection occurs. In the respiratory and urinary systems, as well as the contamination that occurs during surgery, there may also be additional contamination after surgery. For example colonization of the lung and bronchus may be contributed to by a decreased clearance of secretions from the airway after extubation, a decreased cough reflex or decreased inspiration because of pain leading to atelectasis. Colonization of the urinary tract may be contributed to by incomplete emptying of the bladder, urinary retention or to catheter trauma after surgery. Antibiotics with a MIC90 against relevant bacterial pathogens, which is sustained for up to 24 hours, may therefore have an advantage in decreasing these infections. The reduction of chest and urinary infection has been observed in some studies when ceftriaxone, a third generation cephalosporin with a long half-life, has been used (Morris 1993; Morris 1994; de la Hunt et al. 1985; Lumley et al. 1992). Some studies using one versus multiple doses of the same antibiotic have also shown a decrease in urinary infection (McDonald et al. 1988). However multiple doses of prophylactic antibiotics do increase side effects such as postoperative pyrexia (Turano 1992) and Clostridium difficile colonization (Starks et al. 2008), and may also significantly increase the risk of antibiotic resistance. The relationship between the duration of action of a single dose of a prophylactic antibiotic and the likelihood of a subsequent chest or urinary infection is commented on later in the thesis.
1.6 Cost and cost effectiveness of prophylactic antibiotics

Another debated aspect of prophylactic antibiotic use has been the cost effectiveness of different antibiotic regimens. In the economic environment in which we work cost cannot be ignored. The increase in healthcare expenditure over the last two decades (Angell 1985; Bochner et al. 1996; Eddy 1994; Visual Economics 2010) has emphasized the limited availability of resources, and the need for optimizing health benefits within an available budget. Figures from 2010 show that in the USA healthcare expenditure accounts for 15% of the GDP, with an annual cost of 5,771 USD per person (Visual Economics 2010).

The cost effectiveness of appropriate prophylactic antibiotic use is undisputed, as the cost of the administration of a single dose of a prophylactic antibiotic is minimal compared to the cost of an infection. However the comparative cost effectiveness of different antibiotic regimens is less clear. Comparison of the cost of different prophylactic and therapeutic antibiotics has usually concentrated on the direct cost of the antibiotic and the cost of consumables and/or staff time taken in administering the antibiotic (Smyth et al. 1995). The cost savings recorded in many studies may also be determined by the financial perspective of the researcher performing the investigation. For example, pharmaceutical studies looking at the therapeutic drug substitution of third generation cephalosporins may look primarily at the cost savings to the pharmacy (Pearce and Begg 1994). However, if a cheaper antibiotic is not as effective in treating or preventing an infection, then this may contribute to a longer hospital stay. A limited perspective may therefore encourage a ‘false economy.’ A further problem in studies looking at the cost of prophylactic antibiotics has been to assume that the cost of a wound infection is the same for the different antibiotics being studied (Anderson et al. 1996; Davey et al. 1992). This assumption has never been tested. To measure the true cost effectiveness of different prophylactic antibiotic regimens it is important that all costs, especially the cost of the actual infections that are not prevented, are accounted for.

This thesis investigates a number of cost issues as they pertain to prophylactic antibiotics. The first was to assess the total cost of postoperative infection, with an assessment of inpatient (hospital), outpatient and community costs. The second was to assess if cost could be used as a direct measurement of the severity of infection. It is well recognized that some infections are minor or low cost whereas others are very disabling and expensive. Attempts have been made to compare differences in the severity of infection between different antibiotics, by using wound scoring systems (Wilson et al. 1986; Hall and Hall 2004). In this thesis we
wanted to test the hypothesis that the cost of infections not prevented may be used as a tool to demonstrate a difference in antibiotic efficacy.

1.7 Prediction of wound infection

An issue closely related to the prevention of nosocomial infection is the ability of the clinician to predict which patient is likely to go on and develop a postoperative infection. The accurate prediction of wound infection, either before or immediately after surgery, would help concentrate efforts on preventing infection in high risk patients. Alternatively, if the identification of the high risk patient is poor, then attempts to reduce wound infection would require better adherence to an institutional based approach, or to protocols which are based on well-designed studies that are applicable to most patients.

Although individual risk factors for wound infection have been identified (Hayley et al. 1985; Culver et al. 1991; Garibaldi et al. 1991), we were unaware of any work that assessed how good the clinician was at predicting the development of a wound infection. There was also no work that assessed the importance of the clinician changing his/her prediction of a wound infection following surgery. In this thesis the ability of the surgeon to predict the likelihood of a wound infection is assessed immediately before and after surgery. This study also assessed the importance of the doctor changing the prediction of infection as this may identify a group of high risk patients who require additional treatment. Questions arising from the wound infection study lead to a further analysis of the risk factors for wound infection.

1.8 Reasons for studying the use of prophylactic antibiotics to prevent infection after abdominal surgery

Although the use of prophylactic antibiotics is well established, there remain a number of unresolved questions. The questions this thesis has attempted to address are:

1) What is the preferred antibiotic to use in abdominal surgery, where a high proportion of the wounds will be either clean-contaminated or contaminated?
2) Is there any evidence for prophylactic antibiotics decreasing infections (especially pneumonia and urinary tract infection) distant to the wound?
3) Does the cost of infections not prevented vary for different prophylactic antibiotics?
4) How good is the clinician at predicting the likelihood of a wound infection in the individual patient?
5) How important is the global health of the patient, as measured by the ASA classification of physical status, in the development of wound infection?
6) How important is it when the clinician changes his/her estimation of the risk of a wound infection during the operation?

**Part B: Prediction of complications**

**1.9 Links between prediction and prevention**

Accurate knowledge about postoperative infective complications is central to our ability to predict, prevent and to treat postoperative infections. The linkage between prediction and prevention is also important as there is no utility in making predictions unless they result in useful actions which can prevent subsequent infections.

While in Part A the thesis investigates how accurate the surgeon is at predicting wound infection, in Part B the importance of prediction is extended to a much wider canvas and assesses the ability of the surgeon to predict major complications.

**1.10 Prediction of risk is fundamental**

Assessment of risk underpins decision making, and is so fundamental to everyday life that we are doing this, often at an intuitive or subconscious level, all of the time.

In clinical practice risk assessment has significant implications for both the patient and surgeon. Some of these implications are outlined in Figure 1.1. For example preoperative assessment of medical comorbidities, and the threshold for deciding about preoperative investigations (Eagle et al. 2002) is based on a risk analysis. Our assessment of risk (Figure 1.1) will also influence decisions about patient selection for surgery, including deciding about a lesser operation, or a non-operative intervention in very high risk patients. Informed consent (Figure 1.1) also implies that the patient understands not only the likely outcome of surgery, but also about the nature and likelihood of important complications. This will also influence decisions about operative and postoperative management (Figure 1.1). Examples of assessment of risk directly influencing clinical management include the use of prophylactic antibiotics (section 1.3) and the identification of selected high risk patients for admission to an intensive care unit (Curran and Grounds 1998).
Figure 1.1 Influence of risk prediction on clinical practice

Although risk assessment is fundamental to patient assessment and clinical decision making, and although we all intuitively make judgement calls based on our risk assessment, the problem we are faced with when trying to study this is that there is surprisingly little information available on the process the clinician goes through to predict risk, or on how good he/she is at predicting complications.

1.11 Use of risk factors in assessing the risk of major complications after surgery

Many risk factors contribute to postoperative complications. Some of these include:

- Age
- Medical comorbidity
- Malnutrition
- The extent of the underlying pathology
- The timing (acuteness) of surgery
- The magnitude and complexity of surgery
- Anaesthetic issues
- The standard of postoperative care
- Institutional and surgeon volumes
- System problems that may vary from institution to institution
- Random events.
It is therefore not surprising that efforts to improve clinical outcomes, which concentrated on one area, only identified a subgroup of high risk patients. In the 1970s and 1980s the importance of low albumin and malnutrition in causing complications was emphasized. However the demonstration that low albumin was usually caused by sepsis, and that a careful clinical assessment, noting cardiorespiratory disease, pre-existing sepsis and nutritional status, was as effective as any other single preoperative indicator in predicting major complications after surgery (Pettigrew and Hill 1986) emphasized the importance of the ‘bigger picture.’ In terms of risk assessment an approach which incorporated all risk factors was needed. The development of scoring systems, which combine multiple predictors of risk, has attempted to achieve this.

1.12 Scoring systems

Unquestionably scoring systems represent a significant advance in terms of assessment of risk. However the large number of risk models in use also indicates that there is ‘no easy solution’ to finding a simple system that is flexible enough to be used in most situations. Although a number of models as they impact on general surgery were reviewed, in this thesis our discussion is limited to only a few of these models. These will include POSSUM (Physiological and Operative Severity Score for the enUmeration of Mortality and Morbidity) (Copeland et al. 1991; Copeland 2002), APACHE (Acute Physiology And Chronic Health Evaluation) (Knaus et al. 1985; Knaus 2002) and the Veterans Affairs Surgical Risk Model (Daley et al. 1997; Khuri et al. 1997), which later developed into the American College of Surgeons National Surgical Quality Improvement Program (NSQIP).

Risk models aim to make an accurate prediction, based on the risk factors present, of the expected rate of mortality or morbidity. A robust scoring system is one that has been developed from a large database, which has good discrimination and calibration, and which has been well validated in a number of different settings. The expected (risk-adjusted) rate of complications predicted by the model can then be compared to the observed rate of complications. The model therefore acts as a benchmark for comparative audit. By giving us results that are ‘risk adjusted,’ models enable us to identify true performance, and so can provide the opportunity to disseminate best practice. This can be applied to individual surgeons, to hospitals, and can even be used to compare the outcomes in different countries. This can also be used to correct potentially misleading data. For example, league tables of raw
mortality rates in the UK have been shown to incorrectly identify some institutions as ‘underperforming’ (Adab et al. 2002). Models can also be used to monitor progress over time, and have been able to demonstrate improvements in practice, such as demonstrating a risk adjusted reduction in mortality rates for abdominal aortic aneurysm surgery after the introduction of endovascular stents (Bush et al. 2009).

Unfortunately models have not been so helpful when it comes to the individual patient. Although it should be possible to have a good estimation of risk preoperatively, most risk models have limitations when it comes to assessing preoperative risk. Reasons for risk models not helping with decision making in the individual patient include: the scoring system is designed for comparative surgical audit rather than for helping with the prospective management of an individual patient (Copeland et al. 1991; Khuri et al. 1997; Pillai et al. 1999; Sutton et al. 2002), the use of intraoperative factors and postoperative pathology results in generating the score (Copeland et al. 1991; Tekkis et al. 2002), the model is specifically designed for a single disease entity (Tekkis et al. 2002) or only for intensive care unit use (Knaus et al. 1985), the model assesses mortality but not morbidity (Knaus et al. 1985; Sutton et al. 2002; Tekkis et al. 2002) and some scoring systems having a reduced accuracy beyond the environment in which they were generated (DR Jones et al. 1992; Kempthorne et al. 2002). These limitations were one reason for the recent popularity of the Surgical Apgar Score (Regenbogen et al 2009). This is focused on the patient and is designed to ‘flag’ the high risk patient as needing extra attention before they leave the operating theatre. Although there are now many good risk models that provide us with an accurate prediction of risk (mainly mortality), there is still need for a model which can help us with patient management. In this context we were interested in assessing how accurate the surgeon was in predicting complications in the individual patient.

1.13 Prediction of complications by the clinician

There are a number of reasons why prediction of major complications by the surgeon may be helpful. Whereas models look at groups of patients, the clinician is assessing the individual. The clinician will therefore have unique insights into the case that a model does not have. This includes how the patient ‘looks,’ personal knowledge about the operation to be performed and the package of perioperative care that is available. A second factor is that the clinician can recognize risk factors that are important, but uncommon, and incorporate this into the estimation of risk. Potential advantages of the surgeon’s risk assessment therefore
include that it is tailored to the individual surgeon, able to synthesize multiple risk factors and has the potential to incorporate factors which would otherwise be difficult to quantify.

1.14 Use of a visual analogue scale

The surgeon’s prediction of risk is an ‘educated’ assessment based on clinical experience and a precise knowledge of the clinical scenario. Although this is ‘subjective,’ it can be quantified by the use of a VAS, which transforms the global assessment of risk into a quantitatively useful number.

A visual analogue scale is a single line with a label at both ends. Traditionally this does not have any divisions within it. This has been shown to be reliable and sensitive for rating subjective experiences and is used to measure endpoints such as quality of life, health status and the severity of symptoms (Aitken 1969). Because of the subjective nature of these measurements they are easier to measure visually or spatially rather than to be given a number. The VAS is also an excellent tool for measuring changes over time when used by the same individual (Bernhard et al. 2001). The visual analogue scale can be further defined, by using anchors. The anchors define the line, and can improve the consistency with which the visual scale is used by different people.

When this thesis started there were four studies that assessed the risk of surgical complications using a VAS. Two of these four studies assessed preoperative risk (Pettigrew and Hill 1986; Arvidsson et al. 1996), one postoperative risk (Hartley and Sagar 1994) and one both preoperative and postoperative risk (Pettigrew et al. 1987). On three occasions predictions were made in theatre by surgeons, and on one occasion in the preadmission clinic by anaesthetists (Arvidsson et al. 1996). Although there was some variation in the results of these studies, when taken together they confirmed that using a VAS was a useful way to document or score the clinician’s prediction of major complications. Although two small studies using an unanchored VAS which assessed the global ‘gut feeling’ assessment of the surgeon immediately prior to surgery, did not significantly predict complications (Pettigrew and Hill 1986; Pettigrew et al. 1987), a third study (Arvidsson et al. 1996), of over 1000 patients, using a clearly labelled preoperative VAS showed that clinical assessment was more successful at discriminating adverse events than the patients age, a procedural magnitude score and the ASA grade. The two studies which assessed postoperative prediction also
showed a very good prediction of major complications by the surgeon (Pettigrew et al. 1987, Hartley and Sagar 1994).

1.15 Differences between preoperative and postoperative prediction
One important finding in these studies was the difference between preoperative and postoperative prediction (Pettigrew et al. 1987). In Hartley and Sagar’s study (1994) postoperative prediction was able to identify high risk patients as well as POSSUM. Pettigrew and Burns (1987) demonstrated that a comparison of the VAS score immediately before and after surgery could identify important technical intraoperative events that contribute to perioperative complications (Pettigrew et al. 1987). The importance of technical expertise in preventing adverse events is well recognised (Russell 1987; Bancewicz 1990; Ouriel et al. 1990) and is a fundamental principle underpinning surgical training programs (Callahan et al. 2003), and studies examining the relationship between surgical volume and outcomes (Halm et al. 2002; Ingemar 2003). Changes in the VAS score may be one mechanism for important intraoperative events related to surgical technique to be identified.

1.16 Questions about surgical prediction
There were a number of questions about risk prediction that are addressed in this thesis. These included the following:
1) How good is the surgeon at predicting major complications?
2) How important is it when the surgeon changes the prediction of risk postoperatively?
3) Can prediction of complications by the surgeon be improved?
4) Can surgical prediction using a VAS be used as a tool to help influence decision making?
5) Can surgical prediction using a VAS be used to improve models of risk prediction?
Part B of this thesis either directly addresses these questions, or looks at collecting data that will start the process of helping to answer these questions.

1.17 Overview of thesis
This thesis is centred on four studies and a meta-analysis (Figure 1.2). The thesis has been structured around a problem solving approach to questions that arose from the first study. Part A is based around a prospective, randomised, double blinded comparison of the use of prophylactic ceftriaxone and cefotaxime in acute and elective abdominal surgery (Chapters 2 & 3). One of these, ceftriaxone, maintains a therapeutic tissue concentration for twenty-four hours. This enabled us to examine the relationship of the pharmacokinetics of a prophylactic
antibiotic to the prevention of infection beyond the wound. The study was also designed to prospectively assess the cost of infection, enabling us to investigate the question, “Can cost be used as a tool to compare prophylactic antibiotics.”

A parallel study assessing the ability of the surgeon to prospectively predict the development of wound infection using a VAS was performed (Chapter 4). This lead to a retrospective analysis of the importance of the ASA score in predicting wound infection (Chapter 5). A meta-analysis of the effectiveness of ceftriaxone as a prophylactic antibiotic is then presented (Chapter 6).

Part B presents a prospective study on the accuracy of clinical prediction. Although two studies were performed, the first at Dunedin Public Hospital and the second at the Leeds General Infirmary, because of the length of the thesis only the second study is formally presented. This was designed to compare surgical prediction to POSSUM, to assess if surgical prediction can be improved (two options are investigated) and to collect data to see if surgical prediction using a VAS score can be used to strengthen a multifactorial model for assessing risk. A further study, which partly overlapped with the prospective risk study, compared the rate of complications at discharge from hospital with all complications diagnosed at one month. This presents complication data which enables conclusions to be made about the importance of adequate patient follow up in clinical research. This is included as an appendix.
OVERVIEW OF THESIS

Comparison of Frequency and Severity of Infection

STUDY ONE
RCT Comparison of the frequency of infection (Chapter 2)

RCT Comparison of the cost of infection (Chapter 3)

Meta-analysis of the prophylactic use of ceftriaxone against other antibiotics. Comparison of wound infection, chest infection and urinary infection (Chapter 6)

Predicting Wound infection and Risk factors for Wound Infection

STUDY TWO
Prospective study on the prediction of wound infection using a visual analogue scale (Chapter 4)

STUDY THREE
Retrospective review assessing the importance of the ASA score (Chapter 5)

Surgical Prediction of Major Complications

STUDY FOUR
Prospective study of the ability of the surgeon to predict major complications (Chapter 7)
- Accuracy of the prediction of major complication
- Can prediction of complications be improved?
- Comparison of surgical prediction to POSSUM
- Comparison of preoperative and postoperative prediction
- Including surgical prediction in risk models
- Lessons learned with the use of a visual analogue scale

STUDY FIVE
Prospective study assessing the difference between complication rates at discharge from hospital and at one month after surgery (Appendix 3)

Figure 1.2 Overview of thesis
2. A prospective randomised double blinded controlled trial comparing antibiotic prophylaxis with Ceftriaxone and Cefotaxime in abdominal surgery. A comparison of the frequency of infection

2.1 Rationale for study

2.1.1 Choice of two third generation cephalosporins.
Selecting the most appropriate prophylactic antibiotic regimen from the many available options includes consideration of clinical efficacy, safety and cost. Cefotaxime and ceftriaxone are both members of the very broad spectrum class of third generation cephalosporins, and have activity superior to most other single agents against the great majority of gram positive and gram negative pathogens. They are therefore a logical choice for comprehensive surgical prophylaxis.

2.1.2 In-vitro activity of study antibiotics
Although the in-vitro effectiveness of these antibiotics is very similar (Chambers 1992; Chau and Ng 1984), cefotaxime also has an active metabolite, desacetyl cefotaxime (Jones et al. 1987), which has an activity profile similar to cefuroxime. Studies which assess the combined in-vitro activity of cefotaxime and its metabolite desacetyl cefotaxime show a reduction in the MIC 90 against a range of bacteria including enterobacteriaceae, extended spectrum β-lactamase producing Klebsiella species, other gram negative species, anaerobes and Staphylococcus aureus (Jones et al. 1992a; Stratton et al. 1989; Canawati 1992). The additional activity of desacetyl cefotaxime therefore gives cefotaxime an improved in-vitro performance when compared to ceftriaxone. This improvement was proposed to be a ‘synergistic action’ resulting in a promotional campaign by Roussel entitled “Clavforan the power of synergy.”

2.1.3 Pharmacokinetic considerations
In terms of pharmacokinetics ceftriaxone has a protein binding of 90-95% (Stoekel et al. 1981), no active metabolite, a half-life of eight hours and a low total body clearance (Mazzei and Periti 1989). Ceftriaxone’s tissue penetration index (TPI), based on the tissue suction technique is 92% (Mazzei and Periti 1989). In contrast cefotaxime has a protein binding of 32-44% (Esmieu et al. 1980), a half-life of one hour and a TPI of 71% (Mazzei and Periti 1989). The desacetyl metabolite has a half-life of two hours. Antibiotics with a low rate of
serum protein binding (<65%) usually have a good extravascular distribution and a high penetration index. Ceftriaxone however is unique among the cephalosporins in that it combines a low rate of total body clearance with a high protein binding. This results in a prolonged and high concentration of the antibiotic in the vascular compartment, which drives the diffusion of the antibiotic through the capillary bed. The extended high serum concentration of ceftriaxone, in comparison to cefotaxime (and most other antibiotics) is illustrated in Figure 2.1.

![Graph showing serum concentration of Ceftriaxone and Cefotaxime over 24 hours following intravenous administration of 1g of antibiotic.](image_url)

**Figure 2.1 Serum concentration of Ceftriaxone and Cefotaxime over 24 hours following intravenous administration of 1g of antibiotic**

### 2.1.4 Cost
The average wholesale price in New Zealand for 1g of intravenous ceftriaxone was $32 and for 1g of intravenous cefotaxime was $15. The difference in the cost does vary from country to country, with ceftriaxone consistently being the more expensive of the two antibiotics. The cost effectiveness of the prophylactic use of these two antibiotics is compared in the next chapter.

### 2.1.5 Safety profile and single dose prophylaxis
Both antibiotics are safe with few side effects. They have a low incidence of allergy and are not nephrotoxic.
The duration of prophylaxis is also relevant when considering the potential adverse effects of prophylactic antibiotics (Turano 1992). Although colonization with *Clostridium difficile* and a variety of fungi can be identified after a single dose of a third generation cephalosporin (Westh et al. 1991; Privitera et al. 1991), *Clostridium difficile*, diarrhoea and fungal superinfection are more frequent when greater than one dose of antibiotic is given (Starks et al. 2008). Issues relating to *Clostridium difficile* infection and antibiotic resistance are presented in Appendix A.

Overall single dose prophylaxis combines the advantages of providing an adequate cover against wound infection (Song and Glenny 1998; Rowe-Jones et al. 1990; M McDonald et al. 1998) and of minimizing side effects. It also allows us to study if the pharmacokinetic differences (half-life, protein binding, tissue penetration index) between these two antibiotics are important in preventing postoperative infection.

### 2.1.6 Need for a well-designed randomised controlled trial comparing these two antibiotics

The null hypothesis being tested in this chapter is that there is no difference in the frequency of infection after abdominal surgery when prophylactic ceftriaxone or cefotaxime is used. Although both antibiotics are excellent prophylactic agents, studies directly comparing these antibiotics are few and lack statistical power (Anderson et al. 1996; Blatzas et al. 1987; Cardamakis et al. 1991; Germiniani et al. 1988; Kunz et al. 1989; Periti et al. 1984b; Petropoulos et al. 1985). Our objective was to compare the prophylactic use of these agents with a population of sufficient size to achieve the statistical power to detect a difference in efficacy of 5% between the regimens.

### 2.2 Methods of study

#### 2.2.1 Inclusion, randomization, blinding

Inclusion criteria:

Patients admitted into a general surgical unit for acute and elective abdominal surgery were consecutively recruited over a three year period in a prospective, randomised, double blind study.
Exclusion criteria:
1) Active infection requiring treatment before or at the time of surgery,
2) Use of other antibiotics within the 48 hours before surgery,
3) Death within 30 days of surgery when no infection had developed,
4) Cephalosporin or metronidazole allergy,
5) Patient or surgeon withdrawal.
6) Significant deviation from the trial protocol (incorrect timing of antibiotic administration, incorrect administration of metronidazole, incorrect randomized antibiotic being given). Because of the similarities of the two antibiotics it was felt that incorrect administration should be a criteria for exclusion, rather than to include these cases on an ‘intention to treat’ basis.

Randomisation, Stratification, Blinding
A clinical hospital pharmacist, using a blocked randomised technique (cell size of 10) with stratification, independently randomised the patients to antibiotic A or antibiotic B. The "non colorectal" stratification included hepatobiliary, esophageal-gastric-duodenal surgery, pancreatic surgery, splenectomy, small bowel surgery, appendicectomy and vascular surgery. The "colorectal" stratification included all colorectal surgery and stomal surgery.

2.2.2 Antibiotic administration
One gram of intravenous cephalosporin (antibiotic A or B) was administered by the anesthetist at induction of anesthesia. Five hundred milligrams of intravenous metronidazole was also given, at induction of anesthesia, for those in the “colorectal” stratification.

2.2.3 Definition of endpoints
The major endpoint was wound infection with pus formation, or cellulitis. Cellulitis was defined as erythema and induration greater than 1cm laterally, for at least two thirds of the length of the wound or 6cm, whichever was the lesser.

The secondary endpoints included the following:
**Deep Peritoneal infection** required radiological or operative demonstration of a collection with clinical evidence of sepsis or the culture of a pathogen on aspiration. This was further divided into anastomotic morbidity or deep infection. If the deep peritoneal infection was
associated with an anastomotic leak then this was classified as anastomotic morbidity. If there was no demonstrable anastomotic problem then this was classified as a deep infection.

*Chest infection* required at least two of the following: clinical signs of chest infection, purulent sputum with a recognized pathogen and/or radiological demonstration of an infective process. Atelectasis was not included in the definition of a chest infection.

*Urinary tract infection* required a urine with $>10^8$ bacteria/litre or $>10^7$ white blood cells/litre and a positive culture, or correlation with clinical symptoms and signs.

*Febrile morbidity* was defined as an oral temperature greater than 38 degrees Celsius on two occasions, more than 6 hours apart, and greater than 24 hours post operatively.

*Other infections* included septicemia, infective diarrhea with testing for *Clostridium difficile* toxin, yeast superinfection, intravenous line sepsis and drain site infection.

### 2.2.4 Assessment of endpoints

A trained research nurse in conjunction with the surgical team performed clinical evaluation daily from surgery until hospital discharge. Additional investigations such as blood cultures, midstream urine, chest radiology or a wound swab were performed when clinically indicated. Patients with diarrhea had a faecal specimen tested for *Clostridium difficile* toxin. Subsequently a multidisciplinary team comprising of two surgeons, a medical microbiologist, a clinical pharmacist and the research nurse confirmed the assessment of each case on a weekly basis. All patients were reviewed at a minimum of thirty days post operatively. The thirty day assessment was either as an inpatient, at a surgical clinic (surgical outpatient clinic or a wound clinic run in the surgical outpatient department), or they were telephoned by the research nurse using a standardized questionnaire. The surgical audit was also checked, as this provided an independent record of clinically recognized infections.

### 2.2.5 Microbiology

The microbiology of infections was assessed whenever possible. Wound swabs were cultured aerobically only, unless the Gram stain was consistent with an anaerobic or polymicrobial infection. Sputum specimens were screened for “acceptability” before being processed. Specimens from “deep” or normally sterile sites were cultured aerobically and anaerobically. Identification of Gram negative aerobic and facultative anaerobic bacilli was done using a modified agar dilution, replica-plating method (Burman and Ostensson 1978). Other organisms were identified by standard methods (Ballows 1991). Sensitivity testing of most aerobic pathogens was done by agar-dilution method, using NCCLS break-points (NCCLS
1990a&b). Iso-Sensitest agar (Oxoid) was used for the replica-plating method, except for methicillin sensitivity when Mueller-Hinton agar (Gibco) was used. Other aerobic organism sensitivity testing was performed by disc sensitivity (NCCLS 1990a&b). For fastidious organisms Iso-Sensitest agar was supplemented with 5% defibrinated sheep blood. Anaerobic sensitivity testing was not performed routinely (Ballows 1991). All sensitivity results were recorded as sensitive, intermediate or resistant. Breakpoint sensitivity concentrations for both antibiotics were 8 µg/ml and 64 µg/ml. Cefotaxime pure substance was supplied by Roussel. Desacetylcefotaxime was not tested. Ceftriaxone pure substance was supplied by Roche.

2.2.6 Statistical analysis
Sample size: The trial was designed to test for a 5% difference in infection rates (the ability of one antibiotic to reduce the wound infection rate from 10% to 5%), with a confidence of 95%, a type I error of 5% and a power of 80%. Using a Chi square analysis with a two-tailed alternative hypothesis a sample size of 434 was required in both arms of the trial.

The chi-squared test was used to test for differences in endpoint and microbiology results. Yates correction and a two-tailed Fisher’s exact test were used when there were small numbers. The analysis was performed on an intention to treat basis. An interim analysis was to be performed after 400 patients had been recruited and then at intervals of 200.

2.2.7 Ethical approval and consent
Ethics committee approval to undertake this study was obtained from the Otago Regional Ethics Committee. A patient information sheet, explaining the study was given to patients by the clinical staff (medical or nursing) looking after the patient. Questions were answered by staff and written consent for the study was completed at the same time as the consent for surgery was obtained.

2.2.8 Pharmaceutical Involvement in study
Both Roche (Ceftriaxone) and Roussel (Cefotaxime) were involved in the design and funding of this study. Roche provided a monitor who had access to the results. The monitor also had access to the initial draft of the scientific paper, but he was not involved in the analysis of the results and did not make any contribution to publication.
2.3 Results

2.3.1 Inclusion and exclusion
In total 1013 patients were randomised, 93 (9%) were excluded from analysis, 44 in the ceftriaxone group and 49 in the cefotaxime group. The numbers for each exclusion criteria were comparable (Table 2.1).

Table 2.1 Exclusion from study

<table>
<thead>
<tr>
<th>Reasons for exclusion</th>
<th>Ceftriaxone</th>
<th>Cefotaxime</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number Randomised</td>
<td>506</td>
<td>507</td>
</tr>
<tr>
<td>Therapeutic antibiotics</td>
<td>25</td>
<td>24</td>
</tr>
<tr>
<td>Death within 30 days</td>
<td>16</td>
<td>17</td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Final number in analysis</td>
<td>462</td>
<td>456</td>
</tr>
</tbody>
</table>

2.3.2 Patient demographic and operative data
The patient demographic data for the two antibiotics was comparable (Table 2.2). There were however some differences in the operative data. By chance cefotaxime was used significantly more in intermediate rather than major surgery and in acute rather than elective surgery. A significant proportion of this difference was due to the distribution of appendicectomies, which are analyzed separately (see below). With the appendicectomies removed there were no longer any significant differences between the two antibiotics.

2.3.3 Wound infections (superficial and deep surgical site infection)
The overall wound infection rate (Table 2.3) was 8% with ceftriaxone compared to 12% with cefotaxime, p<0.05. At the interim analysis after 800 patients a significant difference was already observed. This difference was due to appendicectomies being performed without metronidazole, with wound infection rates of 6% and 18% for the different antibiotics. The antibiotic code was not broken and subsequent appendicectomies were entered into the "colorectal" strata. At the completion of the study it was identified that for appendicectomy without metronidazole the wound infection was 6% with ceftriaxone and 18% with cefotaxime. When the bias from appendicectomies without metronidazole were excluded, the rates of wound infection were similar (ceftriaxone 8%, cefotaxime 10%), with infection rates
of 6% with ceftriaxone and 10% with cefotaxime for “non colorectal” surgery and of 10% for “colorectal” surgery for both agents.

Table 2.2 Patient demographic and operative data

<table>
<thead>
<tr>
<th></th>
<th>Ceftriaxone</th>
<th>Cefotaxime</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number</strong></td>
<td>462</td>
<td>458</td>
</tr>
<tr>
<td><strong>Age (years ± SD)</strong></td>
<td>51.6 (21.3)</td>
<td>49.6 (21.5)</td>
</tr>
<tr>
<td><strong>Sex, M/F</strong></td>
<td>212/250</td>
<td>226/232</td>
</tr>
<tr>
<td><strong>Duration of surgery (minutes ± SD)</strong></td>
<td>93 (66)</td>
<td>92 (68)</td>
</tr>
<tr>
<td>**Timing of surgery ***</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Non elective</td>
<td>223</td>
<td>261</td>
</tr>
<tr>
<td>Elective</td>
<td>239</td>
<td>197</td>
</tr>
<tr>
<td><strong>Operation Category</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major</td>
<td>333</td>
<td>306</td>
</tr>
<tr>
<td>Intermediate</td>
<td>129</td>
<td>152</td>
</tr>
<tr>
<td><strong>ASA score</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Grade 1</td>
<td>114</td>
<td>129</td>
</tr>
<tr>
<td>Grade 2</td>
<td>105</td>
<td>102</td>
</tr>
<tr>
<td>Grade 3</td>
<td>64</td>
<td>58</td>
</tr>
<tr>
<td>Grade 4 and 5</td>
<td>13</td>
<td>10</td>
</tr>
<tr>
<td><strong>Wound Category</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clean</td>
<td>105</td>
<td>99</td>
</tr>
<tr>
<td>Possibly contaminated</td>
<td>128</td>
<td>124</td>
</tr>
<tr>
<td>Contaminated</td>
<td>169</td>
<td>180</td>
</tr>
<tr>
<td>Dirty</td>
<td>60</td>
<td>55</td>
</tr>
<tr>
<td>Gastrointestinal anastomoses</td>
<td>187</td>
<td>177</td>
</tr>
<tr>
<td>Drains</td>
<td>184</td>
<td>160</td>
</tr>
<tr>
<td><strong>Surgeon</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consultant</td>
<td>196</td>
<td>188</td>
</tr>
<tr>
<td>Resident</td>
<td>266</td>
<td>270</td>
</tr>
</tbody>
</table>

Statistical comparison: Two-tailed chi square test and two tailed Student’s t Test.

ASA = American Society of Anesthesiologists Score

* p<0.05
2.3.4 Deep peritoneal infection (space surgical site infection)
Deep peritoneal infection was divided into deep infection and anastomotic morbidity. There was no significant difference in deep infection. The difference in anastomotic morbidity (Table 2.4) approached significance, being 3% with ceftriaxone and 8% with cefotaxime, p=0.054 following Yates correction.

Table 2.3 Wound infection

<table>
<thead>
<tr>
<th></th>
<th>Ceftriaxone</th>
<th>Cefotaxime</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>All cases (Before removal of ‘non colorectal’ appendicectomies)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Non colorectal stratification”</td>
<td>17/276</td>
<td>36/281</td>
<td>0.007</td>
</tr>
<tr>
<td>“Colorectal stratification”</td>
<td>19/186</td>
<td>18/177</td>
<td>0.87</td>
</tr>
<tr>
<td>Total</td>
<td>36/462</td>
<td>54/458</td>
<td>0.04</td>
</tr>
<tr>
<td>Appendicectomy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“No Metronidazole”</td>
<td>5/83</td>
<td>18/100</td>
<td>0.03</td>
</tr>
<tr>
<td>“Metronidazole”</td>
<td>2/23</td>
<td>3/27</td>
<td>0.85</td>
</tr>
<tr>
<td>Wound infection after removal of ‘non colorectal’ appendicectomies</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>“Non colorectal stratification”</td>
<td>12/193</td>
<td>18/181</td>
<td>0.26</td>
</tr>
<tr>
<td>“Colorectal stratification”</td>
<td>19/186</td>
<td>18/177</td>
<td>0.87</td>
</tr>
<tr>
<td>Total</td>
<td>31/379</td>
<td>36/358</td>
<td>0.45</td>
</tr>
</tbody>
</table>

Statistics: Two-tailed chi square. *Italic* Yates correction used

2.3.5. Distant infection
There were no significant differences in any individual distant infection. The three most commonly identified distant infections: chest infection, urinary tract infection and septicaemia were less frequent with ceftriaxone. Superinfections with *Clostridium difficile*, diarrhea or a yeast infection were less frequent with cefotaxime (Table 2.4). However it should be noted that the majority of cases with a superinfection also had additional antibiotics for another postoperative infection. Febrile morbidity was similar for the two antibiotics. On review approximately half of cases with febrile morbidity had a clinical diagnosis of pulmonary atelectasis in the early postoperative period.

2.3.6 Combined infection
All appendicectomies that did not receive metronidazole were excluded from the analysis of the combined endpoints (Table 2.5) because of the potential bias introduced by cefotaxime
giving inadequate cover for appendicectomy. The number of patients that developed either a chest or urinary infection, or both, was significantly reduced with ceftriaxone (ceftriaxone 6%, cefotaxime 11%, p=0.01). The number of patients developing “any infection”, which included all endpoints except febrile morbidity (see above), was also significantly reduced with ceftriaxone (ceftriaxone 20%, cefotaxime 27%, p=0.03).

### Table 2.4 Peritoneal and distant infections

<table>
<thead>
<tr>
<th></th>
<th>Ceftriaxone</th>
<th>Cefotaxime</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>462</td>
<td>458</td>
<td></td>
</tr>
<tr>
<td>Deep Peritoneal Infection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Deep Infection</td>
<td>3</td>
<td>8</td>
<td>0.22</td>
</tr>
<tr>
<td>Anastomotic morbidity*</td>
<td>6/187</td>
<td>15/177</td>
<td>0.054</td>
</tr>
<tr>
<td>Distant Infection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chest infection</td>
<td>11</td>
<td>19</td>
<td>0.19</td>
</tr>
<tr>
<td>Urinary infection</td>
<td>14</td>
<td>23</td>
<td>0.17</td>
</tr>
<tr>
<td>Clostridium difficile</td>
<td>4</td>
<td>0</td>
<td>0.12*</td>
</tr>
<tr>
<td>Yeast infection</td>
<td>12</td>
<td>7</td>
<td>0.36</td>
</tr>
<tr>
<td>Febrile Morbidity</td>
<td>23</td>
<td>25</td>
<td>0.74</td>
</tr>
<tr>
<td>Septicaemia</td>
<td>2</td>
<td>6</td>
<td>0.18*</td>
</tr>
<tr>
<td>IV line infection</td>
<td>6</td>
<td>9</td>
<td>0.59</td>
</tr>
<tr>
<td>Drain site infection</td>
<td>3/184</td>
<td>5/160</td>
<td>0.48*</td>
</tr>
</tbody>
</table>

Statistics: Two-tailed chi square test, with Yates correction (italics) and two tailed Fischer’s exact (*) test as indicated by small numbers.

IV=intravenous

### Table 2.5 Combined infections

<table>
<thead>
<tr>
<th>Infection</th>
<th>Ceftriaxone</th>
<th>Cefotaxime</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>379</td>
<td>358</td>
<td></td>
</tr>
<tr>
<td>Chest and urinary</td>
<td>23</td>
<td>40</td>
<td>0.01</td>
</tr>
<tr>
<td>CLD and Yeast</td>
<td>12</td>
<td>6</td>
<td>0.28</td>
</tr>
<tr>
<td>Any</td>
<td>77</td>
<td>97</td>
<td>&lt;0.05</td>
</tr>
</tbody>
</table>

CLD=Clostridium difficile diarrhoea.

Any infection is wound infection, deep surgical site infection and all distant infections. Febrile morbidity is not included. Two-tailed chi square test performed. Italics: Yates Correction
2.3.7 Microbiology
There were 66 wound infections with a cultured growth, of which 42% grew skin microbes, 35% enteric microbes and 23% a mixture of both. Twenty-nine (44%) wound infections were polymicrobial.

Differences in wound microbiology (Table 2.6) were due to Staphylococcal infections. *Staphylococcus aureus* was the most important pathogen. When compared to the number of patients receiving an antibiotic, it was cultured from 7 of 462 patients who were given ceftriaxone and in 22 of 458 patients given cefotaxime, p<0.01. The difference was still significant when the upper stratification appendectomy cases were removed. When compared to the number of cultures that grew a pathogen (Table 2.6), it was identified in 29 of 66 cultures, including 29% and 52% of positive wound cultures for ceftriaxone and cefotaxime respectively, p=0.11. It was also, unexpectedly, the most frequently isolated organism with chest infection, being cultured on one occasion with ceftriaxone and on six occasions with cefotaxime.

In contrast the incidence of *Staphylococcus epidermidis* was more common when ceftriaxone was used. The number of cases with a *Staphylococcus epidermidis* wound infection was not significantly different (8/462 v 2/458, p=0.11), but it was isolated in 33% and 5% of positive wound cultures for ceftriaxone and cefotaxime respectively, p<0.01.

The organisms isolated from deep peritoneal and urinary tract infections were comparable. The antibiotic sensitivity profile was similar for both antibiotics. The most frequent resistant organisms cultured were *Enterococcus faecalis* and *Pseudomonas aeruginosa*. No MRSA were isolated.

2.4 Discussion
2.4.1 Wound infection (superficial and deep incisional surgical site infection)
2.4.1.1 Both antibiotics provided effective prophylaxis for wound infection
In terms of the frequency of infection both antibiotics were confirmed to give effective prophylaxis against wound infection when used alone for non colorectal operations and when used with metronidazole for colorectal operations, with rates for wound infection similar to that previously reported (Wittmann and Condon 1991).
Table 2.6 Wound microbiology

<table>
<thead>
<tr>
<th></th>
<th>Ceftriaxone</th>
<th>Cefotaxime</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wound infections</td>
<td>36</td>
<td>54</td>
<td></td>
</tr>
<tr>
<td>Cultures</td>
<td>27</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Cultures with growth</td>
<td>24</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td><strong>Microbes</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Staphylococcus aureus</em></td>
<td>7</td>
<td>22</td>
<td>0.11</td>
</tr>
<tr>
<td><em>Staphylococcus epidermidis</em></td>
<td>8</td>
<td>2</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td><em>Staphylococcus species</em></td>
<td>3</td>
<td>4</td>
<td>0.50*</td>
</tr>
<tr>
<td><em>Bacteroides fragilis</em></td>
<td>8</td>
<td>9</td>
<td>0.44</td>
</tr>
<tr>
<td><em>Escherichia coli</em></td>
<td>4</td>
<td>11</td>
<td>0.56</td>
</tr>
<tr>
<td><em>Enterococcus faecalis</em></td>
<td>3</td>
<td>3</td>
<td>0.77</td>
</tr>
<tr>
<td><em>Streptococcus anginosus</em></td>
<td>1</td>
<td>6</td>
<td>0.39</td>
</tr>
<tr>
<td><em>Klebsiella pneumonia</em></td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td><em>Klebsiella oxytoca</em></td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><em>Proteus mirabilis</em></td>
<td>0</td>
<td>4</td>
<td>0.29*</td>
</tr>
<tr>
<td><em>Pseudomonas aeruginosa</em></td>
<td>3</td>
<td>2</td>
<td>0.34*</td>
</tr>
<tr>
<td><em>Morganella morganii</em></td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>3</td>
<td></td>
</tr>
</tbody>
</table>

Statistics: Two-tailed chi square test with Yates correction (Italics) and two tailed Fischer’s exact test (*)

2.4.1.2 Ceftriaxone provided better wound infection prophylaxis for appendicectomy

The difference in antibiotic performance occurred in the appendicectomy patients that did not receive metronidazole, where cefotaxime alone did not provide satisfactory prophylaxis. This difference was no longer the case when appendicectomy patients were placed into the ‘colorectal’ stratification. It had previously been shown that metronidazole alone helps to reduce wound infection after appendicectomy (Tanner et al. 1980; Rodgers et al. 1979), and that the addition of a cephalosporin further reduces the rate of wound infection (Morris et al. 1980; Miholic et al. 1983). However ceftriaxone alone (without metronidazole), a third generation cephalosporin, with good anaerobic cover, had been used effectively as prophylaxis in both appendicectomy (Lang et al. 1988) and colorectal surgery (Lumley et al. 1992). Although this had not been tested with cefotaxime, the use of cefotaxime was
supported by the observation that its in-vitro activity when combined with its desacetyl metabolite was superior to that of ceftriaxone (Stratton et al. 1989; Canawati 1992). The inadequate cover of cefotaxime for appendicectomy is however consistent with one other study where one dose of cefotaxime alone was used as prophylaxis in colorectal surgery (Jones 1990).

2.4.1.3 Why would ceftriaxone provide better prophylaxis following appendicectomy?
The results of this study demonstrated that using cefotaxime alone as prophylaxis for appendicectomy was a mistake. The question to answer is why was the performance of ceftriaxone, in spite of the better in-vitro activity of cefotaxime and desacetyl cefotaxime, significantly better than cefotaxime? The most likely explanation for this is related to differences in the pharmacokinetics of the two antibiotics. The excellent TPI of ceftriaxone results in a concentration of ceftriaxone in the wound which is greater than the concentration of cefotaxime (Figure 2.1), and is also well above the MIC 90 of most bacteria (Figure 2.2). In other words the differences in the TPI and the concentration of antibiotic in the wound were microbiologically more important that the differences in the in-vitro performance of the two antibiotics. The observation that the difference in the performance of the two antibiotics was no longer present when metronidazole was used points towards the role of bacterial synergism between aerobic and anaerobic organisms (Brook 1985) being important in the establishment of these infections. This observation suggests that ceftriaxone was more effective in eradicating anaerobic organisms and therefore more effective in minimizing the impact of bacterial synergism.

2.4.1.4 Differences in wound microbiology
Although the frequency of wound infection was similar once the “non colorectal” appendicectomies were corrected for, differences in the two antibiotic groups were still observed in the frequency of organisms isolated and in the severity of infection. *Staphylococcus aureus* was more frequently isolated with cefotaxime and *Staphylococcus epidermidis* with ceftriaxone. There were no microbiological differences (in terms of antibiotic resistance) to explain this difference. The presence of *Staphylococcus epidermidis*, a skin commensal, in wound infections when broad-spectrum antibiotics are being used (Lang et al. 1988) suggests it may have the ability to colonize wounds when other bacteria, including other skin commensals and more pathogenic bacteria (including *Staphylococcus*
*aureus* are inhibited. The issue of the severity of infection is studied in the next chapter which reports on the cost of infection.

**Figure 2.2** Serum concentration of ceftriaxone compared to the MIC 90 of most pathogenic bacteria

### 2.4.2 Deep peritoneal infection (surgical space infection)

The difference in anastomotic morbidity fell just short of significance. The possibility of an anastomotic infection resulting in an increase in collagenolytic activity and adversely
influencing anastomotic integrity has been demonstrated (Yamakawa et al. 1971; Irvin 1976), but animal and clinical studies have given inconsistent results (Gutman et al. 1993; Irvin Golligher 1973; Peoples et al. 1990). The most important factors in preventing an anastomotic leak are good blood supply, no tension and a technically well-performed anastomosis. Other factors such as steroids, severe intra-abdominal infection and a postoperative period of hypotension are also occasionally important. The differences in anastomotic morbidity in this study are most likely due to chance differences in blood supply and surgical technique. The influence of this on the overall results is addressed in Section 2.4.4, where the antibiotics are compared after cases with anastomotic morbidity have been excluded.

2.4.3 Distant infection

Although there was no significant difference in any individual distant infection endpoint, it was interesting to note that the trend towards a lower frequency of chest infection, urinary infection and septicaemia were all in the same direction, in favour of ceftriaxone. This is consistent with a significant reduction in either chest infection (Morris 1994; de la Hunt et al. 1985), urinary tract infection (Morris 1993) or both of these infections (Lumley et al. 1992) which has been reported in selected studies with ceftriaxone prophylaxis.

As there is wide variation in the incidence of chest infection and urinary infection in studies looking at prophylactic antibiotic use in abdominal surgery, it was difficult to calculate a sample size to assess for these infections. The study was therefore sized for the primary endpoint, wound infection. Using the frequency infection that observed in this study it would be necessary to have a sample size of approximately 1500 patients in both arms of a study to detect a significant difference in urinary infection and a sample size of 1700 patients to detect a significant difference in chest infection (with a confidence of 95% and a power of 80%). To assess for either chest infection and/or a urinary infection would require a sample size of 478 in both arms of the study. The combined chest and urinary infection endpoint in this study, which was significantly less in the ceftriaxone group, approximated this. Caution needs to be taken when interpreting these results. Although this observation is consistent with previous studies, the study is underpowered to draw a conclusion about ceftriaxone reducing chest infection alone or urinary infection alone. A second reason for causation is that chest and urinary infection were secondary endpoints, and as such may not have been as thoroughly assessed as the primary endpoint. The difference in chest and/or urinary infections that was
noted in this study was the stimulus for performing a meta-analysis of ceftriaxone prophylaxis. This is presented in Chapter 6.

The observation that ceftriaxone was effective in reducing the frequency of some infections remote to the wound, raises the possibility that prophylactic antibiotics may have a greater role than just the prevention of wound infection. For optimal effectiveness prophylactic antibiotics should be at therapeutic tissue concentration at the time of bacterial contamination. For chest infection and urinary infection the time of contamination includes both the intraoperative period and the postoperative period (Section 1.5). The sustained 24-hour (Shinagawa 1988) antimicrobial cover given by a single intravenous dose of ceftriaxone may therefore contribute to reducing the incidence of these infections.

2.4.4 Differences between the two antibiotics when cases with anastomotic morbidity are excluded
Differences between the antibiotics, including differences in the wound microbiology and in the frequency of either a chest and/or a urinary infection, were still significant when cases with anastomotic morbidity were removed from the analysis. The overall difference in ‘any infection’ however was no longer significant, with infection rates of 19% with ceftriaxone and 24% with cefotaxime, p=0.10.

2.5 Limitations of this study
Although this was a well-designed study, some limitations are noted. Firstly there was no critical assessment of Clostridium difficile infection. This was a secondary endpoint, but should have been given more attention. Secondly placing appendicectomy into the upper abdominal stratification (no metronidazole given) proved to be an incorrect decision, and made the interpretation of results for the primary endpoint (wound infection) more complicated. Thirdly with respect to the differences in chest and/or urinary infection it needs to be emphasised that these are secondary endpoints. Fourthly the ‘chance difference’ in anastomotic morbidity had the potential to influence other endpoints and the wound microbiology. The potential impact of this on other results is reported in 2.5.4
2.6 Conclusions

2.6.1 Similarities between the antibiotics
In terms of the primary endpoint (wound infection), the frequency of infection when the two antibiotics were appropriately used (including when metronidazole was used for appendicular surgery) was the same. The null hypothesis is accepted. In terms of secondary endpoints there was no difference in the frequency of ‘any infection’ when cases with anastomotic morbidity were excluded from analysis.

2.6.2 Differences between the antibiotics
However some of the results suggested that ceftriaxone may be more versatile. In terms of the primary endpoint (wound infection) it is less dependent on metronidazole as an adjunct and provides a more effective prophylactic cover against *Staphylococcus aureus*. In terms of secondary endpoints it may have the additional benefit of reducing chest and urinary infections.

2.6.3 The importance of a therapeutic concentration of antibiotic during the time of bacterial contamination
Although cefotaxime and desacetyl cefotaxime had a superior in-vitro performance to ceftriaxone, the study results for some of the secondary endpoints, and for the wound infection when metronidazole was not given, suggested that ceftriaxone performed more robustly than cefotaxime. The most likely reason for this is the excellent tissue penetration index of ceftriaxone, resulting in a therapeutic concentration of antibiotic well above the MIC 90 of most bacteria in the serum, wound, chest and urine (Figure 2.2). Knowledge of the concentration of the antibiotic at the site of contamination throughout the duration of the surgical procedure is important when interpreting in-vitro predictions about the relative effectiveness of different antibiotics.

2.7 Further questions for study
1) In the context of ‘clean-contaminated’ and ‘contaminated’ wounds, does an antibiotic with an antimicrobial activity well above the MIC 90 for most bacterial pathogens perform better than an antibiotic which is just above the MIC 90?
2) Does ceftriaxone prevent chest and urinary infection? Although this study is consistent with previous studies that suggest this is the case, it was underpowered to assess the question. This issue is addressed in Chapter 6.

3) If prophylactic antibiotics can decrease distant infections, is this primarily due to the activity of the antibiotic during the time of surgery, or is more prolonged antibiotic prophylaxis required to achieve this? This leads into the question: Is there an ideal pharmacokinetic profile for a prophylactic antibiotic?

2.8 Publication
CHAPTER 3

3. Using cost of infection as a tool to demonstrate a difference in prophylactic antibiotic efficacy. A prospective randomised double blinded comparison of the pharmacoeconomic effectiveness of Ceftriaxone and Cefotaxime prophylaxis in abdominal surgery.

3.1 Introduction

In the economic environment in which we work the cost of healthcare cannot be ignored. The increase in healthcare expenditure over the last two decades (Angell 1985; Bochner et al. 1996; Visual Economics 2010) has emphasized the limited availability of resources, and the need for optimizing health benefits within an available budget. Postoperative infection may double hospital stay, substantially increase cost (Wenzel 1992; Green and Wenzel 1977) and result in disability which can last for more than six months (Leapee et al. 1991). As the cost of antibiotic prophylaxis is minimal compared to the cost of an infection, the use of prophylactic antibiotics is a good example of a cost effective investment.

The selection of a preferred prophylactic antibiotic should therefore assess cost effectiveness, as well as the ability of an antibiotic to prevent infection. A comprehensive cost effectiveness analysis includes an assessment of all the effects (good and bad) and all the costs of an intervention. It is not a cost containment exercise. It does not focus on just the pharmaceutical cost or the direct costs of antibiotic administration. Nor does it consider the costs from the perspective of one department. Rather it looks at the best value for money spent from the perspective of the whole institution.

Although the relative merit of prophylactic antibiotics is traditionally evaluated by a comparison of the number of infections, it is well recognised that all infections are not the same. Some infections are minor or low cost whereas others are very disabling and expensive. The possibility of comparing differences in severity of infection between antibiotics has been recognized and attempts have been made to quantify this with the use of a scoring system (Wilson et al. 1986; Hall and Hall 2004). Studies have also been performed looking at the cost effectiveness of different prophylactic antibiotics which have assumed that the cost of the same postoperative infection is the same for each antibiotic (Anderson et al. 1996; Davey et al. 1992). However the severity of infection may vary with different antibiotics, and this
difference may be the most important factor in determining the actual cost effectiveness of different prophylactic antibiotic regimes.

In this study the cost of infection was used as a direct measurement of the severity of infection. The null hypothesis being tested in this chapter is that there is no difference in the cost of infection after the prophylactic administration of ceftriaxone or cefotaxime. The alternative hypothesis is that the cost of infections not prevented may be used as a tool to demonstrate a difference in antibiotic efficacy. An accurate assessment of the overall cost of infection was collected prospectively in the setting of a randomised controlled trial (RCT), which also assessed the gold standard of the frequency of infection for the antibiotics being compared.

3.2 Methods

3.2.1 Methods pertaining to the frequency of infection

The comparison of the cost of infection was performed as part of the RCT presented in Chapter 2. The criteria for inclusion and exclusion, methods of randomization and blinding, method of antibiotic administration, the definition of the infective endpoints, the assessment of the infective endpoints and the process of ethical approval and patient consent have been presented.

3.2.2 Assessment of the cost of infection

The cost of postoperative infection was measured in a prospective and blinded manner. Inpatient costs were based on hospital costs for the investigation and treatment of infection. A multidisciplinary team, which included two consultants and a research nurse determined by consensus if extra days spent in hospital were caused by postoperative infection or not. The marginal costs for hospital stay were obtained by allocating each additional hospital day into one of four cost categories, determined by the patient’s daily level of dependency. Laboratory, radiology, pharmaceutical and disposable item expenses were costed using the in-house resource utilization system (RUS). As well as charging for direct medical costs, RUS also incorporated indirect costs, including the cost of hospital resources such as equipment, use of buildings, management and overhead costs. RUS had been developed over a number of years by the Health Board and when this study was performed was accepted to be a validated system for costing within the organization. Direct operating theatre costs and physiotherapy costs were attributed separately. The fulltime research nurse recorded the cost of inpatient
infection while the patient was in hospital. The collection of hospital costs was also assisted by a nursing questionnaire.

Outpatient costs included the cost of outpatient clinics, accident and emergency care and the use of community nursing services (district nursing services). Community costs included loss of wages by the patient or spouse or family, the use of primary care and private healthcare services including laboratory and pharmaceutical charges, insurance expenses and supplements to income by social welfare. These costs were collected by the use of a patient questionnaire and also a prospective community nurse questionnaire. The research nurse was responsible for ensuring these forms were correctly compiled.

3.2.3 Statistics
The power calculation and statistics used for the comparison of the frequency of infection are presented in Chapter 2. For the comparison of the mean and median cost of infection the Statistical Package for Social Sciences (SPSS 11.0) for Windows (SPSS Institute, Cary, NC, USA) was used. The distribution of the cost of infection for each endpoint and its goodness of fit to normal distribution was assessed using descriptive statistics including skewness, kurtosis, Lilliefors test for goodness of fit to the normal distribution curve and a visual inspection of the distribution of the cost of infection. As the cost of infection was non-parametric, the Mann Whitney U (MWU) test was used for antibiotic comparison. A logarithmic transformation was also performed and when the transformed data passed the goodness of fit test, antibiotic comparisons were performed using the Student’s t test (Coyle 1996). If a patient had more than one infection the total cost of infection was included in the analysis of both endpoints, and once in the analysis of the combined endpoints.

3.3 Results
3.3.1 Inclusion and exclusion, patient demographic and operative data
These are presented in Chapter 2

3.3.2 The cost of infection
The total cost of infection for the 920 (ceftriaxone 462, cefotaxime 458) patients was $510,231. The range was from no cost to $64,436. Anastomotic morbidity was the most expensive endpoint totaling $230,936 (range $1,902 - $64,436). Wound infection was the most frequent endpoint, with a total cost of $156,125 ($0 - $31,914). Ninety one percent of
the cost of infection was inpatient, 6% outpatient and 3% community cost. The breakdown of the cost of infection is presented in Figure 3.1. Increase in hospital stay accounted for 56% (95% Confidence Interval, 37%-75%) of the total cost of infection and re-operations accounted for 14% (7-21%) of the total cost of infection. Other inpatient costs were evenly spread over radiology (3%, 2.2-4.1), microbiology (3%, 2.4-3.5) and pharmaceuticals (4%, 2.0-5.3). Community nurse expenses accounted for 73% (46-100%) of outpatient costs. Loss of wages accounted for 77% (0%-100%) of community costs.

3.3.3 Comparison of the cost of infection for each antibiotic

3.3.3.1 Overall cost
Antibiotic comparison showed that ceftriaxone use was associated with 27% and cefotaxime with 73% of the total cost of infection. The mean cost of infection each time ceftriaxone was given was $298. The mean cost each time cefotaxime was given was $814. With the addition of the cost of the antibiotic, its storage and administration (Hall et al. 1988b; Atkinson et al. 1989; Smyth et al. 1995) the respective costs were $335 and $832.

3.3.3.2 Increase in hospital stay
The cost of additional days spent in hospital because of infection was the main contributor to the overall cost of infection. This cost was also significantly different for the two antibiotics (Figure 3.1). When considering all infective endpoints, more patients had severe infections resulting in an increased hospital stay in the cefotaxime group (28 Ceftriaxone and 55 Cefotaxime, p<0.005). The four main infections which caused an increase in hospital stay were wound infection, chest infection, anastomotic morbidity and deep infection. The number of times these four infections caused an increase in hospital stay for each antibiotic was: for ceftriaxone – wound infection 11, chest infection 6, anastomotic morbidity 5, deep infection 3; and for cefotaxime – wound infection 22, chest infection 11, anastomotic morbidity 11, deep infection 8.
Figure 3.1 Breakdown of the total cost of infection for both antibiotics

When a patient had an infection severe enough to result in an increase in hospital stay, the cost of the additional days in hospital was similar for both antibiotics. That is when an infection was severe enough to cause an increase in hospital stay, it was ‘equally severe’ for both antibiotics. The mean cost of the increase in the hospital stay was $2670 (SD ± 2349) for ceftriaxone and $4195 (SD ± 6583) for cefotaxime, p=0.93. The proportion of the extra days that were spent in each of the four levels of dependency was also similar for both antibiotics (Ceftriaxone: I – 22%, II – 14%, III – 53%, IV – 10%. Cefotaxime: I – 17%, II – 28%, III – 40%, IV – 16%). Category IV, intensive care unit stay, accounted for 31% of the hospital stay costs.

For wound infection 11 of 36 ceftriaxone and 22 of 54 cefotaxime wound infections (chi-square, p=0.33), resulted in an increase in hospital stay (the significant difference in the number of patients having an increase in hospital stay which was noted for all infective endpoints was not present for any individual endpoint). The mean increase in stay was 4.4 SD ± 10.2 days, 1.5 SD ± 3.6 days with ceftriaxone and 6.7 SD ± 13.0 days with cefotaxime. A significant proportion of the wound infections were diagnosed after the patient was discharged from hospital (Appendix D discusses this issue further).
3.3.3.3 Assessment of the total cost of infection with anastomotic morbidity removed
This was assessed because of the chance differences in anastomotic morbidity between the two antibiotics, and because anastomotic morbidity was the most expensive endpoint. When deep peritoneal infection (space surgical site infection) due to anastomotic morbidity was excluded the total cost of infection for the 920 patients was $290,069, with a range from no cost to $31,914. The breakdown of the total cost of infection is similar to when anastomotic morbidity was included. Eighty five percent of the cost of infection was inpatient, 10% outpatient and 5% community cost. Increase in hospital stay accounted for 57% (37%-76%) of the total cost. Other costs were spread over reoperation 9% (2%-16%), radiology 3% (2%-4%), microbiology 4% (3%-5%), pharmaceuticals 4% (2%-6%), community nurse expenses 9% (6%-13%) lost wages 4% (0%-8%) and other costs. On antibiotic comparison ceftriaxone was associated with 32% and cefotaxime 68% of the cost of infection. The mean cost of infection, the antibiotic and its administration was $239 each time ceftriaxone was administered and $448 each time cefotaxime was administered. The difference in hospital stay was also maintained with 23 ceftriaxone and 45 cefotaxime treated patients experiencing an increase in hospital stay due to infection, p<0.005. The mean increase in hospital stay was 3.3 days (SD ±9 days) being 1.6 (SD ± 5.6) days with ceftriaxone and 4.6 (SD ± 11) days with cefotaxime.

3.3.3.4 Wound infection
There was a significant difference in the frequency of appendicectomy wound infection when metronidazole was not given (ceftriaxone 6%, cefotaxime 18%, p<0.03). The costs of these infections however were similar (Table 3.2). For all other cases there was no difference in the frequency of wound infection (ceftriaxone 8%, cefotaxime 10%) but there was a significant difference in the cost (ceftriaxone $887, SD ± $1743, cefotaxime $2995, SD ± $6592, p<0.05) and in the microbiology of these infections.

The details of the wound microbiology are presented in Chapter 2. Staphylococcus aureus (STA) was identified more frequently when cefotaxime was used and Staphylococcus epidermidis (STE) isolated more frequently in ceftriaxone wounds. There was no difference in the mean cost of Staphylococcus aureus and Staphylococcus epidermidis infections (STA $2374 [SD ± $7049] v STE $954 [SD ± $1660], p=0.82). However the cost for both Staphylococcal infections was significantly less expensive when ceftriaxone was used,
compared to when cefotaxime was used. (ceftriaxone $506, SD ± $1301 v cefotaxime $2950, SD ± $7667, p<0.05).

Table 3.1 Cost of superficial and deep incisional surgical site infections

<table>
<thead>
<tr>
<th>All cases (excluding appendectomy without metronidazole)</th>
<th>Ceftriaxone</th>
<th>Cefotaxime</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>379</td>
<td>358</td>
<td></td>
</tr>
<tr>
<td>Number with infection</td>
<td>31 (8%)</td>
<td>36 (10%)</td>
<td>0.37</td>
</tr>
<tr>
<td>Mean Cost $ (±SD)</td>
<td>887 (±1743)</td>
<td>2995 (±6592)</td>
<td>0.015*</td>
</tr>
<tr>
<td>Median Cost $ (range)</td>
<td>170 (19-6767)</td>
<td>824 (21-31914)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Appendectomy without metronidazole</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>83</td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Number with infection</td>
<td>5 (6%)</td>
<td>18 (18%)</td>
<td>0.01</td>
</tr>
<tr>
<td>Mean Cost $ (±SD)</td>
<td>994 (±1101)</td>
<td>878 (±1368)</td>
<td>0.33*</td>
</tr>
<tr>
<td>Median Cost $ (range)</td>
<td>614 (66-2860)</td>
<td>330 (0-4557)</td>
<td></td>
</tr>
</tbody>
</table>

Chi square test used to compare frequency of infection with Fischer’s exact test to adjust for small cell numbers.

* Student’s t test performed on logarithmically transformed data

3.3.3.5 Distant infection: Individual endpoints

There was no significant difference in the cost (Figure 3.2) or the frequency of any individual distant infection. The details of the frequency of infection are presented in Chapter 2. Anastomotic morbidity was the most expensive endpoint, with a mean cost of $7370 (SD ± 3839) with ceftriaxone and $12448 (SD ± 16846) with cefotaxime, p=0.76.

3.3.3.6 Combined endpoints

The number of patients having either or both a chest and a urinary infection were significantly reduced with ceftriaxone compared to cefotaxime (6% v 11%, p=0.01), but the cost of these infections was similar (Table 3.3). As in the frequency of infection study appendicectomy cases not receiving metronidazole were excluded.

Both the frequency and cost of all infections, which for the cost effectiveness study also includes febrile morbidity, (Table 3.4) was significantly less with ceftriaxone (ceftriaxone 24%, cefotaxime 31% p<0.05, cost ceftriaxone $1441 SD ± $2686, cefotaxime $3331 SD ±
$8392, p<0.05). When anastomotic morbidity was excluded the frequency of all infection was no longer significantly different, but the cost of infection was (ceftriaxone 22%, cefotaxime 27%, p NS, cost ceftriaxone $910 SD ± $1809, cefotaxime $1872 ± SD $4539, p<0.05).

![Mean Cost of Infection](image)

**Figure 3.2** Comparison of the mean cost of different infections following the use of prophylactic antibiotics in abdominal surgery

Mean cost of infection with 95% confidence intervals

### 3.4 Discussion

#### 3.4.1 Frequency and cost of infection are complementary measurements

**3.4.1.1 Overview**

Frequency of infection is the gold standard for antibiotic comparison. In this study cost of infection has been used as a measurement of the severity of the infection. These two ‘tests’ of antibiotic performance examine different aspects of the antibiotic performance, and in this
study were able to independently discriminate between the two antibiotics being examined. Both the frequency and cost (severity) of infections not prevented were different for each prophylactic antibiotic. The ability of the cost of infection to demonstrate a significant difference between prophylactic antibiotics had not previously been demonstrated in the context of a randomised controlled trial. Differences in the cost of infections not prevented can therefore be used as a tool to assess differences in the performance of prophylactic antibiotics.

**Table 3.2 Cost of chest and urinary infection**

<table>
<thead>
<tr>
<th></th>
<th>Ceftriaxone</th>
<th>Cefotaxime</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>379</td>
<td>358</td>
<td></td>
</tr>
<tr>
<td>Number with infection</td>
<td>23 (6%)</td>
<td>40 (11%)</td>
<td>0.01</td>
</tr>
<tr>
<td>Mean Cost $ (±SD)</td>
<td>1272.91 (±2338)</td>
<td>1614.66 (±4083)</td>
<td>0.64</td>
</tr>
<tr>
<td>Median Cost $ (range)</td>
<td>284 (59-9874)</td>
<td>258 (5-22861)</td>
<td></td>
</tr>
</tbody>
</table>

**Table 3.3 Cost of all infection**

<table>
<thead>
<tr>
<th></th>
<th>Ceftriaxone</th>
<th>Cefotaxime</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>379</td>
<td>358</td>
<td></td>
</tr>
<tr>
<td>All Infection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>90 (24%)</td>
<td>110 (31%)</td>
<td>0.03</td>
</tr>
<tr>
<td>Mean Cost $ (±SD)</td>
<td>1441 (±2686)</td>
<td>3331 (±8392)</td>
<td>0.04*</td>
</tr>
<tr>
<td>Median Cost $ (range)</td>
<td>197 (0-11521)</td>
<td>610 (0-64436)</td>
<td></td>
</tr>
<tr>
<td>All Infection excluding infection associated with anastomotic morbidity</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frequency</td>
<td>84 (22%)</td>
<td>96 (27%)</td>
<td>0.14</td>
</tr>
<tr>
<td>Mean Cost $ (±SD)</td>
<td>910 (±1809)</td>
<td>1872 (±4539)</td>
<td>0.03*</td>
</tr>
<tr>
<td>Median Cost $ (range)</td>
<td>170 (0-9872)</td>
<td>306 (0-31914)</td>
<td></td>
</tr>
</tbody>
</table>

Appendicectomy cases not receiving metronidazole are excluded
Chi square test used to compare frequency of infection
* Student’s t test performed on logarithmically transformed data

The frequency and cost of infection were also ‘complementary’ with the differences in the frequency of infection and in the cost of infection always being in in favour of the same
antibiotic. Statistically this would not have to always be the case, but it is likely that if a prophylactic antibiotic does not perform well against infection that both more frequent and more severe infections will be present.

The cost of infection was also able to demonstrate differences not noted by the frequency of infection. In this study, for some endpoints, the cost of infection was significantly different when the frequency of infection was the same. The cost of infection was therefore demonstrated to increase the discriminatory power of a study comparing prophylactic antibiotic performance, over and above the gold standard of the observed frequency of infection.

### 3.4.1.2 Wound infection

The complementary nature of the frequency and cost of infection were demonstrated for the primary endpoint, wound infection. There were two groups, due to appendicectomy initially being placed in the ‘noncolorectal’ stratification (section 2.4.1). For appendicectomy when metronidazole was not given the frequency of wound infection was greater when cefotaxime was used, but the cost of wound infections was the same. For all other cases the frequency of wound infection was the same, but the cost or severity of infection was significantly greater with cefotaxime. The microbiology did demonstrate some differences between the wound infections (although it was interesting to note that this difference in the cost of infection was not primarily due to the differences demonstrated in the wound microbiology). Both frequency and cost of infection were therefore able to demonstrate a complementary difference in antibiotic performance.

### 3.4.1.3 Combined distant infection and all infections

The effectiveness of both the frequency and cost of infection in showing differences in antibiotic performance was also demonstrated for infections other than wound infection. Ceftriaxone decreased the frequency of chest infection and urinary infection, but the severity or cost of these infections was the same. In contrast ‘all infection’ with all infections considered together, showed a significant difference in both the frequency and cost of infection in favor of ceftriaxone.
3.4.2 Impact of removing anastomotic morbidity

To control for the potential impact of chance differences on other endpoints it was decided to analyse the results both with and without anastomotic morbidity. When cases with anastomotic morbidity were excluded there was no significant change in the breakdown of inpatient, outpatient and community costs or in the ratio of the costs of infection for each antibiotic. The significant differences in wound infection and in chest and urinary infection were also independent of anastomotic morbidity. However the results for ‘all infection’ changed. The frequency of ‘all infection’ was no longer significantly different, although the cost of ‘all infection’ did remain significantly different. This again demonstrates the additional discriminatory power of ‘cost of infection’.

3.4.3 Other methods for assessing the severity of infection

Different scoring systems have been used to grade the severity of wound infection. For example, Johnson et al. proposed a four class classification of wound morbidity following infrainguinal bypass surgery (Johnson et al. 1988) which combined both infection and ischaemia. Wilson et al. described a scoring system for assessing the severity of wound infection in cardiac surgery called ASEPSIS (Wilson et al. 1986). This is an acronym for Additional treatment, Serous discharge, Erythema, Purulent exudate, Separation of deep tissues, Isolation of bacteria and duration of inpatient Stay. This was designed to be a reproducible way of grading the severity of wound infection in clinical trials of prospective antibiotic prophylaxis. It gives additional information about the nature and extent of wound morbidity, but is also influenced by non-infective events such as wound haematoma, and has not yet been demonstrated to enhance the ability of a study to discriminate between different antibiotics (Hall and Hall 1996b; Greif et al. 2000).

Hall et al. have also prospectively developed a scoring system for breast surgery (Hall and Hall 2004). This identified five criteria associated with infection of the wound (purulent exudate, serous discharge, presence of pathogens, cellulitis, wound dehiscence) and four criteria that were a consequence of infection (extra hospital stay, wound dressings, antibiotic use, surgical drainage), giving a possible wound score out of one hundred. One of the benefits of this scoring system was that it helped to identify a category of wound pathology due to infection which did not fit the conventional criteria of wound infection. Using the criteria of discharge of pus or a serous discharge containing pathogens, 12 clinical wound infections (5.5%) were identified. However there were also 44 cases with cellulitis, of whom 27% had a
wound score of greater than 40 points. In another study the median wound score for patients with a ‘conventional’ wound infection was 60 (Hall et al. 2006). However to date no scoring system assessing the severity of infection has been able to demonstrate a statistically significant difference in the severity of infection between different prophylactic antibiotics.

Using cost of infection therefore has advantages over ASEPSIS (and other scoring systems for grading the severity of wound morbidity) as it can increase the ability of a study to discriminate between different antibiotics.

3.4.4 Implications of these results for cost effectiveness studies on antibiotic use

The finding that there is a difference in the cost of infections not prevented by different prophylactic antibiotics highlights a fundamental problem in the design of many studies comparing antibiotic cost effectiveness. Firstly, the most important cost is the cost of infection. Studies looking at savings in drug cost and administration in isolation (Murphy and Omo 1992; Pearce and Begg 1994), such as those looking at the cost-effectiveness of the introduction of an automatic therapeutic substitution policy, do not even begin to address the most important part of the cost effectiveness equation. Secondly, the cost of infection is not the same for different antibiotics. It has previously been recognized that the cost of wound infection varies with the type of operation (Green and Wenzel 1977; Lynch et al. 1992b) but our study has also demonstrated that the cost of the same postoperative infection is different with different antibiotics. The results of our study confirm that an adequate comparison of the cost effectiveness of antibiotics cannot be performed unless the cost of the infection prevented, or the cost of the infection being treated, is documented for each of the antibiotics being assessed. Although measuring differences in the cost of infection is difficult, this is an essential requirement for accurately determining the real cost effectiveness of different antibiotic regimens.

3.4.5 Methodology for assessing the cost of infection

In terms of assessing the cost of infection, this study included four important methodological features. These were assessment of infection in the context of a randomised controlled trial, assessing all aspects of the cost of infection, measuring the marginal cost of hospital stay, and following up patients for an adequate length of time.
3.4.5.1 Measuring cost as part of a randomised controlled trial

Although it has been recommended that the prospective costing of infection is the best method (Lynch et al. 1992b; Davey et al. 1995) for assessing antibiotic cost effectiveness, we are not aware of any other study that has recorded both the frequency and cost of infection using a prospective, randomised double blind design. Although this increases the complexity of a RCT it does add economic relevance and also increases the ability of the study to discriminate between antibiotics. On those occasions when there is apparent drug equivalence then any significant cost differences will determine antibiotic choice.

Another advantage of measuring cost of infection in a RCT is that a direct comparison is made between the antibiotics, which enable conclusions about relative antibiotic efficacy to be made, without reference to other studies being necessary. Variations in study protocols, endpoint definitions, length of follow-up and different methods of hospital charging between countries (Davey et al. 1992; Solomon et al. 1994; Wakefield et al. 1988; Pena et al. 1996) makes the direct comparison of results from different studies difficult, and potentially misleading. In the case of this study, both the cost of infection (Green and Wenzel 1977; Davey et al. 1992; Lynch et al. 1992b; Daschner 1989; Hayley et al. 1981; Coello et al. 1993; Kirkland et al. 1999; Miller et al. 1989; Mason, and Garcia 1984; Pinner et al. 1982; Couch et al. 1978) and length of hospital stay (Green and Wenzel 1977; Hayley et al. 1981; Coello et al. 1993; Kirkland et al. 1999; Haley et al. 1980; Penin and Ehrenkranz 1988; Delgado-Rodriauez et al. 1997) were generally lower than previously reported, being less than reported elsewhere when ceftriaxone was used, but within the range of previous reports when cefotaxime was used.

3.4.5.2 Method for identifying and costing extra days in hospital

One reason for our lower than expected length of hospital stay due to infection (and therefore lower costs) was the opportunity to diagnose infections up to 30 days postoperatively, resulting in a number of infections being diagnosed after discharge from hospital. Another reason was our method of determining the number of extra hospital days attributable to infection at a weekly review by our multidisciplinary team. Increase in hospital stay due to social circumstances or due to medical problems unrelated to any postoperative infection was therefore identified, and not counted as additional days in hospital because of infection. This is consistent with work comparing different methods for attributing extra hospital days due to infection (Hayley et al. 1980; Penin and Ehrenkranz 1988), which has shown that the
physician’s assessment of extra days in hospital is less than a more general calculation of extra days in hospital based on comparing patients in different diagnostic groups with and without infection.

As increase in hospital stay was the greatest individual cost, it was important to correctly identify the actual cost of extra days spent in hospital because of infection. The recognition of four levels of dependency (level of care required) gave a more accurate indication of the actual (or marginal) cost of each extra day in hospital, rather than the alternative method of charging the average bed cost for each additional hospital day. The multidisciplinary team was responsible for assigning a dependency level to each extra day spent in hospital.

3.4.5.3 Assessing all aspects of the cost of infection
A full itemization of inpatient, outpatient and community costs was performed. Inpatient costs were the most important category, accounting for approximately 90% of the total cost of infection. The observation that carefully recorded outpatient costs are low compared to hospital costs was confirmed (Davey et al. 1992). The main community cost, loss of earnings, was less than expected. The main reason for this was that the majority of patients in the study were not in active employment.

3.4.5.4 Duration of patient assessment
In terms of follow-up, the importance of identifying all infections in the first 30 postoperative days is emphasized in a number of studies that confirm that up to 40-70% of wound infections may not be clinically apparent at the time the patient is discharged from hospital (Davey et al. 1988; Law et al. 1990; Lynch et al. 1992a; Mitchell et al. 1999). However in terms of the cost, these infections are usually less severe (not as expensive) than infections that develop prior to discharge from hospital (Davey et al. 1992; Hall 1999).

3.4.5.5 Resources required to gather good data on the cost of infection
Measurement of the total cost of infection is both time-consuming and expensive (Hall 1999). In this study a full time clinical nurse researcher combined with weekly team meetings to discuss and assess all cases with infective complications. Others have used trained surveillance nurses in conjunction with a physician epidemiologist (Haley et al. 1981). The accurate documentation of the cost of infection involved frequent inpatient clinical review, the use of questionnaires, and the follow-up of all patients at approximately day 30 (Mitchell et al.
1999). This required a motivated full time clinical researcher, good communication with the nursing staff and a committed surgical team.

3.4.6 Limitations of this study
Setting up the methodology for collecting inpatient, outpatient and community costs of infection was challenging, especially as this had not been done previously. In the methods an overview of this process is given, but the details are not presented. Difficulties included developing three questionnaires, one for the ward nursing staff, one for the patient and one for the district nursing service. There was no study performed to validate these questionnaires.

However the main limitation of using cost of infection as a tool to demonstrate a difference in the performance of prophylactic antibiotics is the time, effort and cost that this involves. Simpler ways of measuring severity of infection would have obvious advantages. One option we assessed was the number of patients spending additional days in hospital. This was significantly different when considering all infections. However it was not as reliable as the total cost of infection, which was also able to identify significant differences for individual endpoints (for example wound infection). Further work on assessing if the number of additional days spent in hospital is a potentially useful tool for identifying differences in the severity of infection between antibiotics would be useful.

3.5 Conclusions
3.5.1 Using cost of infection as a tool to compare prophylactic antibiotics
The study supported the alternative hypothesis, that the cost of infections not prevented may be used as a tool to demonstrate a difference in antibiotic efficacy. This study demonstrated the ability of both the frequency and cost of infection to demonstrate differences in antibiotic performance. In two situations the frequency of infection was significantly different but the cost of infection was similar (‘non colorectal’ appendectomy wound infection and chest and/or urinary infection). In two situations the frequency of infection was similar but the cost of infection was significantly different (wound infection excluding ‘non colorectal’ appendectomy and all infection excluding anastomotic morbidity). On one occasion both the frequency and cost of infection were significantly different (all infection including anastomotic morbidity). The differences in both the frequency and cost of infection were complimentary as they were always in favour of the same antibiotic.
Cost of infection is an effective tool in demonstrating a difference in antibiotic efficacy. This has an advantage over other methods of assessing the severity of infection, as it may be possible to demonstrate a statistically significant difference in the cost of infection even when the frequency of infection is not significantly different. This study therefore supports including both the frequency of infection and the cost of infection in prospective trials of prophylactic antibiotic efficacy.

3.5.2 Comparison of efficacy of antibiotics
Although both antibiotics are effective prophylactic agents this second study has also demonstrated that ceftriaxone has some advantages over cefotaxime. These included a lower frequency of wound infection for appendectomy when metronidazole was not used, a lower cost of all other wound infections, a lower frequency of chest and urinary infections and a lower cost of ‘all infection.’

3.6 Further questions for study
As this is the first RCT comparing both the frequency and cost of infection there are a number of issues that require further research and development. These include…
1) The development and validation of the best methods for recording the total cost of infection (inpatient, outpatient and community costs).
2) Further assessment of different ways of measuring the severity of infection.
3) Validation of the above results, which would require a further RCT assessing the cost of infection to be performed.

3.7 Publication
CHAPTER 4
4. Accuracy of the surgeon’s clinical prediction of wound infection using a visual analogue scale

4.1 Introduction
4.1.1 Why predicting wound infection may be useful
Wound infections account for approximately a quarter of all nosocomial infections (Nichols 2001), and may cause significant morbidity, increasing both the length of the hospital stay and the cost of the operative procedure (Woodfield et al. 2005). The accurate prediction and documentation of wound infection therefore continues to be an important clinical objective. This has practical implications for the hospital, surgeon and patient, and may lead to measures that best minimize infection, particularly in high risk patients.

4.1.2 Overview of risk factors for wound infection
Risk factors which contribute to wound infection can be divided into host factors, microbiological and environmental factors, and surgical factors (Perl and Roy 1995).

Host factors are related to the ability of the host to fight infection and to the efficiency of wound healing. These functions require an adequate wound oxygen tension, and are also affected by a range of comorbidities. These comorbidities include: pre-existing diseases (diabetes, obesity, smoking, renal failure, liver failure, solid and hematologic neoplasia, infection at a concurrent site, autoimmune diseases, AIDS), concomitant therapies (corticosteroid, cytotoxic agents) and poor nutritional status.

Microbiological factors include the dose and the virulence of the bacteria introduced into the wound. This is influenced by the pattern of bacterial resistance in the hospital, and is an important factor in patients with a prolonged stay in hospital prior to surgery (Mishriki et al. 1990).

Surgical factors include the classification of the wound (Cruse and Foord 1973; Cruse and Foord 1980; Garibaldi et al. 1991; Hayley et al. 1985; Culver et al. 1991), duration of the surgical procedure (Garibaldi et al. 1991; Hayley et al. 1985; Culver et al. 1991), surgical site preparation (Cruse and Foord 1980; Tanner et al. 2007; Darouiche et al. 2010), intraoperative
contamination (Garibaldi et al. 1991), use of prophylactic antibiotics, and the technical expertise of the surgeon.

The combination of adverse host, microbiological and surgical risk factors significantly increases the incidence of a postoperative wound infection (Culver et al. 1991; Platell and Hall 2001).

4.1.3 Timing of prediction
Some of these risk factors are known preoperatively and can be used to predict risk, to determine the need for antibiotic prophylaxis, and to make decisions about preoperative intervention (in an attempt to reduce the impact of medical comorbidities on infection). However other factors, which may be estimated preoperatively, can only be accurately quantified at the time of surgery. It is only during the operation that the severity of any underlying infective process and the extent of the contamination of the wound can be quantified. The surgeon also influences the risk of a wound infection intraoperatively by technical steps taken to minimize wound contamination, avoid tension in wound closure, clear the wound of devitalized tissue and to prevent wound haematoma (Leeper 1995). As the surgeon is most aware of these events he/she may be in an ideal position to assess the likelihood of a wound infection developing. If the surgeon is able to predict a substantially increased risk of a wound infection during surgery, this may also have implications for modifying prophylaxis during the perioperative procedure (Bates et al. 1989).

4.1.4 Prediction of wound infection using a visual analogue scale
Although individual risk factors for wound infection have been identified, infection is influenced by the interaction of many factors, and it was unclear how accurate the clinical prediction of wound infection would be in the individual patient. The aim of this prospective ‘pilot study’ was to assess the ability of the surgeon to predict the risk of a wound infection immediately before and after surgery using a VAS (section 1.14). This method had been shown to be helpful in predicting major complications following surgery, but had not previously been used to assess the ability of the surgeon to prediction of wound infection.
4.2 Methods of this study

4.2.1 Prospective study, performed as part of a randomised controlled trial

A total of 1013 consecutive patients undergoing abdominal surgery, who were eligible for entry into a study on prophylactic antibiotic use (Woodfield et al 2003), were also entered into this study assessing the likelihood of a wound infection.

4.2.2 Use of a visual analogue scale

The surgeon prospectively estimated the risk of a wound infection in the operating room immediately before and after surgery by marking his/her prediction on a 100mm VAS. If the score was changed postoperatively the surgeon was asked to document the reason for this. For the purposes of analysis a difference in the VAS score of >5mm was categorised as a change in the prediction of the risk of an infection.

4.2.3 Definitions, data collection, clinical evaluation, ethics approval

This is presented in Chapter 2. Additional demographic data and risk factors for wound infection including the wound classification, duration of surgery, the acuteness of surgery, the organ system being operated on and the patient’s age were also prospectively recorded.

4.2.4 Statistics

The Mann-Whitney U test was used to compare the VAS score of patients with an infection against those who did not have a wound infection. The chi square test was used to assess associations between categorical variables and the proportions of patients with a wound infection. For ordered categorical variables the significance of the trend between an increasing variable score and an increasing incidence of wound infection was tested using the Cochran-Armitage test. Different methods for predicting a wound infection were compared using the area under the receiver operating characteristic curve (c statistic). Logistic regression analysis of the different risk factors contributing to the development of a wound infection was performed in the Statistical Package for the Social Sciences version 11.0 (SPSS Institute, Cary, NC, USA). The surgeon’s VAS score, patient’s age and the duration of surgery were entered as continuous variables. The acuteness of surgery was entered as three categorical variables and the wound classification as four categorical variables. Possible contributions from interactions between the independent variables were tested and the presence of multi-collinearity was checked by using variance inflation factors.
4.3 Results

4.3.1 Patient demographics and operative details
Patient demographics and operative details are presented in section 2.3.2. The study included a wide range of elective and acute abdominal procedures. Of these, 252 operations were classified as emergencies, 291 as urgent and 470 as arranged. Types of surgery were: vascular 79, upper gastrointestinal 65, biliary 126, small bowel 75, colorectal 292, appendix 247, liver/spleen/pancreas 34, hernias 22, gynaecological 38 and ‘other’ 35.

4.3.2 Accuracy of preoperative and postoperative prediction
The surgeon’s preoperative VAS score for the risk of a wound infection was completed in 1013 patients and the postoperative score was completed in 1011 patients. There were 98 wound infections. Although the surgeon’s preoperative prediction of wound infection was non-discriminating; the surgeon’s postoperative prediction did significantly predict wound infection (p=0.01, Table 4.1).

Table 4.1 VAS scores for patients with and without a wound infection

<table>
<thead>
<tr>
<th></th>
<th>Preoperative Scores</th>
<th>Postoperative Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WI</td>
<td>No WI</td>
</tr>
<tr>
<td>Number</td>
<td>98</td>
<td>915</td>
</tr>
<tr>
<td>Mean (standard deviation)</td>
<td>22.4 (15.5)</td>
<td>20.6 (13.6)</td>
</tr>
<tr>
<td>Median (range)</td>
<td>19.5 (2-74)</td>
<td>17 (0-95)</td>
</tr>
<tr>
<td>p Value</td>
<td>P=0.41</td>
<td>p=0.01</td>
</tr>
</tbody>
</table>

The Mann-Whitney U test was used for comparison of those with and without complications

WI is wound infection

The postoperative VAS score for a wound infection was increased in 190 patients, not changed in 718 and decreased in 103 patients (Figure 4.1). An increase in the VAS score was associated with 7 more SSI being correctly identified. The improved prediction was predominately in patients with a low preoperative VAS score (Wound infection rate of 9%)
Figure 4.1 Flow diagram demonstrating the incidence of wound infection in low and high risk preoperative categories and the influence of changing the prediction of wound infection postoperatively

WI is wound infection

Low risk: Up to and including the median preoperative prediction, 0-17mm

High risk: Above the median preoperative prediction, >17mm
in whom the risk of an infection was perceived to have increased during surgery (Wound infection rate of 17%). Changing prediction in the high risk preoperative group was not associated with a better prediction of wound infection.

**Table 4.2 Reasons for changing the risk of wound infection postoperatively**

<table>
<thead>
<tr>
<th>Category</th>
<th>Increased Risk</th>
<th>Decreased risk</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cases with change in risk</td>
<td>190</td>
<td>101</td>
<td>291</td>
</tr>
<tr>
<td>Number of cases with a reason stated</td>
<td>136</td>
<td>28</td>
<td>164</td>
</tr>
<tr>
<td>Reason not stated</td>
<td>54</td>
<td>73</td>
<td>127</td>
</tr>
</tbody>
</table>

**REASONS GIVEN FOR CHANGING THE VAS SCORE**

<table>
<thead>
<tr>
<th>Reason</th>
<th>Number of cases with a reason stated</th>
<th>Number of cases with &gt;1 reason</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of reasons</td>
<td>183</td>
<td>30</td>
<td>213</td>
</tr>
<tr>
<td>Number with &gt;1 reason</td>
<td>35</td>
<td>2</td>
<td>37</td>
</tr>
</tbody>
</table>

**TECHNICAL**

<table>
<thead>
<tr>
<th>Reason</th>
<th>Number of cases with a reason stated</th>
<th>Number of cases with &gt;1 reason</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of operation</td>
<td>21</td>
<td></td>
<td>21</td>
</tr>
<tr>
<td>Difficulty of surgery</td>
<td>15</td>
<td>7</td>
<td>22</td>
</tr>
<tr>
<td>Size of operation</td>
<td>16</td>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>Anastomotic</td>
<td>9</td>
<td>3</td>
<td>12</td>
</tr>
<tr>
<td>Other stated technical</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Contamination</td>
<td>51</td>
<td>4</td>
<td>55</td>
</tr>
<tr>
<td>Bleeding</td>
<td>4</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Injury to organ</td>
<td>10</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Other</td>
<td>4</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Total for technical</td>
<td>130</td>
<td>23</td>
<td>153</td>
</tr>
</tbody>
</table>

**EXTENT OF DISEASE**

<table>
<thead>
<tr>
<th>Reason</th>
<th>Number of cases with a reason stated</th>
<th>Number of cases with &gt;1 reason</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Contamination/sepsis</td>
<td>33</td>
<td>5</td>
<td>38</td>
</tr>
<tr>
<td>Ischaemia</td>
<td>6</td>
<td>1</td>
<td>7</td>
</tr>
<tr>
<td>Malignancy</td>
<td>11</td>
<td></td>
<td>11</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Wrong diagnosis</td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total for extent of disease</td>
<td>52</td>
<td>7</td>
<td>59</td>
</tr>
</tbody>
</table>

**ANAESTHETIC/MEDICAL**

<table>
<thead>
<tr>
<th>Reason</th>
<th>Number of cases with a reason stated</th>
<th>Number of cases with &gt;1 reason</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAS = visual analogue scale</td>
<td>1</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>
4.3.4 Reasons given by the surgeon for changing the risk of a wound infection at the completion of the operation

The reasons given for changing the surgeon’s VAS score following surgery were due to technical, disease related or anaesthetic/medical reasons in 71%, 28% and 1% respectively (Table 4.2).

The five main technical reasons stated were intraoperative contamination, the magnitude of the surgical procedure, the technical difficulty of the surgery, the duration of the surgery and anastomotic issues. It was not documented whether the stated technical reasons were inevitable, if they were contributed to by the technical competence of the surgeon, or if they were due to a technical error. The main disease related reason for changing the surgeon’s VAS score was the extent of sepsis found at surgery. When no reason was given for a change in the VAS score the magnitude of the change was smaller (p<0.05) and there was more likely to be a reduction (rather than an increase) in the risk of infection (p<0.05).

4.3.5 Accuracy of other predictors of wound infection

There was a significant association between the wound classification (Cruse & Foord 1980) and the rate of wound infection. Wound infection occurred in 7% of clean wounds, 9% of clean-contaminated wounds, 11% of contaminated wounds and in 13.5% of dirty wounds. Other risk factors including the duration of surgery, acuteness of surgery, organ system being operated on, and the age of the patient did not predict the development of a subsequent wound infection (Table 4.3).

4.3.6 Assessment of the discrimination of prediction and risk factors for predicting wound infection

The area under the Receiver Operating Characteristic (ROC) curves was 0.52 for preoperative prediction, 0.57 for wound classification, and 0.58 for postoperative prediction (Figure 4.2). A result of 0.5 represents random discrimination (the same as tossing a coin), between 0.7 and 0.8 represents reasonable discrimination and a value exceeding 0.8 very good discrimination (Tekkis et al. 2002).
<table>
<thead>
<tr>
<th>Category</th>
<th>Prevalence</th>
<th>Wound Infection (%)</th>
<th>p value *</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Wound classification</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clean</td>
<td>209</td>
<td>12 (5.7)</td>
<td>0.01**</td>
</tr>
<tr>
<td>Clean Contaminated</td>
<td>292</td>
<td>26 (8.9)</td>
<td></td>
</tr>
<tr>
<td>Contaminated</td>
<td>384</td>
<td>43 (11.2)</td>
<td></td>
</tr>
<tr>
<td>Dirty</td>
<td>126</td>
<td>17 (13.5)</td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-39</td>
<td>318</td>
<td>32 (10.1)</td>
<td>0.72**</td>
</tr>
<tr>
<td>40-64</td>
<td>321</td>
<td>29 (9.0)</td>
<td></td>
</tr>
<tr>
<td>65-74</td>
<td>229</td>
<td>27 (11.8)</td>
<td></td>
</tr>
<tr>
<td>≥75</td>
<td>145</td>
<td>10 (6.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Duration of surgery (minutes)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-60</td>
<td>406</td>
<td>37 (9.1)</td>
<td>0.44**</td>
</tr>
<tr>
<td>61-120</td>
<td>392</td>
<td>35 (8.9)</td>
<td></td>
</tr>
<tr>
<td>121-180</td>
<td>130</td>
<td>16 (12.3)</td>
<td></td>
</tr>
<tr>
<td>≥180</td>
<td>82</td>
<td>10 (12.2)</td>
<td></td>
</tr>
<tr>
<td><strong>Urgency of Surgery</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arranged</td>
<td>464</td>
<td>43 (9.3)</td>
<td>0.27**</td>
</tr>
<tr>
<td>Urgent</td>
<td>296</td>
<td>24 (8.1)</td>
<td></td>
</tr>
<tr>
<td>Emergency</td>
<td>253</td>
<td>31 (12.2)</td>
<td></td>
</tr>
<tr>
<td><strong>Organ System</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Vascular</td>
<td>75</td>
<td>5 (6.7)</td>
<td>0.44*</td>
</tr>
<tr>
<td>Hepatobiliary</td>
<td>133</td>
<td>8 (6.0)</td>
<td></td>
</tr>
<tr>
<td>Appendix</td>
<td>232</td>
<td>25 (7.4)</td>
<td></td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>493</td>
<td>53 (10.7)</td>
<td></td>
</tr>
<tr>
<td>Gynaecology, Hernia</td>
<td>76</td>
<td>7 (10.8)</td>
<td></td>
</tr>
</tbody>
</table>

* Chi square test
** Cochran-Armitage test
On univariate analysis the two significant predictors of wound infection were the postoperative VAS prediction by the surgeon and the wound classification. On multivariate analysis the only significant predictor of wound infection was a contaminated or a dirty wound (Table 4.4).
Table 4.4 Logistic regression analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>β coefficient</th>
<th>Wald statistic</th>
<th>p Value (for Wald)</th>
<th>Odds Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Postoperative prediction</td>
<td>.005</td>
<td>.63</td>
<td>.43</td>
<td>1.01 (0.99-1.02)</td>
</tr>
<tr>
<td>Duration of surgery</td>
<td>.002</td>
<td>1.6</td>
<td>.21</td>
<td>1.002 (0.99-1.01)</td>
</tr>
<tr>
<td>Wound classification Clean</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Clean-contaminated</td>
<td>.597</td>
<td>2.5</td>
<td>.12</td>
<td>1.8 (0.86-3.8)</td>
</tr>
<tr>
<td>Contaminated</td>
<td>.776</td>
<td>4.5</td>
<td>.03</td>
<td>2.2 (1.06-4.4)</td>
</tr>
<tr>
<td>Dirty</td>
<td>.924</td>
<td>4.2</td>
<td>.04</td>
<td>2.5 (1.04-6.1)</td>
</tr>
<tr>
<td>Acuteness of Surgery Elective</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Urgent</td>
<td>-.156</td>
<td>.32</td>
<td>.57</td>
<td>.86 (.50-1.5)</td>
</tr>
<tr>
<td>Emergency</td>
<td>.268</td>
<td>.90</td>
<td>.34</td>
<td>1.3 (.75-2.3)</td>
</tr>
</tbody>
</table>

4.4 Discussion

4.4.1 An overview of risk factors used to predict wound infection

The first system widely used to clinically predict the risk of a wound infection was the wound classification system, which identified four categories of wound contamination (Cruse and Foord 1973 and 1980). The initial reported rates of wound infection of 1.5% for clean, 7.7% for clean-contaminated, 15.2% for contaminated, and 40.0% for dirty wounds has subsequently been reduced to <10% for contaminated operations by the introduction of prophylactic antibiotics (Leaper 1995; Sandusky 1980, Wittmann and Condon 1991). However when it was observed that there was a significant variation in postoperative infection rates for clean procedures, broader predictive models, which also included surgical and host risk factors, were developed. The Study on the Efficacy of Nosocomial Infection Control (SENIC) developed an index for risk stratification (Hayley et al. 1985) based on four independent risk factors; laparotomy, duration of surgery over two hours, contaminated or dirty operation, and three or more diagnoses at discharge. This was further modified into the National Nosocomial Infection Surveillance (NNIS) System Index (Culver et al. 1991), which grades the risk for postoperative infection according to three factors: an American Society of Anesthesia (ASA) score of physical status of 3, 4, or 5; a contaminated or dirty operation; and an operation lasting longer than T, where T is the 75th percentile for the time taken for the
procedure to be performed. A further study (Garibaldi et al. 1991) has identified the duration of surgery (odds ratio 3.0), intraoperative contamination (3.0), wound classification (2.7) and an ASA score of physical status (2.4) as independent risk factors on logistic regression analysis. These and other studies have supported the observation that in clean wounds the rate of infection doubles for every hour of surgery (Cruse and Foord 1980).

4.4.2 Comparison of surgical prediction with other predictors of wound infection

In this study both the surgeon’s prediction and established risk factors were poor predictors of wound infection. When asked to prospectively predict the likelihood of a wound infection using a VAS the surgeon’s preoperative prediction, with a c statistic of 0.52, was approximately the same as tossing a coin. The postoperative prediction with a c statistic of 0.58, was slightly better, and was statistically significant, but this result is not discriminating enough for surgeon prediction to be used as a predictive test. The poor discrimination of prediction is confirmed by a difference in median score in the VAS of only 2mm between patients who did, and did not, go on to develop a wound infection.

The results also showed that the established risk factors of wound infection were also unhelpful at predicting wound infection. The duration of surgery was not predictive. The ASA score had not been assessed (See chapter 5). The wound classification system, which along with postoperative prediction, was the only other statistically significant risk factor for wound infection, had a c statistic of 0.57, which was very similar to c statistic for the surgeon’s postoperative prediction.

At logistic regression analysis a contaminated or dirty wound was the only predictor of a subsequent wound infection. This suggests that a major component of the surgeon’s postoperative prediction was determined by the degree of wound contamination, and that once this was adjusted for the other components of the surgeon’s prediction were no longer significant.

4.4.3 Possible reasons why established risk factors for wound infection performed poorly in this study

Our data was collected in the context of a RCT comparing ceftriaxone and cefotaxime. In terms of the duration of surgery descriptive statistics confirms that the range for the time to complete the operation was 10-565 minutes, with 13% of operations lasting 2-3 hours and 8%
lasting for longer than 3 hours. Ceftriaxone has a half-life of 8 hours resulting in a therapeutic tissue concentration which is sustained for 24 hours (Shinagawa 1988). Cefotaxime has a half-life of one hour. However it is metabolized to desacetyl-cefotaxime which is an active antimicrobial agent with a half-life of two hours (Jones et al. 1992a), which extends the time period of antimicrobial prophylaxis. The extended duration of activity of these two cephalosporin antibiotics is the most likely explanation for the duration of surgery not being predictive of a wound infection, as there was a good concentration of antibiotic in the wound for at least three hours after the prophylactic antibiotics were given. Similarly, in terms of the four categories of wound contamination the timely administration of these antibiotics may minimize the impact of the bacterial contamination of the wound.

But why would our results be any different from SENIC, NNIS and other studies? One explanation is that our data was collected in the context of ‘optimal conditions’ a RCT comparing two very effective antibiotics. In contrast, the data used to generate the above risk factors comes from a large number of hospitals operating under ‘normal’ conditions. This means that a wide range of antibiotics (both appropriate and inappropriate), would have been used, and that compliance with antibiotic protocols would range from institution to institution. This may have resulted in less effective antibiotic prophylaxis, enabling these risk factors to have a stronger impact on the development of subsequent wound infections.

4.4.4 Literature supports the observation that the prediction of wound infection is poor

Our review of the literature also showed that the overall prediction of wound infection is poor. In the NNIS study it was possible to categorise patients into low, medium and high risk categories (Culver et al. 1991), but even in the high risk group the risk of a wound infection was only 13%. More recently Brandt and colleagues studied 214,271 procedures and evaluated a number of risk factors; including ASA score, wound class, duration of surgery, endoscopic surgery, gender and age (Brandt et al. 2004). They found that the predictive power of their different logistic regression models was low (with a c statistic between 0.55 and 0.71), and that for most procedures the prediction was only slightly better than the NNIS System risk index.

These studies confirm that even when using statistically significant and well validated predictors of wound infection, the predictive ability and discrimination of these risk factors is
insufficient to help predict between patients who will, or will not, go on to develop a wound infection.

4.4.5 Reasons why the clinical prediction of wound infection is difficult

Five reasons, some general, and some specific to this study, may help to explain why the clinical prediction of wound infection is difficult. Firstly it is important to emphasise that the limited available data available in the medical literature shows that wound infection is difficult to predict (section 4.4.4). The next two reasons are general reasons related to prediction. An event which is relatively uncommon (single number percent risk) is more difficult to predict using any criterion than an event which occurs more frequently. Therefore uncommon events such as predicting falls in the elderly (Oliver, Britton et al 1997) or assessing the risk of suicide (Capodanno, Targum 1983) are difficult. Nevertheless, in contrast to prediction in the individual patient, the discrimination of prediction depends on the ability to identify patients who are more likely to develop a complication from those who will not develop a complication (the prediction is not 0% v 100%), so even with infrequent events if the risk factors are specific enough then prediction with a reasonable discrimination is still possible. A second reason related to prediction is the observation we make during this thesis (chapter 7), that more severe complications, where the magnitude of the event being predicted is greater, are easier to predict than minor complications. For example in a parallel study, performed in the same patients by the same surgeons (section 7.2.1), the c statistic for the postoperative prediction of major complications was 0.69. In both studies in chapter 7 it was also noted that the discrimination of prediction (of major complications) was best when this prediction was used to discriminate between patients who did or did not die, was good when used to discriminate between patients who did or did not develop major complication, and was less discriminating when used to discriminate between any complication or not complication (Table 7.1, Table 7.8). The combination of ‘an infrequent’ and ‘a relatively minor’ complication therefore contributes to the difficulty in predicting wound infection.

The next two reasons are related to the design and methods used in this thesis. The first is our finding that effective prophylactic antibiotics reduce the ability of the clinician to predict wound infection. In this case good prophylaxis reduced the impact of contaminated and dirty wounds, and also provided adequate prophylaxis cover for the duration of the operation. Wound contamination and duration of surgery are risk factors you would expect to be incorporated into the surgeon’s global view of the risk of a wound infection. Reducing the
impact of these risk factors, therefore made wound infection less predictable. The second reason is related to our findings later in the thesis (8.4.2.1.2) that some clinicians did not use the VAS as a linear scale. This inconsistent use of the VAS would have contributed to the prediction of wound infection being under-estimated.

4.4.6 Poor preoperative prediction emphasises the importance of good evidence based protocols
The poor clinical prediction of wound infection means that prior to surgery clinical prediction alone is unlikely to contribute to patient management, especially when appropriate antibiotic prophylaxis is followed. Improvements are more likely to be achieved by ensuring that evidence based protocols are in place and are followed. Adherence to protocols that ensure that an appropriate antibiotic is given at the correct time is the most obvious example of this. Other evidence based interventions that make a difference should also be built into protocols to minimise wound infection. Examples of such interventions would include taking steps to eradicate MRSA colonization, especially prior to vascular surgery (Pearl and Roy 1995; Trautmann et al. 2008; Bode et al. 2010, Bandyk 2008), and optimising blood glucose control in diabetics (Ata et al. 2010; McConnell et al. 2009).

Another approach, that has been shown to make a difference, is to give clinical feedback on infection rates to individual medical staff and consultants (Cruse and Foord 1980; Smyth and Emmerson 2000; Rodriguez et al. 2006). This is discussed further in section 7.8.3.

4.4.7 Improved postoperative prediction. Does this provide an opportunity for intervention?
The improved prediction of wound infection postoperatively highlights that additional information can be identified and that this can potentially contribute to clinical decision making. It is significant that the two main reasons given for changing the risk of a wound infection were intraoperative contamination and the extent of sepsis found at surgery. It is well recognised that when there is intraoperative contamination that the rate of intra-abdominal sepsis and wound infection is increased (Periti et al. 1989; Garibaldi et al. 1991). There is therefore a subset of cases where single dose prophylaxis will not provide adequate antibiotic cover. The present study, where there was a significant increase in wound infection in low risk patients with an increased prediction of infection postoperatively (from 9% to 17%) [these patients were not given additional intraoperative or perioperative antibiotics], supports
this observation. In this setting a clinical decision, whether to provide an additional intraoperative dose of antibiotic (Bates et al. 1989), or to begin a course of therapeutic antibiotics, needs to be made by the surgeon.

The flexibility of intraoperative assessment may therefore provide an opportunity to ‘add’ to the benefit of good antibiotic prophylaxis. The threshold at which additional antibiotic treatment should be recommended and the impact of such a policy on the rates of wound infection requires further study.

4.4.8 Limitations of this study
This was the first time that the clinical prediction of wound infection using a VAS has been assessed. In terms of methodology, it may be that some clinicians did not use the VAS as a linear scale. Although all surgeons were instructed to use the VAS as a linear scale, a further study (section 8.2.2) suggests this may not have been the case. However the preoperative prediction was so poor that is unlikely that using an anchored VAS would have led to a major improvement in the accuracy of prediction.

In terms of antibiotic use, in the primary prophylactic antibiotic study (Chapter 2) it was shown that cefotaxime alone did not give adequate cover for appendicectomy, until an interim analysis was performed. This was not predicted and would also have contributed to a less accurate prediction of wound infection.

Additional data collection, especially of patient co-morbidities and the ASA score of physical status would also have improved the quality of this study (see Chapter 5).

4.5 Conclusions
4.5.1 Prediction using established risk factors and clinical prediction was poor
In this study the wound classification system was only weakly predictive of wound infection, and other established risk factors such as the duration of surgery were not significantly associated with wound infection. One reason for this is that the administration of optimal prophylactic antibiotics diminishes the predictive ability of these established predictors of wound infection.
In addition to this surgical prediction of wound infection was poor, and a review of the literature, which included a logistic regression analysis of recognised risk factors, also confirmed that the ability predict wound infection was poor. Reasons for this include wound infection being an infrequent and relatively minor complication.

These results highlight the importance of using robust preoperative protocols in an effort to further reduce the incidence of wound infection.

4.5.2 Improved postoperative prediction and intraoperative contamination
At the completion of surgery prediction by the surgeon improved, and a change in prediction identified a group of high risk patients. This highlights the importance of identifying intraoperative contamination. The optimal response to the surgeon changing his/her risk prediction of a wound infection intraoperatively requires further study.

4.6 Further questions for study
1) Is the prediction of wound infection, using a visual analogue scale, a useful exercise? Although preoperative surgical prediction alone was unhelpful, there is evidence that the clinical prediction of wound infection using a VAS may be useful because of its ability to identifying patients with an increase in risk of infection postoperatively.
2) What is the optimal response to intraoperative events that in the mind of the surgeon increase the risk of infection? Questions to examine include: How much contamination is required to give additional antibiotics? At what threshold is one extra dose sufficient and at what threshold is a therapeutic course of antibiotics necessary?
CHAPTER 5

5. Does ASA classification of physical status predict wound infection?

5.1 Introduction

5.1.1 How important is the global health of the patient in the development of a wound infection?

Although risk factors which contribute to wound infection can be divided into host factors, microbiological and environmental factors, and surgical factors (Perl and Roy 1995), it was our impression that surgeons place a greater importance on surgical and microbiological factors than on host factors when predicting the likelihood of a wound infection. Whereas most surgeons would recognize that certain systemic diseases, such as diabetes, are important risk factors for wound infection, the importance of the host or of the global health of the patient is often not considered. In the previous study (chapter 4) the prediction of wound infection by the surgeon was poor. It is possible that a more careful consideration of patient co-morbidity would have resulted in a more accurate prediction of infection. Because of this it was decided to assess if the global health of the patient was an independent predictor of wound infection.

5.1.2 The American Society of Anesthesiologists classification of physical status as a measurement of the global health the patient

The American Society of Anesthesiologists (ASA) classification of physical status is a measurement that is contributed to by the functional status, co-existing medical diseases and physiological stability of the patient (ASA 1963). Although this does not identify or directly measure the host mechanisms responsible for preventing wound infection, it does provide an excellent measurement of the global health of the patient. One reason for the widespread popularity and use of this classification system is its simplicity. Although the repeatability of the ASA classification is open to significant variation (Haynes and Lawler 1995), it has consistently been demonstrated to be predictive of postoperative mortality (Vacanti et al. 1970) and morbidity (Menke et al. 1993), and is widely used in models for predicting surgical risk (Tekkis et al. 2002; Sutton et al. 2002; Klotz et al. 1996; Daley et al. 1997).
5.1.3 Impact of ASA classification of physical status in the context of optimal prophylaxis
The ASA score has also been used in a number of studies to predict wound infection (Garibaldi et al. 1991; Culver et al. 1991). In these studies a broad range of patients and procedures were included. There was also no standardised protocol for who would receive prophylactic antibiotics, what antibiotic would be given, and when the antibiotic would be administered. This study was intended to document the association of the ASA classification of physical status with wound infection, in the context of optimal antibiotic prophylaxis in abdominal surgery. The null hypothesis is that the ASA classification is not predictive of wound infection. The alternative hypothesis is that in the context of optimal antibiotic prophylaxis that the ASA classification is the strongest predictor of wound infection.

5.2 Method
5.2.1 Retrospective review of a prospectively performed RCT database
A retrospective review of a prospective study on prophylactic antibiotic use in abdominal surgery was performed. The design of the study, comparing ceftriaxone or cefotaxime given at the induction of anaesthesia, is described in chapter 2. As part of this study demographic and surgical data including the patient’s age, wound classification, acuteness of surgery, and the duration of surgery were prospectively recorded.

The ASA classification of physical status was assigned by the anaesthetist in the operating theatre. Although the ASA score was not a field in the study’s initial data collection (this was not being routinely documented in the hospital when the study commenced), it was increasingly recorded with the patient’s co-morbidities throughout the duration of the study. This resulted in a subset of patients having an ASA score recorded in their antibiotic trial database folder. This data was used for this retrospective study.

Appendicectomies that did not receive metronidazole were not adequately covered by the use of cefotaxime alone (Chapter 2). To prevent this from confounding the current study all appendicectomies that did not receive metronidazole were excluded from this study.

5.2.2 Statistics
The chi square test and Fisher’s exact test were used to assess associations between categorical variables and the proportions of patients with a wound infection.
Table 5.1 Associations between risk factors for wound infection and the incidence of wound infection

<table>
<thead>
<tr>
<th>Category</th>
<th>Prevalence</th>
<th>Wound Infection (%)</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-39</td>
<td>69</td>
<td>6 (8.7)</td>
<td>0.71*</td>
</tr>
<tr>
<td>40-64</td>
<td>135</td>
<td>11 (8.1)</td>
<td></td>
</tr>
<tr>
<td>65-74</td>
<td>110</td>
<td>13 (11.8)</td>
<td></td>
</tr>
<tr>
<td>≥75</td>
<td>68</td>
<td>6 (8.8)</td>
<td></td>
</tr>
<tr>
<td><strong>ASA</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>74</td>
<td>1 (1.4)</td>
<td>0.02*</td>
</tr>
<tr>
<td>II</td>
<td>162</td>
<td>15 (9.3)</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>118</td>
<td>14 (12.0)</td>
<td></td>
</tr>
<tr>
<td>IV/V</td>
<td>29</td>
<td>6 (20.7)</td>
<td></td>
</tr>
<tr>
<td><strong>Wound Classification</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clean</td>
<td>101</td>
<td>6 (5.9)</td>
<td>0.046*</td>
</tr>
<tr>
<td>Clean Contaminated</td>
<td>112</td>
<td>10 (8.9)</td>
<td></td>
</tr>
<tr>
<td>Contaminated</td>
<td>143</td>
<td>14 (9.8)</td>
<td></td>
</tr>
<tr>
<td>Dirty</td>
<td>26</td>
<td>6 (23.1)</td>
<td></td>
</tr>
<tr>
<td><strong>Duration of Surgery (minutes)</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0-60</td>
<td>87</td>
<td>6 (6.9)</td>
<td>0.13*</td>
</tr>
<tr>
<td>61-120</td>
<td>189</td>
<td>16 (8.5)</td>
<td></td>
</tr>
<tr>
<td>121-180</td>
<td>64</td>
<td>9 (14.1)</td>
<td></td>
</tr>
<tr>
<td>≥180</td>
<td>42</td>
<td>5 (11.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Acuteness of Surgery</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arranged</td>
<td>225</td>
<td>17 (7.6)</td>
<td>0.14**</td>
</tr>
<tr>
<td>Urgent</td>
<td>94</td>
<td>9 (9.6)</td>
<td></td>
</tr>
<tr>
<td>Emergency</td>
<td>63</td>
<td>10 (15.9)</td>
<td></td>
</tr>
<tr>
<td><strong>Organ System operated on</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gynecology, Hernia</td>
<td>32</td>
<td>4 (12.5)</td>
<td>0.76***</td>
</tr>
<tr>
<td>Hepatobiliary</td>
<td>67</td>
<td>6 (8.9)</td>
<td></td>
</tr>
<tr>
<td>Gastrointestinal</td>
<td>221</td>
<td>23 (9.6)</td>
<td></td>
</tr>
<tr>
<td>Appendix</td>
<td>17</td>
<td>1 (5.9)</td>
<td></td>
</tr>
<tr>
<td>Vascular</td>
<td>45</td>
<td>2 (4.4)</td>
<td></td>
</tr>
</tbody>
</table>

* Cochran-Armitage test, ** Chi square test, *** Fischer’s exact test
For ordered categorical variables the significance of the trend between an increasing variable score and an increasing incidence of wound infection was tested using the Cochran-Armitage test. Logistic regression analysis of the different risk factors that contributed to the development of a wound infection was performed in SPSS version 11.0 (SPSS Institute, Cary, NC, USA). The patient’s age and the duration of surgery were entered as continuous variables. The ASA score, the wound classification and the acuteness of surgery were entered as categorical variables. Possible contributions from interactions between the independent variables were tested, and the presence of multi-collinearity was checked, by using variance inflation factors.

5.3 Results
5.3.1 Number of patients
From an initial 1013 patients recruited into the prophylactic antibiotic study, 483 patients had been assigned an ASA score. Of these there were 101 patients that had an appendicectomy but were not given metronidazole. These were withdrawn. The remaining 382 patients who were the basis for this study included 176 males and 206 females, with the mean age of 58.6 (95% confidence interval 56.9-60.4). Further details, including the type of operative procedures performed are included in Table 5.1.

5.3.2 Association between risk factors and the rate of wound infection
The wound infection rate was 9.4% (36 infections). The association between risk factors and the rate of wound infection is shown in Table 5.1. As the ASA classification of physical status score increased, there was a corresponding increase in wound infection (Cochrane Armitage test p=0.02). Similarly there was a relationship between the wound classification system and wound infection (Cochrane Armitage test, p=0.046). The trend between wound infection and other risk factors, including the duration of surgery and the acuteness of surgery was not significant.

5.3.3 Interaction between the ASA score and the classification of the wound
Combining the ASA classification of physical status and the wound classification helped to stratify the risk of wound infection. For example, there were no wound infections in the 23 cases with a clean wound and an ASA status score of I, and two wound infections in the four cases with a dirty wound and an ASA status score of IV. Table 5.2 illustrates that when the ASA score is increased from I/II to III/IV/V (Culver et al. 1991), the rate of wound infection
approximately doubles for clean, clean-contaminated, and contaminated wounds, and more than doubles in dirty wounds.

Table 5.2 The interaction between ASA classification of physical status and the wound classification system on the incidence of wound infection

<table>
<thead>
<tr>
<th>WOUND CLASSIFICATION</th>
<th>ASA I, II</th>
<th>ASA III, IV, V</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clean</td>
<td>2/49 (4.1)</td>
<td>4/52 (7.7)</td>
</tr>
<tr>
<td>Clean-contaminated</td>
<td>5/75 (6.7)</td>
<td>5/37 (13.5)</td>
</tr>
<tr>
<td>Contaminated</td>
<td>7/95 (7.4)</td>
<td>7/48 (14.6)</td>
</tr>
<tr>
<td>Dirty</td>
<td>2/17 (11.8)</td>
<td>4/9 (44.4)</td>
</tr>
</tbody>
</table>

Infection rate as a percentage is shown in parentheses

Table 5.3 Logistic regression analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>β coefficient</th>
<th>Wald statistic</th>
<th>p Value (for Wald)</th>
<th>Odds Ratio (95% CI)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ASA classification of physical status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>8.572</td>
<td>.036</td>
<td>1.0</td>
</tr>
<tr>
<td>ASA II</td>
<td>1.840</td>
<td>3.116</td>
<td>.078</td>
<td>6.295 (0.82-48.55)</td>
</tr>
<tr>
<td>ASA III</td>
<td>2.286</td>
<td>4.794</td>
<td>.029</td>
<td>9.839 (1.27-76.18)</td>
</tr>
<tr>
<td>ASA IV/V</td>
<td>2.937</td>
<td>6.924</td>
<td>.009</td>
<td>18.859 (2.12-168.11)</td>
</tr>
<tr>
<td>Wound classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall</td>
<td></td>
<td>5.807</td>
<td>.121</td>
<td>1.0</td>
</tr>
<tr>
<td>Clean</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clean-contaminated</td>
<td>.630</td>
<td>1.288</td>
<td>.256</td>
<td>1.878 (0.63-5.58)</td>
</tr>
<tr>
<td>Contaminated</td>
<td>.714</td>
<td>1.832</td>
<td>.176</td>
<td>2.042 (0.73-5.74)</td>
</tr>
<tr>
<td>Dirty</td>
<td>1.565</td>
<td>5.767</td>
<td>.016</td>
<td>4.782 (1.33-17.15)</td>
</tr>
</tbody>
</table>

5.3.4 Logistic regression analysis

When logistic regression analysis was applied to all these variables (Table 5.3), the only significant risk factor was the ASA classification of physical status. Overall the wound classification system was not significant, although a dirty wound, when entered as an
independent categorical variable was significantly associated with the development of a wound infection.

5.4 Discussion

5.4.1 ASA classification of physical status as a predictor or wound infection in other studies

In this study the ASA classification of physical status and the patient’s age (Hall and Hall 1996a) were used as indirect measurements of the host’s global health. The four wound categories were used as a measurement of the wound microbiology. The duration of surgery, acuteness of surgery, and the organ system being operated on partly assessed some of the surgical factors that may influence wound infection. Among all of these variables the ASA classification of physical status was the strongest predictor of wound infection, emphasising the importance of the host in the prevention of wound infection.

The ASA classification of physical status and its use as a risk factor for wound infection has been presented in section 4.4.1. In these studies the ASA status was an independent risk factor of wound infection with a similar predictive ability to other identified risk factors. In the NNIS study (Culver et al. 1991), the similar predictive ability of the wound classification, ASA status and duration of surgery (time) enabled these risk factors to be combined into a risk index, so that the rate of wound infection was 1.5% when there were no risk factors, 2.9% with one risk factor, 6.8% with two risk factors and 13% with three risk factors. In the study by Gabribaldi et al. the four independent risk factors on logistic regression analysis were again of a similar predictive ability, with odds ratios between 2.4 and 3.0 (Garibaldi et al. 1991).

5.4.2 The impact of optimal prophylactic antibiotic use on the predictive ability of the ASA classification of physical status

In this study the ASA classification of physical status improved from being ‘one of a number of risk factors,’ to becoming the strongest predictor of wound infection. The differences in the design of our study, and the SENIC and NNIS studies are presented in section 4.4.3 and section 5.1.3. As ceftriaxone and/or cefotaxime provide both an effective and prolonged antibiotic cover, this lessens the impact of bacterial contamination, lessening the predictive impact of both the wound classification system and the duration of surgery. In contrast the
impact of host factors on wound infection is not modified by prophylactic antibiotics, enabling these risk factors to become ‘relatively’ more important.

It is in the context of optimal prophylactic antibiotic use that wound classification and duration of surgery become less important, and the severity of the illness and other host factors, which are indirectly measured by the ASA classification of physical status, become ‘relatively’ more important in determining if a wound infection will occur.

5.4.3 Implication of these results

5.4.3.1 Improving the ASA classification of physical status
The direct implication of this study is that reducing the ASA classification of physical status is likely to reduce the rate of wound infection. There are two problems with this implication. Firstly although it is implied, we have not performed an interventional study to prove this. Secondly this is not a practical suggestion, as in most patients the ASA status is unlikely to be reducible.

5.4.3.2 Is there a role for giving additional prophylactic antibiotics to patients with a high ASA score?
Another approach is to recognise that patients with a high ASA status are at a higher risk of wound infection. Asking the question, “Can anything be done to decrease wound infection in high risk patients?” may lead to a new or slightly different perspective on the use of prophylactic antibiotics. For example, if patients with a higher ASA status score developed an infection after a smaller inoculation of bacteria into the wound, then the use of prophylactic antibiotics in these patients should have a lower threshold. For example, a patient undergoing a clean procedure with an ASA score of IV may benefit from prophylactic antibiotics, whereas a patient with an ASA score of I having the same procedure should not be given prophylactic antibiotics. Alternatively, there may a role for giving a larger dose of prophylactic antibiotic to a patient undergoing a procedure with a contaminated wound when they have an ASA score of IV in comparison to having an ASA score of I. The possibility of a higher ASA score being used as a criteria for increasing the use of prophylactic antibiotics is an area which has not been studied and would benefit from further research. While attempting to decrease the ASA score in many patients would be futile, altering the use of prophylactic antibiotics in-line with the ASA score of a patient is a simple and achievable intervention. Such an intervention would need to be properly assessed in a RCT as the use of additional...
prophylactic antibiotics without a proven benefit may increase the complications of antibiotic use.

5.4.3.3 ASA classification of physical status and improving the host’s defence against infection
Thirdly, if the ASA classification of physical status is taken as an indirect measurement of the host’s ability to prevent infection, then the implication of this finding is that efforts to reduce wound infection need to consider the importance of the host. In this study, in the context of optimal prophylaxis, host factors as measured by a higher ASA score were significantly associated with the development of wound infection. The implication of this finding is that the role of host factors in preventing wound infection should not be ignored.

Although the advantage of the ASA score is that it is an excellent general measurement of the global health of the patient, this is also a weakness, in that it does not identify or quantify any specific host problem that may be predisposing to the development of a wound infection. This raises an interesting question. Is the ASA status a surrogate measurement for an underlying problem (which can be specifically addressed), or is it the overall poor health of the patient which limits the effectiveness of the host to mount an optimal immune response (through a series of mechanisms), and so predisposes to wound infection? How can this finding help us to optimise the host’s defences against bacterial infection?

5.5 The relationship between wound infection and different components of the ASA score
To explore the implications of this finding it is helpful to look at the different components of the ASA score: the functional status, co-existing medical diseases and physiological stability of the patient.

5.5.1 Functional status and wound infection
A decreased level of functional status has been identified to be a risk factor in a number of studies looking at nosocomial infection (Mylotte et al. 2000; Pittet et al. 1999) and wound infection (Giles et al. 2010; Veeravagu et al. 2009).
5.5.2 Medical comorbidities, nutritional intervention and wound infection

There is also a clear association between the number of medical co-morbidities and an increased risk of wound infection. For example, in terms of general co-morbidity there is an association between having three or more diagnoses at the time of discharge from hospital (Hayley et al. 1985) and developing a wound infection. In terms of vascular co-morbidity, there is an association between the number of cardiac medications a patient is taking at the time of admission to hospital (Kennedy and van Rij 1998) and wound infection.

There are also many specific co-morbidities that are associated with an increase in wound infection. Some of the most commonly identified co-morbidities associated with wound infection are obesity (Giles et al. 2010; Hedrick et al. 2007; Cheadle 2006), smoking (Giles et al. 2010; Veeravagu et al. 2009; Sørensen et al. 2010) diabetes [we discuss these three co-morbidities later] and malnutrition.

The importance of good preoperative and postoperative nutrition for the prevention of infective complications in malnourished patients is well established (Hill 1987; Beattie et al. 2000). However, a more clinically relevant issue in 2011 is whether there is a role for nutritional supplementation in people who are well nourished. A recent study suggested that preoperative carbohydrate loading probably makes no difference, especially after laparoscopic surgery (Mathur et al. 2010). The other nutritional question is whether boosting the immune system by the use of immunonutrition can prevent infective complications. Immunonutrition includes glutamine, arginine, omega-3 polyunsaturated fatty acids and ribonucleic acids. These can boost the immune system by a number of mechanisms. Some studies have shown a benefit in reducing all infections (Braga et al. 2002; Okamoto et al. 2009; Zheng et al. 2007; Oguz et al. 2007) as well as reducing wound infection (Oguz et al. 2007). Other studies demonstrated no reduction in infective complications after surgery (Helminen et al. 2007; Yeh 2008; Klek et al. 2008). The role of using immunonutrition in healthy patients has not yet been resolved.

5.5.3 Acute physiology and wound infection

The focus of recent work looking at preventing wound infection has been on optimising the ability of the host to prevent infection by providing an optimal environment to fight infection. These studies are focusing on the physiology of the patient and the homeostasis of the wound. The three main physiological parameters studied have been oxygen (with the option of
providing supplemental oxygen), temperature (with active warming to achieve intraoperative normothermia) and the optimal control of blood glucose.

Evidence that surgical wound infections can be directly related to the oxygenation of the wound (Kabon et al. 2004) has led to a number of studies assessing the impact of supplemental oxygen on wound infection. Although hypoxia clearly contributes to infection, does increasing the oxygen tension in the wound to higher levels, by giving oxygen supplementation, help to prevent infection? The initial studies, which used 80% inspired oxygen in patients undergoing colorectal surgery were very positive (Belda et al. 2005). However a wide range of results have since been published. These included some studies that showed no benefit (Gardella et al. 2008; Mayzler et al. 2005, Meyhoff et al. 2009) and one study that showed a worse outcome (Pryor et al. 2004). Two meta-analyses have also reached different conclusions (Al-Niaimi and Safdar 2009; Qadan et al. 2009). The use of supplemental oxygen requires further clarification.

Temperature control has also been studied. Perioperative hypothermia is common and can adversely affect a range of clinical outcomes due to its effect on a range of homeostatic functions. At least one good study shows that active warming can maintain normothermia and that this results in a significant reduction in wound infection (Kurz et al. 1996). A more recent review on this subject concludes that although the reduction in wound infection, blood loss and perioperative pain with warming is promising, that more evidence from good-quality prospective randomised controlled trials is needed to evaluate the role of warming in improving overall morbidity and mortality (Kumar et al. 2005).

A number of studies have shown an association between a high blood glucose level and wound infection (Giles et al. 2010). The threshold at which an elevated glucose contributes to wound infection varies in different studies, but lies within the rage of 8mmol/l to 11mmol/l (Ata et al. 2010; Park et al. 2009; McConnell et al. 2009). However other studies have looked specifically at interventions to achieve optimal or ‘normal’ glucose control. For example in the context of liver surgery it has also been shown that intensive insulin therapy results in a better control of glucose and a reduction in wound infection when compared to ‘normal’ methods of glucose control (Okabayashi et al. 2009).
There is some evidence that a reduced oxygen tension in the wound, below an ideal threshold for fighting infection, may prove to be a common pathway by which a number of host risk factors for wound infection work, for example, obesity, smoking and hypothermia. In obesity the fat tissue mass is increased without a concomitant increase in blood flow per cell, resulting in a relative hypoperfusion with decreased tissue oxygenation. It has been demonstrated that even when supplemental oxygen is given, that the oxygen tension in the wound of obese patients is reduced to levels that are associated with a substantial increase in infection risk (Kabon et al. 2004). Smoking decreases oxygen tension in the wound secondary to vasospasm (Sørensen et al. 2010). It also appears that the ability of warming (maintaining normothermia) to prevent wound infection may be partly mediated through increased cutaneous blood flow and an improved oxygen tension at tissue level (Leaper 2006).

Although the association of a high ASA score with wound infection does not ‘answer the question’ of how we prevent wound infection, a useful application of this finding is that it points in the direction of improving host factors, which includes interventions that aim to improve the physiology of the patient and the homeostasis of the wound.

5.5.4 Merits of a combined approach

Although the most effective individual intervention to improve the homeostasis of the wound is unclear, an attempt can be made to improve as many of the ‘physiological’ risk factors as possible. This approach has been used in a number of protocol based studies. The main interventions in these ‘package of care’ studies include a more strict adherence to antibiotic protocols, better control of blood glucose and interventions to maintain normothermia (Liau et al. 2010; Hedrick et al. 2007; Forbes et al. 2008; Dellinger et al. 2005). Some studies have also included the use of supplemental oxygen (Dellinger et al. 2005), the use of clippers instead of shavers for the removal of hair (Liau et al. 2010; Dellinger et al. 2005), and the placement of deep subcutaneous drains in obese patients (Hedrick et al. 2007). Many of these studies have shown a significant reduction in wound infection (Liau et al. 2010; Hedrick, Heckman et al 2007; Dellinger, Hausmann et al 2005). One disadvantage to this approach is that it is not possible to demonstrate what proportion of the beneficial effect is due to a greater awareness of infection by medical staff, the better use of antibiotics, to steps taken to improve the host’s ability to fight infection, or to a more specific intervention.
5.6 Limitations of this study

With respect to the collection of data there are two reasons why there should be caution in interpreting the results of this study. The first is that the primary purpose of the database was to assess the impact of prophylactic antibiotics on wound infection, not to assess the impact of the ASA score on wound infection. The second is that this is a retrospective study. Both the design of the study and the collection of ASA status data would have been more focused if the impact of the ASA classification of physical status had been assessed in a prospective study.

Another limitation is the choice of the ASA status score to represent the general health of the host. Firstly the observation that the ASA status score is associated with an increase in wound infection has no immediate practical application, as in most patients the ASA status is unlikely to be reducible. Secondly the ASA status score is being used as a general indicator for the importance of the host in preventing wound infection. It is therefore ‘vague’ as it does not pinpoint any specific problem or immediate intervention that would help to decrease wound infection.

5.7 Conclusion

This retrospective study rejected the null hypothesis, and confirmed the alternative hypothesis, that in the context of optimal antibiotic prophylaxis the ASA classification of physical status becomes the strongest predictor of wound infection. Although the ASA score does not identify any specific problem that may predispose a patient to the development of a wound infection, it does highlight the importance of the host and the global health of the patient in preventing wound infection, and therefore emphasises that host factors need to be assessed more rigorously when trying to reduce the incidence of wound infection. This observation has been reinforced by recent interventional studies which have concentrated on reducing wound infection by optimising different aspects of the physiology of the patient and the homeostasis of the wound.

5.8 Further questions for study

1) The ASA classification of physical status is such a general measurement that it does not identify a specific problem. A reasonable question is therefore ‘are there any specific measures which can be undertaken to enhance the host’s immune response in patients with a high ASA score’?
2) Will steps to improve glucose control, wound oxygen tension and ‘normothermia’ be equally effective in all patients, or will they be more effective in reducing wound infection in patients with a higher ASA score?

3) Will increasing the prophylactic cover for patients with a higher ASA score result in a reduction in wound infection? This question could be explored by using animal models and also by interventional clinical studies which stratify patients according to the level of wound contamination and their ASA status.

5.9 Publication
CHAPTER 6

6. A meta-analysis of randomised controlled trials assessing the prophylactic use of Ceftriaxone. A study of wound, chest and urinary infections.

6.1 Introduction

6.1.1 Should ceftriaxone be considered as a first-line prophylactic antibiotic for the prevention of postoperative infections?

Ceftriaxone is a very effective prophylactic antibiotic which has been widely used as prophylaxis against wound infection following surgery. Advantages include its excellent cover against a wide range of gram positive and gram negative organisms, a half-life of eight hours and a tissue penetration index (Mazzei and Periti 1989) of 92%. This results in a sustained concentration of antibiotic above the minimal inhibitory concentration (MIC90) of most bacteria for 24 hours (Shinagawa 1989). However for an antibiotic to be preferred above many other reasonable options, it needs to be shown to have a definite advantage.

Ceftriaxone was more widely used as a first line prophylactic antibiotic in the 1990’s than now. Some authors have claimed that ceftriaxone is no more effective than first or second generation cephalosporins in preventing wound infection (Kreter and Woods 1992) and have recommended that a less expensive antibiotic should be used. There may however be a case for stratifying antibiotics so that different prophylactic antibiotics are recommended for low risk and high risk procedures. Others have suggested ceftriaxone should be kept in reserve for the treatment of serious infections. There are also important concerns about patterns of antibiotic resistance and of Clostridium difficile infection (see section 6.7 and Appendix A).

This chapter reviews the available evidence about the relative efficacy of ceftriaxone prophylaxis is preventing infection in comparison to other antibiotics.

6.1.2 A study of wound, chest and urinary infections.

In this chapter we assess wound, chest and urinary infections. The evidence for using prophylactic antibiotics to treat wound infection is overwhelming, and is not represented here. Another potential benefit of prophylactic antibiotics is in the prevention of infections at other sites, such as chest infection and urinary tract infection (UTI). This had previously been noted in some studies using ceftriaxone prophylaxis (De la Hunt et al. 1985; Lumley et al. 1992; Morris 1993). For pneumonia and UTI the time of contamination may extend beyond the time
of surgery (see section 1.5). Ceftriaxone, with its longer duration of action, may therefore be a suitable antibiotic for preventing these infections. If this can be confirmed in a meta-analysis it would strengthen the case for using ceftriaxone as a first line prophylactic antibiotic.

6.1.3 Why a meta-analysis now?
There were two reasons for performing a meta-analysis now:
1) A meta-analysis is a cost effective way of assessing uncommon but potentially serious events, such as pneumonia. To perform a RCT with pneumonia as a primary endpoint would require a large study. In our RCT ceftriaxone reduced chest and urinary infection as a combined endpoint (Woodfield et al. 2003), but the low incidence of these infections (both less than 5%) meant a larger study was needed to assess the ability of ceftriaxone to reduce both of these infections (section 2.4.3).

2) Many RCTs looking at the prophylactic performance of ceftriaxone have been performed, so there should be sufficient data to clearly identify if the performance of ceftriaxone is ‘good enough’ for it to be used as a first line prophylactic antibiotic, both for wound infection, and also for chest and/or urinary infection.

6.1.4 Questions to be answered by the meta-analysis
The meta-analysis is designed to compare the overall efficacy of ceftriaxone against other prophylactic antibiotics in the prevention of wound infection, chest infection and urinary infection. This will also enable a comparison of antibiotic performance across a range of low and high risk surgical procedures.

The null hypothesis is that ceftriaxone is no more effective than other appropriately selected antibiotics in decreasing wound infection, chest infection or urinary infection. The alternative hypothesis is that ceftriaxone is more effective than other appropriately selected antibiotics in decreasing wound infection, chest infection and urinary infection.

6.2 Methods
6.2.1 Search strategy
Medline, Embase, the Cochrane controlled trials register and the Cochrane review register were reviewed by two reviewers, the author and a clinical researcher. Additional relevant cited references were identified from the retrieved papers and review articles. Proceedings from meetings were also studied, Roche was contacted (but never replied!), and email was
used to contact authors to clarify points or to help locate additional references. The duration of the search was from 1983 (first published paper on ceftriaxone) to January 2005. No language or age restrictions were used. Abstracts as well as published articles were assessed. Search terms used included ceftriaxone, randomised controlled trial and surg$ or operat$. The search was then repeated using perioperat$.

6.2.2 Inclusion and exclusion criteria
All randomised trials assessing the operative prophylactic use of ceftriaxone were assessed. The inclusion criteria included both elective and emergency surgery, single or multiple doses of antibiotics and more than one antibiotic being used. Papers comparing the use of ceftriaxone with placebo were also included (to help assess if ceftriaxone was able to reduce chest or urinary infection) but were analysed separately. Studies were excluded if they were non randomised prospective studies or retrospective studies, if they were assessing pharmacokinetic endpoints only, if there were major methodological problems, if the endpoint results were not clearly stated or if they were documenting the treatment of infection rather than prophylaxis. In terms of randomisation open studies were excluded. If the nature of randomisation was not clearly stated the paper was included in the initial meta-analysis. Studies included in more than one publication were only included once. Studies using additional anaerobic cover, such as metronidazole were included if this was used equivalently in both arms of the study, or if it was used in the antibiotic regimen being compared against ceftriaxone. Studies with no documented infections were not included.

6.2.3 Analysis of papers
Data extracted from the papers included the date, method of randomisation, method of blinding, details of antibiotic administration, type of surgery, duration of follow-up, number of patients in each arm, number lost to follow-up, definition of endpoints and the number of endpoint infections. Endpoint infections were compared to the 2004 CDC definition of infections (Horan and Gaynes 2004). The validity of each study was assessed using the Jadad scoring system (Jadad et al. 1996), which assesses the quality of randomisation, blinding and the documentation of patient withdrawal. Data was independently reviewed by two people, by the author and by Dr Nagy Beshay. Final inclusion of articles was determined by consensus.
6.2.4 Analysis of the data

A separate meta-analysis comparing ceftriaxone against other prophylactic antibiotics was performed for wound infection, for urinary tract infection and for pneumonia. A further meta-analysis compared the effectiveness of Ceftriaxone against placebo studies.

The meta-analysis was performed in Comprehensive Meta-analysis v2 (Biostat, Englewood NJ, 2005). The pooled effect/summary statistics were expressed as an odds ratio with 95% confidence intervals using the fixed-effects model. A funnel plot was used to check for publisher bias and to assess heterogeneity. Heterogeneity was also assessed by using Cochran’s Q statistic and I².

![Flow diagram](image-url)  
**Figure 6.1 Flow diagram for identification and inclusion of papers**
In each meta-analysis the following steps were followed. Outlying studies were identified using a funnel plot and were removed from further analysis. The summary statistic for the remaining studies was then generated. A series of sensitivity analyses were then performed. These included removing each individual study, assessing only the studies where it was clearly stated that double blinding had occurred, assessing only the studies that used the CDC definition of infections, and assessing only those studies with a Jadad score of 4 or 5. A sub-analysis was carried out comparing the performance of ceftriaxone against different groups of antibiotics and assessing the performance of ceftriaxone across different types of surgical procedures. Finally an analysis combining studies with double blinding and a CDC definition of infection was performed.

6.3 Results

6.3.1 Results for search strategy

The number of abstracts identified using Medline, Embase and the Cochrane registers was 364. Using the exclusion criteria, 116 abstracts were identified and the papers retrieved (Figure 6.1). The references of these papers and of relevant review articles were then searched with an additional 82 papers being identified. A further 33 articles were identified from the proceedings of scientific meetings, resulting in a total of 231 papers. The exclusion criteria were reapplied to the 231 retrieved papers, following which 90 studies were selected to be included in the meta-analysis (Childs et al. 1983; Ronconi et al. 1983; Beam et al. 1984; Finkelstein et al. 1984; Kellum et al. 1984; Periti et al. 1984a; Periti et al. 1984b; de la Hunt et al. 1985; Geroulanos et al. 1985; Hemsell et al. 1985; Luscher and Bruhwiler 1985; Petropoulos et al. 1985; Tripi et al. 1986; Weaver et al. 1986; Blatzas et al. 1987; Burdon and Keighley 1987; Cucchiaro et al. 1987; Harmes and Schumacker 1987; Harnoss et al. 1987; Parker et al. 1987; Stoll 1987; Winter et al. 1987; Wohlfart and Siedeh 1987; Alley 1988; Ferraz et al. 1988; Germiniani et al. 1988; Kiff et al. 1988; Lang et al. 1988; Petersen et al. 1988; Rausis and Amato 1988; Tristaino et al. 1988; Badel and Schmuziger 1989; Bekse et al. 1989; El-Mufti et al. 1989a; El-Mufti et al. 1989b; Franceschini et al. 1989; Garcia and Pedroso 1989; Garotta et al. 1989; Geroulanos et al. 1989; Kujath 1989; Kunz et al. 1989; Periti and Jacchia 1989; Karachalios et al. 1990; Periti and Tonelli 1990; Stiver et al. 1990; Thomas 1990; Thorsteinsson et al. 1990; Cardamakis et al. 1991; Hall et al. 1991; Hjortrup et al. 1991; Luke et al. 1991; Rotman et al. 1991; Lumley et al. 1992; Mamsen et al. 1992; Papageorgiou et al. 1992; Botto et al. 1993; Fejgin, et al. 1993; Fingerhut and Hay 1993; Hall et al. 1993; Jimenez-Cruz and Broseta 1993; Jimenez-Cruz et al. 1993; Kracht 1993;

Of these 44 were sourced from Medline, an additional 8 from Embase, 10 from the Cochrane registers, 10 from scientific meetings and 18 from identified references. Fifty non English articles, including French, German, Italian, Spanish, Russian and Chinese were retrieved. Of these 16 were included in the analysis. A number of these fifty identified studies had also been published in English (they were published separately in more than one language).

6.3.2 Meta-analysis for wound infection

The funnel plot identified three studies (Weaver et al. 1986; Matikainen and Hiltunen 1993; Thomas et al. 1999) that were either on or outside the 95% confidence interval, all in favour of ceftriaxone. These were excluded from further analysis (Figure 6.2, Figure 6.3).
Figure 6.2 Funnel plot for all wound infection before exclusion of three studies on or outside the 95% confidence interval line

Note that the two studies that were on the 95<sup>th</sup> percentile line favouring an antibiotic being compared against ceftriaxone are now inside the 95<sup>th</sup> percentile line. These were therefore left in the meta-analysis. Another option would have been to omit all five studies, but when there were two options the option which did not favour ceftriaxone was selected (on the assumption that if its performance was indeed better this should still show at the time of the final analysis).

Figure 6.3 Funnel plot after removing three studies on or outside the 95% percentile line

The remaining 61 studies were evenly distributed throughout the funnel plot. The meta-analysis showed a difference in favour of ceftriaxone (Figure 6.4)
Figure 6.4 Forest plot with odds ratio and 95% confidence intervals for all studies assessing wound infection

Table 6.1 Summary of meta-analysis of studies assessing wound infection

<table>
<thead>
<tr>
<th>Category</th>
<th>No of studies</th>
<th>WI</th>
<th>No WI</th>
<th>WI</th>
<th>No WI</th>
<th>OR</th>
<th>95% CI</th>
<th>p Value</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Studies</td>
<td>61</td>
<td>314</td>
<td>1011</td>
<td>478</td>
<td>1032</td>
<td>0.68</td>
<td>0.53-0.70</td>
<td>&lt;0.001</td>
<td>0.93</td>
</tr>
<tr>
<td>Single dose of Ceftriaxone compared to single or multiple doses of other antibiotics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>18</td>
<td>119</td>
<td>2833</td>
<td>199</td>
<td>3978</td>
<td>0.62</td>
<td>0.48-0.79</td>
<td>&lt;0.001</td>
<td>0.79</td>
</tr>
<tr>
<td>Multiple</td>
<td>41</td>
<td>187</td>
<td>7075</td>
<td>265</td>
<td>7140</td>
<td>0.73</td>
<td>0.60-0.89</td>
<td>0.002</td>
<td>0.87</td>
</tr>
<tr>
<td>Method of blinding: Not stated or poorly outlined against double blinding</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stated</td>
<td>21</td>
<td>174</td>
<td>3979</td>
<td>273</td>
<td>4142</td>
<td>0.66</td>
<td>0.54-0.81</td>
<td>&lt;0.001</td>
<td>0.50</td>
</tr>
<tr>
<td>Not stated</td>
<td>40</td>
<td>140</td>
<td>6132</td>
<td>205</td>
<td>6181</td>
<td>0.70</td>
<td>0.56-0.89</td>
<td>0.003</td>
<td>0.96</td>
</tr>
<tr>
<td>Definition of wound infection: Not stated or poorly defined against CDC definition</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defined</td>
<td>40</td>
<td>267</td>
<td>7994</td>
<td>383</td>
<td>8163</td>
<td>0.72</td>
<td>0.69-0.85</td>
<td>&lt;0.001</td>
<td>0.78</td>
</tr>
<tr>
<td>Not defined</td>
<td>21</td>
<td>47</td>
<td>2117</td>
<td>95</td>
<td>2160</td>
<td>0.53</td>
<td>0.37-0.76</td>
<td>0.001</td>
<td>0.95</td>
</tr>
<tr>
<td>Comparison against different types of antibiotics</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aminoglycosides</td>
<td>3</td>
<td>15</td>
<td>250</td>
<td>37</td>
<td>266</td>
<td>0.39</td>
<td>0.21-0.74</td>
<td>0.004</td>
<td>0.61</td>
</tr>
<tr>
<td>1st &amp; 2nd generation cephalosporins</td>
<td>28</td>
<td>136</td>
<td>5426</td>
<td>187</td>
<td>5257</td>
<td>0.70</td>
<td>0.55-0.89</td>
<td>0.003</td>
<td>0.64</td>
</tr>
<tr>
<td>3rd generation cephalosporins</td>
<td>13</td>
<td>42</td>
<td>1230</td>
<td>97</td>
<td>1670</td>
<td>0.67</td>
<td>0.48-0.94</td>
<td>0.02</td>
<td>0.75</td>
</tr>
<tr>
<td>Penicillins</td>
<td>9</td>
<td>53</td>
<td>1400</td>
<td>64</td>
<td>1405</td>
<td>0.84</td>
<td>0.58-1.23</td>
<td>0.37</td>
<td>0.88</td>
</tr>
<tr>
<td>Assessment of performance for different types of surgical procedures</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal – gastrointestinal</td>
<td>32</td>
<td>184</td>
<td>5040</td>
<td>321</td>
<td>5329</td>
<td>0.61</td>
<td>0.51-0.75</td>
<td>&lt;0.001</td>
<td>0.89</td>
</tr>
<tr>
<td>Pelvic</td>
<td>10</td>
<td>35</td>
<td>1551</td>
<td>49</td>
<td>1495</td>
<td>0.65</td>
<td>0.41-1.02</td>
<td>0.06</td>
<td>0.50</td>
</tr>
<tr>
<td>Clean</td>
<td>15</td>
<td>78</td>
<td>3300</td>
<td>88</td>
<td>3294</td>
<td>0.88</td>
<td>0.64-1.12</td>
<td>0.44</td>
<td>0.78</td>
</tr>
</tbody>
</table>

Heterogeneity is the p value for Cochran’s Q test
WI = wound infection
OR = odds ratio
Abdominal – gastrointestinal includes studies looking at ‘all abdominal surgery’, appendicectomy, biliary surgery, upper gastrointestinal surgery and colorectal surgery.
Pelvic – Obstetric, gynaecological, open urology
Clean – cardiac surgery, neurosurgery, orthopaedic surgery
Two studies are omitted from single dose of ceftriaxone category as more than one dose of ceftriaxone was administered.
Pelvic surgery (Periti et al. 1984a; Periti et al. 1984b; Hemsell et al. 1985; Kunz et al. 1989; Cardamakis et al. 1991; Papageorgiou et al. 1992; Fejgin et al. 1993; Von Mandach et al. 1993; Lehapa et al. 1999) showed a non significant trend in favour of ceftriaxone (p=0.06).

There was no difference in antibiotic performance for clean operations, which included cardiac surgery (Beam et al. 1984; Geroulanos et al. 1985; Badel and Schmuziger 1989; Geroulanos et al. 1989; Hall et al. 1993; Sisto et al. 1994; Ilic et al. 1997; Salminen et al. 1999; Kriaras et al. 2000), neurosurgery (Zhu et al. 2001) and orthopaedic operations (Winter et al. 1987; Periti and Jacchia 1989; Karachalios et al. 1990).

Randomised studies with double blinding and a CDC definition of infection (Figure 6.5) were then analysed (Beam et al. 1984; Hemsell et al. 1985; Burdon and Keighley 1987; Lang et al. 1988; Kunz et al. 1989; Hall et al. 1991; Luke et al. 1991; Rotman et al. 1991; Lumley et al. 1992; Hall et al. 1993; Morris 1993; Morris 1994; Anderson et al. 1996; Zhu et al. 2001; Woodfield et al. 2003).

<table>
<thead>
<tr>
<th>Study name</th>
<th>Odds ratio and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson [13]</td>
<td></td>
</tr>
<tr>
<td>Beam [69]</td>
<td></td>
</tr>
<tr>
<td>Burdon [16]</td>
<td></td>
</tr>
<tr>
<td>Hall [29]</td>
<td></td>
</tr>
<tr>
<td>Hall [28]</td>
<td></td>
</tr>
<tr>
<td>Hemsell [31]</td>
<td></td>
</tr>
<tr>
<td>Kunz [39]</td>
<td></td>
</tr>
<tr>
<td>Lang [41]</td>
<td></td>
</tr>
<tr>
<td>Luke [43]</td>
<td></td>
</tr>
<tr>
<td>Lumley [6]</td>
<td></td>
</tr>
<tr>
<td>Morris [7]</td>
<td></td>
</tr>
<tr>
<td>Morris [45]</td>
<td></td>
</tr>
<tr>
<td>Rotman [55]</td>
<td></td>
</tr>
<tr>
<td>Woodfield [9]</td>
<td></td>
</tr>
<tr>
<td>Zhu XL [67]</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6.5 Forest plot for studies with double blinding and CDC definition of infection assessing incisional SSI
This also showed an advantage for ceftriaxone \((n=15, \text{OR} 0.68, \text{CI} 0.51-0.84, p<0.001)\), which was predominantly in the abdominal-gastrointestinal group of operations (Burdon and Keighley 1987; Lang et al. 1988; Hall et al. 1991; Luke et al. 1991; Rotman et al. 1991; Lumley et al. 1992; Morris 1993; Morris 1994; Anderson et al. 1996; Woodfield et al. 2003).

### 6.3.3 Meta-analysis for urinary tract infection


The funnel plot identified three studies (Cucchiaro et al. 1987; Parker et al. 1987; Periti et al. 1984b) that were either on or outside the 95% confidence interval, two (Cucchiaro et al. 1987; Parker et al. 1987) in favour of ceftriaxone. These three studies were excluded from further analysis.

Six studies (Periti et al. 1984a; Tripi et al. 1986; Thorsteinsson et al. 1990; Botto et al. 1993; Jimenez-Cruz and Broseta 1993; Viitanen et al. 1993) assessed the ability of antibiotics to reduce UTI after endoscopic surgery. These showed a difference in favour of ceftriaxone \((\text{OR} 0.77, \text{CI} 0.51-0.99, p=0.04)\).

The remaining 35 studies assessed UTI after a range of surgical procedures. These studies were evenly distributed throughout the funnel plot (Figure 6.6). The meta-analysis showed a difference in favour of ceftriaxone (Figure 6.7).
This was unchanged when the sensitivity analysis (Table 6.2) looked at studies with double blinding (Beam et al. 1984; Anderson et al. 1996; Kellum et al. 1984; Lang et al. 1988; Garotta et al. 1989; Kunz et al. 1989; Cardamakis et al. 1991; Luke et al. 1991; Rotman et al. 1991; Lumley et al. 1992; Papageorgiou et al. 1992; Morris 1993; Nyam et al. 1995; Zhu et al. 2001; Woodfield et al. 2003) and CDC definitions of infection (Figure 6.8) (Beam, et al. 1984; Harmes and Schumacker 1987; Lang et al. 1988; Beksac et al. 1989; Periti and Tonelli 1990; Stiver et al. 1990; Luke et al. 1991; Lumley et al. 1992; Kriaras et al. 2000; Woodfield et al. 2003).

The result also remained unaltered when looking at studies with a Jadad score of four or more (n=8, OR 0.58, CI 0.38-0.90, p=0.014) (Beam et al. 1984; Lang et al. 1988; Lumley et al. 1992; Nyam et al. 1995; Anderson et al. 1996; Zhu et al. 2001; Woodfield et al. 2003). There was minimal heterogeneity.

Sub-analysis confirmed a difference in antibiotic performance across most groups of antibiotics. The difference in antibiotic performance was significant following general abdominal-gastrointestinal operations, including studies looking at a range of abdominal surgery (Luke et al. 1991; Kracht 1993; Schweizer et al. 1994; Woodfield et al. 2003), and colorectal surgery (Franceschini et al. 1989; Garcia and Pedroso 1989; Thomas 1990; Lumley et al. 1992; Morris 1993; Nyam et al. 1995). The difference was not present following upper gastrointestinal (Blatzas et al. 1987; Periti and Tonelli 1990; Rotman et al. 1991; Anderson et al. 1996) or biliary surgery (Kellum et al. 1984; de la Hunt et al. 1985; Harnoss et al. 1987; El-Mufti et al. 1989a).
Figure 6.7 Forest plot with odds ratio and 95% confidence intervals for all studies assessing urinary tract infection (excluding endoscopic urology)

The differences were most significant after colorectal surgery (n=6, OR 0.30, CI 0.16-0.58, p<0.001) and obstetric and gynaecological (O&G) surgery (n=9, OR 0.53, CI 0.34-0.69) (Harmes and Schumacker 1987; Beksac et al. 1989; Kunz et al. 1989; Stiver et al. 1990;

Table 6.2 Summary of meta-analysis of studies assessing urinary tract infection

<table>
<thead>
<tr>
<th>Category</th>
<th>Ceftriaxone</th>
<th>Other Antibiotics</th>
<th>Odds ratio summary statistic</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No of studies</td>
<td>UI</td>
<td>No UI</td>
<td>UI</td>
</tr>
<tr>
<td>All Studies</td>
<td>35</td>
<td>188</td>
<td>6518</td>
<td>338</td>
</tr>
<tr>
<td>Single dose of Ceftriaxone compared to single or multiple doses of other antibiotics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>17</td>
<td>76</td>
<td>2445</td>
<td>136</td>
</tr>
<tr>
<td>Multiple</td>
<td>17</td>
<td>111</td>
<td>3922</td>
<td>202</td>
</tr>
<tr>
<td>Method of blinding: Not stated or poorly outlined against double blinding</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stated</td>
<td>15</td>
<td>80</td>
<td>2748</td>
<td>140</td>
</tr>
<tr>
<td>Not stated</td>
<td>20</td>
<td>108</td>
<td>3770</td>
<td>198</td>
</tr>
<tr>
<td>Definition of urinary infection: Not stated or poorly defined against CDC definition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defined</td>
<td>10</td>
<td>51</td>
<td>2200</td>
<td>83</td>
</tr>
<tr>
<td>Not defined</td>
<td>25</td>
<td>137</td>
<td>4318</td>
<td>225</td>
</tr>
<tr>
<td>Comparison against different types of antibiotics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st &amp; 2nd generation cephalosporins</td>
<td>17</td>
<td>107</td>
<td>3678</td>
<td>199</td>
</tr>
<tr>
<td>3rd generation cephalosporins</td>
<td>10</td>
<td>66</td>
<td>1519</td>
<td>93</td>
</tr>
<tr>
<td>Penicillins</td>
<td>7</td>
<td>16</td>
<td>753</td>
<td>25</td>
</tr>
<tr>
<td>Assessment of performance for different types of surgical procedures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal – gastrointestinal</td>
<td>19</td>
<td>80</td>
<td>3640</td>
<td>161</td>
</tr>
<tr>
<td>Pelvic</td>
<td>9</td>
<td>92</td>
<td>1225</td>
<td>151</td>
</tr>
<tr>
<td>Clean</td>
<td>6</td>
<td>13</td>
<td>1572</td>
<td>19</td>
</tr>
</tbody>
</table>

This does not include urinary infection after endoscopic urological surgery

UI = urinary tract infection

There was no difference in antibiotic performance for clean operations, which included mainly cardiac surgery (Beam et al. 1984; Geroulanos et al. 1989; Kriaras et al. 2000) and orthopaedic operations (Winter et al. 1987; Periti and Jacchia 1989).
Randomised studies with double blinding and a CDC definition of infection were then analysed. As many studies did not have a CDC defined symptomatic urinary tract infection there were only five studies in this category (Beam et al. 1984; Lang et al. 1988; Rotman et al. 1991; Lumley et al. 1992; Woodfield et al. 2003). A difference in favour of ceftriaxone was still present (OR 0.48, CI 0.32-0.71, p<0.001).

6.3.4 Meta-analysis for pneumonia

Schweizer et al. 1994; Sisto et al. 1994; Nyam et al. 1995; Anderson et al. 1996; Ilic et al. 1997; Salminen et al. 1999; Kriaras et al. 2000; Zhu et al. 2001; Woodfield et al. 2003). The funnel plot (Figure 6.9) identified one study (de la Hunt et al. 1985) that was outside the 95% confidence interval. This was in favour of ceftriaxone, and was excluded from further analysis. The remaining 37 studies were evenly distributed throughout the funnel plot. The meta-analysis (Figure 6.10) showed a difference in favour of ceftriaxone.

This was unchanged when sensitivity analysis was performed looking at studies with double blinding (Kellum et al. 1984; Ferraz et al. 1988; Lang et al. 1988; Garotta et al. 1989; Kunz et al. 1989; Luke et al. 1991; Rotman et al. 1991; Lumley et al. 1992; Morris 1993; Morris 1994; Nyam et al. 1995; Anderson et al. 1996; Zhu et al. 2001; Woodfield et al. 2003), CDC definitions (including those studies which made a genuine attempt to correlate clinical, laboratory and radiological parameters) of infection (Geroulanos et al. 1985; Geroulanos et al. 1989; Anderson et al. 1996; Kriaras et al. 2000; Zhu et al. 2001; Woodfield et al. 2003) and a Jadad score of four or more (Figure 6.11, n=7, OR 0.60, CI 0.42-0.87, p=0.007) (Lang et al. 1988; Lumley et al. 1992; Nyam et al. 1995; Anderson et al. 1996; Zhu et al. 2001; Woodfield et al. 2003):
Figure 6.10 Forest plot with odds ratio and 95% confidence intervals for all studies assessing pneumonia
<table>
<thead>
<tr>
<th>Study name</th>
<th>Odds ratio and 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson [13]</td>
<td></td>
</tr>
<tr>
<td>Kellum [77]</td>
<td></td>
</tr>
<tr>
<td>Lang [41]</td>
<td></td>
</tr>
<tr>
<td>Lumley [6]</td>
<td></td>
</tr>
<tr>
<td>Nyam [46]</td>
<td></td>
</tr>
<tr>
<td>Woodfield [9]</td>
<td></td>
</tr>
<tr>
<td>Zhu [67]</td>
<td></td>
</tr>
</tbody>
</table>

Figure 6.11 Forest plot for studies with a Jadad score of 4 or 5 assessing pneumonia

Throughout the sensitivity analysis (Table 6.3) the p value for heterogeneity was usually greater than 0.40. Sub-analysis demonstrated that the difference in antibiotic performance was greatest when using first or second generation cephalosporins. The difference in antibiotic performance was significant in the general abdominal-gastrointestinal group of operations (Anderson et al. 1996; Kellum et al. 1984; Blatzas et al. 1987; Cucchiaro et al. 1987; Ferraz et al. 1988; Lang et al. 1988; Tristaino et al. 1988; El-Mufti et al. 1989a; Franceschini et al. 1989; Garcia and Pedroso 1989; Periti and Tonelli 1990; Thomas 1990; Hjortrup et al. 1991; Luke et al. 1991; Rotman et al. 1991; Lumley et al. 1992; Kracht 1993; Morris 1993; Morris 1994; Schweizer et al. 1994; Nyam et al. 1995; Woodfield et al. 2003). This was significant for studies looking at upper gastrointestinal and biliary surgery (Anderson et al. 1996; Kellum et al. 1984; Blatzas et al. 1987; Tristaino et al. 1988; El-Mufti et al. 1989a; Periti and Tonelli 1990; Hjortrup et al. 1991; Rotman et al. 1991; Morris 1994).

Colorectal surgery showed a non-significant trend, p=0.07 (Franceschini et al. 1989; Garcia and Pedroso 1989; Thomas 1990; Lumley et al. 1992; Morris 1993; Nyam et al. 1995).
Table 6.3 Summary of meta-analysis of studies assessing pneumonia

<table>
<thead>
<tr>
<th>Category</th>
<th>Ceftriaxone</th>
<th>Other Antibiotics</th>
<th>Odds ratio summary statistic</th>
<th>Heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No of studies</td>
<td>CI No CI</td>
<td>CI No CI</td>
<td>OR 95% CI</td>
</tr>
<tr>
<td>All Studies</td>
<td>37</td>
<td>198 6777</td>
<td>284 6991</td>
<td>0.66 0.54-0.81</td>
</tr>
<tr>
<td>Single dose of Ceftriaxone compared to single or multiple doses of other antibiotics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Single</td>
<td>10</td>
<td>74 1934</td>
<td>121 2176</td>
<td>0.69 0.51-0.93</td>
</tr>
<tr>
<td>Multiple</td>
<td>27</td>
<td>104 4843</td>
<td>163 4815</td>
<td>0.65 0.49-0.85</td>
</tr>
<tr>
<td>Method of blinding: Not stated or poorly outlined against double blinding</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stated</td>
<td>14</td>
<td>99 2627</td>
<td>179 2871</td>
<td>0.59 0.45-0.77</td>
</tr>
<tr>
<td>Not stated</td>
<td>23</td>
<td>79 4150</td>
<td>105 4120</td>
<td>0.78 0.57-1.065</td>
</tr>
<tr>
<td>Definition of chest infection: Not stated or poorly defined against CDC definition</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Defined</td>
<td>7</td>
<td>54 2314</td>
<td>107 2646</td>
<td>0.62 0.44-0.88</td>
</tr>
<tr>
<td>Not defined</td>
<td>30</td>
<td>124 4463</td>
<td>177 4345</td>
<td>0.67 0.53-0.80</td>
</tr>
<tr>
<td>Comparison against different types of antibiotics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1st &amp; 2nd generation cephalosporins</td>
<td>22</td>
<td>93 4042</td>
<td>126 3936</td>
<td>0.71 0.53-0.95</td>
</tr>
<tr>
<td>3rd generation cephalosporins</td>
<td>6</td>
<td>15 983</td>
<td>27 955</td>
<td>0.56 0.30-1.06</td>
</tr>
<tr>
<td>Penicillins</td>
<td>4</td>
<td>33 541</td>
<td>51 535</td>
<td>0.63 0.39-1.01</td>
</tr>
<tr>
<td>Assessment of performance for different types of surgical procedures</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal – gastrointestinal</td>
<td>22</td>
<td>138 3755</td>
<td>221 3990</td>
<td>0.67 0.53-0.85</td>
</tr>
<tr>
<td>Pelvic</td>
<td>3</td>
<td>1 476</td>
<td>4 467</td>
<td>0.42 0.07-2.39</td>
</tr>
<tr>
<td>Clean</td>
<td>10</td>
<td>37 2435</td>
<td>56 2429</td>
<td>0.65 0.42-1.01</td>
</tr>
</tbody>
</table>

CI = chest infection or pneumonia
Clean surgery also showed a non-significant trend which was equally contributed to by cardiothoracic surgery (Geroulanos et al. 1985; Badel and Schmuziger 1989; Geroulanos et al. 1989; Sisto et al. 1994; Illic et al. 1997; Salminen et al. 1999; Kriaras et al. 2000), neurosurgery (Zhu et al. 2001) and orthopaedic surgery (Periti and Jacchia 1989).

There was no difference in antibiotic performance following pelvic surgery (Kunz et al. 1989; Fejgin et al. 1993; Periti et al. 1984).

Randomised studies with double blinding and a CDC definition of infection were then selected. As few studies used a CDC definition of pneumonia there were only three studies in this category (OR 0.63, CI 0.42-0.94, p=0.023) (Anderson et al. 1996; Zhu et al. 2001; Woodfield et al. 2003)

6.3.5 General observations and heterogeneity

6.3.5.1 General observations

1) Single and multiple doses of antibiotic being compared to ceftriaxone.

When a single dose of ceftriaxone was compared against a single dose of another antibiotic, there was always a difference in favour of ceftriaxone. This difference was not altered by giving additional doses of the other antibiotic. This suggests that the difference in antibiotic performance is not due to the long half-life of ceftriaxone.

2) When the mechanism of blinding was not clearly defined there was no increase in the magnitude of the difference (in favour of ceftriaxone) between the antibiotics being studied. This is in contrast to authorities (Jadad et al. 1996) who claim that when there is no double blinding the results usually favour the new intervention being studied.

3) Similarly, when the CDC definitions of infection were not used there was no consistent change in the difference in the performance of ceftriaxone and the other antibiotics being studied.

6.3.5.2 Heterogeneity

In terms of heterogeneity the p value for Cochran’s Q statistic was rarely less than 0.75 in the meta-analyses assessing wound infection and UTI, and rarely less than 0.50 in the meta-
analysis for pneumonia. Similarly, $I^2$ was rarely greater than 1 in the meta-analyses assessing wound infection and UTI, and rarely greater than 10 in the meta-analysis for pneumonia (a value of 50 demonstrates statistically significant heterogeneity (Mahid et al. 2006). In no case were the results of the most important individual studies outside the funnel plot or at variance with the overall results of the meta-analysis.

6.3.6 Results comparing ceftriaxone against placebo

An analysis of the studies which compared ceftriaxone with a placebo treatment confirmed the ability of ceftriaxone to reduce the incidence of wound infection (Childs et al. 1983; Luscher and Bruhwiler 1985; Harnoss et al. 1987; Kiff et al. 1988; Beksac et al. 1989; Cardamakis et al. 1991; Boxma et al. 1996; Dormann et al. 1999; Luthje et al. 2000; Mamsen et al. 1992; Papageorgiou et al. 1992; Ronconi et al. 1983; Petersen et al. 1988; Steiner et al. 2003) ($n=14$, OR 0.30, CI 0.22-0.42, $p<0.001$), pneumonia (Harnoss et al. 1987; Kiff et al. 1988; Boxma et al. 1996; Dormann et al. 1999; Luthje et al. 2000; Ronconi et al. 1983; Petersen et al. 1988; Mamsen et al. 1992; Steiner et al. 2003) ($n=9$, OR 0.38, CI 0.23-0.66, $p<0.001$) and urinary infection ($n=12$, OR 0.26, CI 0.18-0.36, $p<0.001$) (Luscher and Bruhwiler 1985; Harmes and Schumacker 1987; Harnoss et al. 1987; Kiff et al. 1988; Beksac et al. 1989; Cardamakis et al. 1991; Boxma et al. 1996; Luthje et al. 2000; Mamsen et al. 1992; Papageorgiou et al. 1992; Petersen et al. 1988; Steiner et al. 2003). It was not surprising to note that the degree of difference was greater than when a prophylactic antibiotic was being compared against ceftriaxone.

6.4 Discussion

6.4.1 Confirmed advantages for the use of ceftriaxone as a prophylactic antibiotic

This meta-analysis has demonstrated ceftriaxone to be a very effective and versatile prophylactic antibiotic. The following observations are supported:

1) Ceftriaxone effectively reduced the incidence of wound infection over and above the reduction of infection provided by most other prophylactic antibiotics. This difference was most noticeable when contaminated operations involving the colon or clean-contaminated operations involving the upper gastrointestinal tract or biliary system were performed.

2) There was no advantage in using ceftriaxone in clean surgery (cardiac, neurosurgery, orthopaedic).

3) Ceftriaxone can reduce the incidence of infections which are remote to the site of surgery, such as pneumonia and urinary tract infections.
4) The benefit in reducing pneumonia was most clearly seen in upper abdominal (upper gastrointestinal and biliary) surgery.

5) The benefit in reducing urinary tract infection was most clearly seen in obstetric and gynaecological surgery and colorectal surgery.

6) Ceftriaxone is very versatile in reducing infection after abdominal surgery, having the potential to reduce wound infection (especially when the surgery is not clean), to reduce pneumonia after upper abdominal surgery and to reduce urinary tract infection after colorectal and pelvic surgery.

6.4.2 Why would ceftriaxone have this advantage?

The consistent nature of these results is remarkable when we consider that ceftriaxone is being compared to other effective and appropriately chosen antibiotics. The difference in antibiotic performance is not because ceftriaxone provided a superior in-vitro cover (RN Jones et al. 1992; Woodfield et al. 2003). It is more likely to be related to the pharmacokinetics of ceftriaxone. With its exceptional tissue penetration (Mazzei and Periti 1989), the concentration of ceftriaxone throughout the skin, soft tissue, lung and bladder are well above the MIC90 and the minimal bactericidal concentration of most bacteria. The importance of this is that ceftriaxone is present at a higher concentration (relative to the MIC or MBC for most pathogens) than are other antibiotics which have been administered at their normal prophylactic dosage (see Figures 2.1 and 2.2). This would enable ceftriaxone to suppress bacterial activity more effectively than many other antibiotics, and is likely to account for its improved performance. The same mechanism may explain why a single ‘high dose’ of gentamicin provides more effective prophylaxis than three ‘standard doses’ of gentamicin over 24 hours (Zelenitsky et al. 2000).

In terms of UTI and pneumonia, it had been postulated in this thesis that ceftriaxone may provide an advantage over other prophylactic antibiotics because of its prolonged duration of action. In this meta-analysis the advantage for both UTI and pneumonia was the same when a single dose or multiple doses of the comparative antibiotic was given. This would suggest that the advantage of ceftriaxone in reducing chest and urinary infection is (at least partly) due to its better performance during the operative period, rather than being primarily due to its longer (24 hour) duration of action.
6.4.3 Robustness of results: Heterogeneity and sub-analysis

One difficulty in this meta-analysis is the inclusion of a diversity of studies. Differences such as the range of ‘other’ antibiotics being compared to ceftriaxone and the different types of operations being performed were able to be addressed by the sub analyses looking specifically at these differences. However there were also differences in the design and execution of studies. For example there were significant variations in the detection rate of infection, with the rate of chest infection following abdominal surgery ranging from <5% to 20%. There were also differences in the duration of follow-up, with many studies assessing infection in hospital and others assessing infection for 30 postoperative days. In this context it was encouraging that the funnel plots showed an even distribution of cases and that the heterogeneity figures for the majority of comparisons were robust. Equally important was the observation that the sensitivity analyses consistently gave very similar results. This adds to the confidence in the overall findings of the study. The consistency of the findings across a divergent patient population and over a range of conditions implies that the results of the meta-analysis are likely to be widely applicable.

6.4.4 Problems with CDC definitions of infection

Another difficulty was with the CDC definition of infection. The CDC definition of an incisional surgical site infection is intuitive, and was correctly defined in most studies. For urinary infection the CDC definition distinguishes between symptomatic urinary tract infection (SUTI) and asymptomatic bacteriuria (ASB). Of 35 studies, 21 defined a UTI, but in only 11 was this consistent with the definition of a SUTI (Beam et al. 1984; Harmes and Schumacker 1987; Lang et al. 1988; Beksa et al. 1989; Periti and Tonelli 1990; Stiver et al. 1990; Luke et al. 1991; Lumley et al. 1992; Kriaras et al. 2000; Woodfield et al. 2003). For chest infection the CDC definition requires a correlation of symptoms, signs, laboratory and radiological findings. The underlying pathology is also important (Garner et al. 1988). Of the 37 studies, 18 attempted to define pneumonia but in only 7 cases was this close to the CDC definition of infection. One reason for this difficulty is that some studies predated the 1988 CDC definitions of nosocomial infection (Garner et al. 1988). The range of definitions used was one reason why it was elected to perform a sensitivity analysis using otherwise high quality studies with a Jadad score of four or five. This analysis gave very similar results to the main analysis and other sensitivity analyses.
6.4.5 Literature on ceftriaxone meta-analysis

There have been two other meta-analyses of ceftriaxone antibiotic prophylaxis (Dietrich et al. 2002; Esposito S et al. 2004). The study by Dietrich compared ceftriaxone against other cephalosporins in 43 randomised studies, including open studies, performed between 1986 and 1996. Wound, urinary and chest infections were included. Ceftriaxone gave significantly better protection against wound infection and against UTI (difference in UTI was in all studies, but not for those with a CDC definition of UTI). There was a non-significant reduction in chest infection.

The second study, by Esposito et al, was found subsequent to our own meta-analysis being performed. This was similar to our meta-analysis as it compared ceftriaxone against all other prophylactic antibiotics in RCT performed between 1984 and 2003. It identified 48 studies, and showed that the performance of ceftriaxone was better than other antibiotics for wound, chest and urinary infections. This study noted that ceftriaxone was better in clean surgery for wound infection and pneumonia (but not urinary infection), and that it performed better in clean-contaminated surgery for all three infections.

Our meta-analysis, which included non-cephalosporin antibiotics, identified a number of studies not included in the other two meta-analyses. The main difference in comparison to the first study was the better performance of ceftriaxone against chest infection. The studies performed between 1996 and 2005, tended to be of a higher quality (Song and Glenny 1998), and more consistently identified a difference in chest infection.

The main difference in the second study was in the sub-analysis of clean and clean-contaminated surgery. In this study open randomised trials were included, all acute operations were excluded and the reference to their definition of a clean procedure was unhelpful. In our sub-analysis using ceftriaxone in clean surgery (rather than another appropriately selected antibiotic) did not reduce surgical site infections. This sub-analysis of clean surgery included studies looking at cardiac surgery, neurosurgery and orthopaedic surgery. Both studies confirmed that in clean surgery ceftriaxone had no advantage over other antibiotics in preventing urinary tract infection.
6.4.6 Both the efficacy and cost effectiveness of ceftriaxone support using this as a first line prophylactic agent

How important are these findings? Compared to the reduction in wound infection that occurred when prophylactic antibiotics were introduced, the impact of changing one appropriate prophylactic antibiotic to a more effective option is relatively small. Using a number needed to treat analysis, using all studies included in the meta-analysis, the number needed to treat (NNT) to prevent one wound infection is 70, one urinary infection is 49 and one chest infection is 87. However, if we select high risk groups the NNT to prevent one infection reduces. For wound infection in patients undergoing gastrointestinal surgery it is 42, for urinary infection in patients undergoing colorectal or gynaecological surgery it is 19, and for chest infection after upper gastrointestinal or hepatobiliary surgery it is 33. This difference emphasizes that the benefits of ceftriaxone become more pronounced when the risk of infection is increased.

In terms of cost effectiveness, as the cost of an infection is vastly greater than the cost of the antibiotic, significant savings can be made (Woodfield et al. 2005). The cost of ceftriaxone is also likely to become more competitive as generic forms of the medicine are now available.

6.5 Limitations of this study

Three potential problems were identified and have been discussed in this chapter. The heterogeneity of studies included in the meta-analysis is discussed in figure 6.4.3. It is noted that the heterogeneity statistics and funnel plots suggest that this was not a major problem. Issues with the CDC definition of infection are discussed in 6.4.4 and issues pertaining to Clostridium difficile infection are discussed in 6.5.

6.6 Conclusion

In conclusion this meta-analysis has accepted the alternative hypothesis, that ceftriaxone is more effective than other appropriately selected antibiotics in decreasing wound infection, chest infection and urinary infection. The advantages identified included a reduction in wound infection after abdominal surgery (but not after clean surgery), a reduction in urinary tract infection - especially after colorectal and O&G surgery, and a reduction in pneumonia after upper abdominal surgery. In these settings these results would support using ceftriaxone as a first line prophylactic antibiotic.
6.7 Clostridium difficile infection, antibiotic resistance and caution about how to best apply these findings

Two important concerns about the use of third generation cephalosporins is the increasing incidence of antibiotic resistance and Clostridium difficile infection. Unfortunately this meta-analysis, as well as other meta-analyses looking at prophylactic antibiotics, have not been able to give additional information on these issues. The reason for this is that most of the RCTs looking at the prophylactic use of ceftriaxone were performed before 2000. The incidence of Clostridium difficile infection was very low, with no cases being documented in many studies. MRSA infection and antibiotic resistance to most of the traditionally sensitive pathogenic bacteria were also infrequent events in the majority of these studies. Other reasons why it has been difficult to get good information on these two issues are discussed briefly in Appendix A.

The findings of this meta-analysis therefore need to be interpreted in the context of changing patterns of Clostridium difficile infection, MRSA infection and the changing patterns of antibiotic resistance.

6.8 Further questions for study

To better understand the mechanism of the additional prophylactic efficacy of ceftriaxone, work in the following areas would be helpful:

a) To determine what the ideal tissue concentration of an antibiotic above the MIC 90 or MBC should be to prevent wound infection in the different scenarios of clean, clean-contaminated and contaminated wounds.

b) To determine the ideal pharmacokinetic profile of an antibiotic in reducing chest and urinary infection. Is the main effect due to the intraoperative action of an antibiotic, and what is it the contribution from antimicrobial cover in the post-operative period?

The answers to these questions may help us to provide better antibiotic prophylaxis in the future.

6.9 Publication

Part B

Prediction of Complications
CHAPTER 7
7. A prospective study of the ability of the surgeon to predict major complications

7.1 Introduction
Although prediction is fundamental to clinical practice, there is a lack of objective data on how good the surgeon is at predicting the likelihood of major complications after surgery. The question of whether surgical prediction of major complications can be used as a tool to help influence decision making is an interesting one, but cannot be addressed without adequate data on the accuracy of prediction.

Our approach to collecting this data was to perform two prospective studies. Study One assessed the accuracy of both the preoperative and postoperative prediction of major complications. This is summarised in section 7.2.1. Study Two was designed to assess the accuracy of prediction in comparison to the ‘gold standard’ of POSSUM, and to assess if the surgeon’s prediction of major complications could be improved. Chapter 7 describes this study.

7.2 Background information for Study Two: Summary of previous studies looking at clinical prediction of complications
7.2.1 Study One: The Dunedin Study
Study One is summarized here as this sets the scene for the clinical questions that were investigated in Study Two. It is not presented in detail however as the main focus of this chapter is on Study Two.

In Study One the surgeon prospectively estimated the risk of a major complication on a 100mm VAS in the operating room immediately before and after surgery in 1013 patients. If the score was changed postoperatively the surgeon was asked to document the reason for this. A change of >5mm was categorized as a change in the prediction of a major complication. Clinical evaluation for complications was performed daily until hospital discharge and all patients were reviewed at a minimum of thirty days post operatively. Definitions of major complications were the same as had been used previously (Pettigrew and Hill 1996). There were 31 deaths (3.06%), 96 cases with major morbidity (9.48%) and 245 cases with minor morbidity (24.19%). Altogether, 127 (12.4%) patients had a major complication.
The three main areas investigated were the accuracy of prediction, the importance of the surgeon changing the VAS prediction at the completion of surgery, and the impact of incorporating the VAS prediction into a risk model for predicting major complications.

The surgeons’ prediction discriminated between patients who did and did not develop a complication (Table 7.1).

**Table 7.1 Surgeons’ VAS scores for patients with and without complications**

<table>
<thead>
<tr>
<th>Mortality</th>
<th>Preoperative Scores</th>
<th>Postoperative Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mortality</td>
<td>No Mortality</td>
</tr>
<tr>
<td>Number of predictions</td>
<td>31</td>
<td>982</td>
</tr>
<tr>
<td>Median score (range)</td>
<td>42 (5-74)</td>
<td>20 (0-94)</td>
</tr>
<tr>
<td>p Value</td>
<td>P&lt;0.001</td>
<td>P&lt;0.001</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Major Complication</th>
<th>Preoperative Scores</th>
<th>Postoperative Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of predictions</td>
<td>127</td>
<td>886</td>
</tr>
<tr>
<td>Median (range)</td>
<td>27 (2-80)</td>
<td>19 (0-94)</td>
</tr>
<tr>
<td>p Value</td>
<td>P&lt;0.001</td>
<td>P&lt;0.001</td>
</tr>
</tbody>
</table>

Major complication is mortality plus major morbidity

The Mann Whitney U test was used for comparison of those with and without complications

Although the surgeon was able to make a meaningful prediction, the discrimination of prediction, or the c statistic, as determined by the area under the ROC curve, showed only moderate discrimination. (A c statistic of 0.7 to 0.8 represents reasonable discrimination (Tekkis et al 2002)). The c statistic for identifying who would or would not develop a major complication was 0.67 for preoperative VAS prediction, 0.69 for postoperative VAS prediction, 0.69 for the patient’s age and 0.64 for the duration of surgery. The surgeon’s prediction was therefore similar to other individual indicators of risk. In this study it was not sensitive or specific enough to reliably describe surgical risk, or to be used to influence perioperative decision making.
The importance of the surgeon changing his/her prediction of complications (Pettigrew et al. 1997) was confirmed. When the surgeon increased the VAS score there was a statistically significant doubling of the rate of major complications. Patients with an increase in VAS score had a higher mortality rate (6.3% v 2.4%, p=0.006), ‘major complication’ rate (20.1% v 11.0%, p=0.001) and ‘all complication’ rate (48.3% v 34.3%, p=0.001) than those with no change in prediction or a decreased VAS score. The reasons given for changing the VAS score are categorized in Table 7.2.

### Table 7.2 Reasons given by the surgeon for changing the VAS score for major complications

<table>
<thead>
<tr>
<th>Category</th>
<th>Increased Risk</th>
<th>Decreased risk</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of cases with change in risk</td>
<td>174</td>
<td>128</td>
<td>302</td>
</tr>
<tr>
<td>Reason stated</td>
<td>132</td>
<td>53</td>
<td>185</td>
</tr>
<tr>
<td>Reason not stated</td>
<td>42</td>
<td>75</td>
<td>117</td>
</tr>
<tr>
<td>Number of stated reasons</td>
<td>181</td>
<td>60</td>
<td>241</td>
</tr>
<tr>
<td>Technical Reasons</td>
<td>133</td>
<td>44</td>
<td>177</td>
</tr>
<tr>
<td>Extent of Disease</td>
<td>43</td>
<td>16</td>
<td>59</td>
</tr>
<tr>
<td>Anaesthetic/Medical</td>
<td>5</td>
<td></td>
<td>5</td>
</tr>
</tbody>
</table>

The main technical reasons included the operation being more technically demanding or more extensive than anticipated, gastrointestinal or biliary contamination, anastomotic issues and the duration of surgery. The main disease related reasons for changing the VAS score were the extent of sepsis, the extent to which the underlying malignancy had spread, and a new diagnosis of intestinal ischemia. It was not documented whether the reason given for increasing the postoperative score was inevitable, contributed to by ‘technical proficiency’ or due to a technical error. However these results did confirm that the surgeon can identify a group of high risk patients, and that what the surgeon sees and does intraoperatively is important in determining the outcome of surgery.

The surgeon’s prediction was then combined with the Complexity and Risk Adjusted model (Pillai et al. 1999) for predicting major complications after all general surgery. This model, which is based on twelve variables prospectively collected on the Otago Surgical Audit for General Surgery (Department of Surgery Clinical Audit Research Unit 1996), is designed to
be adjusted for different population groups. In Table 7.3 the prediction of complications using the general model (general surgical model) and the model after optimization for the patient mix used in the study (abdominal surgery model) are presented. This shows that the Otago Audit Model’s discrimination was modest and that the addition of the surgeon’s preoperative VAS score resulted in a small improvement in the goodness of fit and also an improvement in the discrimination of the model. No collinearity was present between the surgeon’s VAS score and other prognostic variables (variance inflation factors <2).

Table 7.3 Summary of logistic regression analysis for the prediction of operative risk using the Otago Audit model

<table>
<thead>
<tr>
<th></th>
<th>General Surgery Model</th>
<th></th>
<th>Abdominal Surgery Model</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model only</td>
<td>Adding VAS</td>
<td>Model only</td>
<td>Adding VAS</td>
</tr>
<tr>
<td>Chi-Square for covariates</td>
<td>$\chi^2=154.5$ with 19 DF, $p=0.001$</td>
<td>$\chi^2=169.3$ with 22 DF, $p=0.001$</td>
<td>$\chi^2=160.7$ with 12 DF, $p=0.001$</td>
<td>$\chi^2=172.2$ with 15 DF, $p=0.001$</td>
</tr>
<tr>
<td>Hosmer and Lemeshow Goodness of fit</td>
<td>$\chi^2=15.1$ with 8 DF, $p=0.056$</td>
<td>$\chi^2=8.6$ with 8 DF, $p=0.374$</td>
<td>$\chi^2=6.1$ with 8 DF, $p=0.631$</td>
<td>$\chi^2=3.9$ with 8DF, $p=0.863$</td>
</tr>
<tr>
<td>Area under ROC curve</td>
<td>0.76</td>
<td>0.77</td>
<td>0.76</td>
<td>0.77</td>
</tr>
</tbody>
</table>

DF is degrees of freedom

The finding of the absence of collinearity between the surgeon’s VAS score and the other prognostic variables was an important and unexpected finding. It was unexpected as it had been anticipated that the surgeon’s VAS score would be based on the known prognostic factors already included in the model. This result however confirmed that the surgeon’s risk assessment incorporates other factors, and that surgical prediction can improve the prediction of perioperative complications. This work was published in the World Journal of Surgery (Woodfield et al. 2007).

7.2.2 Other studies looking at clinical prediction

Four other studies, assessing the prediction of complications using a VAS, were available when study two was commenced. These have been introduced in section 1.14. Two small studies using an unanchored VAS (a visual analogue scale with no markings to define how the scale should be used), which assessed the global ‘gut feeling’ assessment of the surgeon
immediately prior to surgery, did not significantly predict complications (Pettigrew and Hill 1986; Pettigrew et al. 1987). However in one of these studies (Pettigrew and Hill 1986) a structured clinical assessment looking at cardiorespiratory disease, pre-existing sepsis and nutritional status did predict complications. In the third study (Arvidsson et al. 1996), of over 1000 patients, using a clearly labelled preoperative VAS which was scored at a preadmission clinic, the VAS score, ASA grade, the patient’s age and a procedural magnitude score were all significantly correlated with adverse events, with the VAS score being the most successful discriminator for severe adverse events. The suggestion that a structured or systematic clinical assessment may result in a more accurate preoperative prediction of risk than an intuitive assessment alone was the basis for introducing a multifactorial VAS in phase II of our study (section 7.4.6.2).

Two of these four studies had also looked at postoperative risk assessment. Both of these showed a very good prediction of major complications by the surgeon (Pettigrew et al. 1987, Hartley and Sagar 1994). In one the clinical prediction was also noted to be comparable to POSSUM in its ability to accurately identify a group of high risk patients (Hartley and Sagar 1994).

7.3 Rationale for Study Two: The Leeds Study

7.3.1. Outstanding Questions

Following study one a number of questions remained:

1) How good is surgical prediction? Although the risk assessment by the surgeon had been shown to be a useful predictor of outcome, in terms of discrimination this was only moderate, and was not sensitive or specific enough to be used to influence perioperative decision making. However there was no control or comparison to compare surgical prediction against. In this context we wanted to reassess the accuracy of surgical prediction and to compare this against a ‘gold standard’ such as POSSUM.

2) Can surgical prediction be improved?

3) Can surgical prediction be used to improve a model assessing the risk of complications?

4) What is the importance of the surgeon changing the risk of complications postoperatively?

One aspect of this question which was not addressed in study one was how much of the change in prediction was related to the performance of the surgeon (due to technical problems), and if using a change in VAS prediction may be a helpful way of highlighting technical issues.
Study Two was designed to help answer these questions. We were also interested in repeating the study in a different setting – with different surgeons, in a different country, working in a different health system.

### 7.3.2. Approach to answering these questions

#### 7.3.2.1 Improving prediction

There are two main approaches to predicting risk. The first is an intuitive assessment, based on an overview or global assessment of the clinical problem. Alternatively one could take a more structured approach which consciously takes into account the different components of risk. Previous studies had implied that a structured approach may be more accurate in predicting risk (section 7.2.2). In this study I wanted to assess if prediction could be improved by the surgeon using a more structured approach. This was investigated by using a multifactorial VAS, which asked the surgeon to assess a number of domains of risk, each with their own VAS scale, before making a global prediction of complications. The null hypothesis was that introducing a more structured assessment would not improve the prediction of complications.

I was also interested in the possibility of feedback improving the prediction of risk. There are medical examples of feedback resulting in an improvement in performance in the areas of improving wound infection and in reducing surgical complications (section 7.8.3). In these examples feedback resulted in a change in action. This is slightly different to our question, where we are trying to assess if feedback can result in a change in assessment or cognition. This was investigated by giving meaningful outcome feedback to each surgeon. The null hypothesis was that providing feedback on complications would not improve the prediction of complications.

#### 7.3.2.2 Technical proficiency and a change in prediction

To help us understand the nature of events that influenced a change in prediction it was decided to ask the surgeon if a change in prediction was either inevitable, related to technical proficiency or due to an error.
7.3.2.3 Surgical prediction and models assessing the risk of complications
To investigate this it was decided to develop a model for predicting the major complications that occurred in the study. This would then enable us to assess if adding the surgeon's VAS score would improve the discrimination and/or goodness of fit of the model. The null hypothesis was that the addition of the surgeon’s VAS score would not improve the performance of the model.

7.4 Design
7.4.1 Overview
A prospective study on the prediction of major complications involving three phases was designed. These were as follows:
Phase I: Initial global assessment, repeating the design of Study One. The design of phase I was kept as similar as possible to study one. Some adjustments, such as the definition of major complications, were made because of new recommendations in the medical literature.
Phase II: Structured assessment. This was to assess if a more structured assessment would improve prediction. In this phase a multifactorial VAS, with a set of VAS scales for testing different domains of risk, was introduced.
Phase III: Impact of feedback. This addressed the question of whether clinical outcome feedback would improve prediction. It was planned to give quality outcome feedback to all participating surgeons at the end of phase II. The study would then be continued using the same methodology as in phase II.

7.4.2 Inclusion criteria
As the design was kept as similar to study one as possible; all major abdominal surgery and vascular surgery, including acute surgery was included. The case-mix was determined by the main specialties at the Leeds General Infirmary (LGI) at the time. These were colorectal, upper gastrointestinal and vascular surgery. Inclusion of cases depended on the willingness of the consultants to be involved. All except one surgeon participated.

7.4.3 Definition of complications
In line with changes of definitions in the medical literature a major complication was determined primarily by the severity of the intervention required to treat the complication (Dindo et al. 2004). A major complication was therefore defined as a grade III to V complication according to the Clavien-Dindo classification, where grade III requires an
operation or a major radiological or endoscopic procedure; grade IV requires organ failure and/or admission to an intensive care unit and grade V complications result in death. Also included in our definition of major complications were potentially life threatening events such as a pulmonary embolus, myocardial infarction or stroke (even if these problems were managed in a general hospital ward) and a complication which was likely to lead to a major intervention at a later stage, such as an enterocutaneous fistula. The definition was therefore in line with what a clinical would consider to be a ‘major event’. The definitions were carefully explained, and an information sheet (Figure 7.1) with definitions and illustrative examples were kept in theatre and also distributed to all participating surgeons.

**Definitions of Complications**

Conceptually a major complication is any major deviation from the normal postoperative course. However when it comes to defining exactly what a major complication is (what to include in the definition of a major complication), there is no agreed upon criteria. For the study we have defined a major complication according to the following three criteria

1) **Severity of intervention required to treat the complication**

   **Any complication that requires a significant intervention to treat**

   In Ann Surg 2004, Vol 240, pg 205-213 a system of grading complications according to the severity of the intervention required to treat the complication is proposed. This is summarised as

   I:  No need for pharmacological, surgical, endoscopic or radiological intervention
   II:  Pharmacological treatment, blood transfusion or TPN required
   III: A surgical intervention or a major endoscopic or major radiological intervention is required (includes returning to theatre, percutaneous drainage of an abscess under radiological guidance, angioplasty, and endoscopic stent placement)
   IV: Life threatening complication with failure of a single organ or multiple organs. This requires a level of care that cannot be given in the ward, resulting in a transfer to ICU (or HDU).
   V: Death

   **For this study levels III to V are defined as major complications.**
Examples.

- A chest infection which responds to antibiotics in the ward would not be a major complication, but if the patient became sufficiently sick to required transfer to HDU for closer monitoring or CPAP then it would be a major complication.
- A wound infection which settles with antibiotics and dressings would not be a major complication, but if the patient returned to theatre for wound debridement it would be a major complication.
- Postoperative observation and rapid weaning of support in ICU would not be considered to be a major complication, but if on-going support was required it would become a major complication.

2) *Any complication which is potentially life threatening*

Examples of this would include any myocardial infarction, PE or stroke. These complications have a wide spectrum of severity, ranging from having minimal impact on the patient to sudden death. Because of the potential of sudden death they are included as major complications.

3) *An unplanned surgical event with a major impact on the patient, and usually with the potential for a further significant intervention at a later stage*

Examples would include a postoperative enterocutaneous fistula or a vascular graft infection.

Although this may not involve a major intervention within the first postoperative month (TPN or long course of antibiotics being a level II rather than a level III intervention), such a complication is a significant unplanned event, has a major impact on the patient and may well lead to further interventions in the future.

**Timeframe of prediction of major complication**

The prediction is for any event occurring up to 30 days after surgery

---

**Figure 7.1** Form used to describe the definition of a major complication
7.4.4 Documentation of risk factors

7.4.4.1 Documentation of preoperative risk factors

A preoperative risk form (Figure 7.2), which included the clinical risk factors for a number of predictive models, including POSSUM, was prospectively collected. The collection was performed by members of each clinical team, and the completion of the forms was checked by the primary clinical researcher.

![Figure 7.2 Form for preadmission and preoperative risk factors](image)

Figure 7.2 Form for preadmission and preoperative risk factors
7.4.4.2 Documentation of intraoperative risk factors

Intraoperative risk factors, including those used in a number of predictive models, was collected by filling out an ‘operative details’ form (Figure 7.3). This was completed by a member of the surgical team at the completion of surgery.

Figure 7.3 Form for intraoperative risk factors
7.4.5 Documentation of complications

Hospital complications were assessed daily by the clinical team and also three times a week by a clinical researcher (Figure 7.4).

![Figure 7.4 Form for documentation of complications](image-url)
This included the clinical researcher performing a ward round where both the clinical notes and patients were reviewed. Points that needed clarification were discussed with the clinical team looking after the patient.

After 30 days patients who had been discharged were contacted using a telephone survey. A structured questionnaire was used. Patients were asked if they had any additional problems since discharge from hospital (as a general question), if they had any problems with their wound (as a specific question) and if they had been to visit their general practitioner (in case they had forgotten to mention something!). Outpatient notes were reviewed if there were any uncertainties after the phone questionnaire. The main reasons for this were if the patient could not be contacted by phone, or if the patient was not able to ‘medically’ describe the nature of problem they had experienced at home.

Although this study was designed to assess the prediction of major complications, all complications were documented. Clinical outcomes were recorded in a separate database to the VAS score and the patient’s risk factors.

7.4.6 Documenting the prediction of a major complications

7.4.6.1 VAS form used in Phase I

The surgeon estimated the risk of a major complication on a 100mm visual analogue scale in the operating room immediately before and after surgery (Figure 7.5). If the score was changed postoperatively the surgeon was asked to document the reason for this. Both the primary surgeon and the assisting surgeon were invited participate. This meant that in a number of cases two predictions were made.

The linear nature of the VAS was discussed with each participant before involvement in the study. Written instructions were also attached to the VAS form (Figure 7.5a). In order to check the reproducibility of the VAS being used as a linear scale a questionnaire was given to all surgeons who contributed in phase I of the study (section 8.2.2, Figure 8.1).

7.4.6.2 Prediction using multiple VAS scales in Phases II and III

A new VAS form (Figure 7.6) was introduced in the second phase of the study. This was included an additional six subscales, each of 100mm, corresponding to the different components of risk.
Completion and Interpretation of the Visual Analogue Scale (VAS)

The scale is 100mm long. The prediction of the doctor is being compared to a POSSUM prediction of morbidity, which gives a percentage prediction of morbidity. The VAS should be used as a linear scale, with a 25% prediction of a major complication being a quarter of the way along the VAS and a 50% prediction half way along the VAS. This will enable an accurate interpretation and comparison to be made.

Figure 7.5 (a) Form for surgeon’s assessment of major complications in Phase I
(b) Instructions about completion of the global VAS
Prediction of Perioperative Complications

Patient Details
(or place patients identification sticker)
Name
Date of Birth
Hospital Number

Operative Details
Operative Procedure
Date
Surgeon

Preoperative Risk Assessment

Contribution of different risk factors to the risk of major perioperative complications....

Risk Factor
Functional Status of patient
Low risk (State of well being, physical activity, walking distance, emotional and social wellbeing)
High Risk

Severity of pathology
Low Risk (nutrition, weight loss, albumin, extent of sepsis, extent of malignancy)
High Risk

Cardiovascular & respiratory co-morbidity
Low Risk (HID, Valve disease, CCF, severity of COPD, chest infection)
High Risk

Complexity of surgery
Low Risk (degree of technical difficulty, duration of surgery)
High Risk

Acuity of surgery
Low Risk (How quickly must the patient be in theatre, if performing surgery during the night is felt to be important)
High Risk

Other risk factors
Low Risk
State main reason
(cirrhosis, morbid obesity, surgeons experience, hospital system problems.....)
High Risk

Global Assessment of Risk of Major complication
This was designed to address the question, “Can clinical prediction be improved by providing prompts to encourage the clinician to think about specific areas of risk before making a global
assessment, or is a structured assessment better than ‘gut feeling’ assessment?’ The six domains of risk that the clinician was invited to assess before making a global assessment were: functional status, cardiac and respiratory morbidity, severity of pathology, complexity of surgery, urgency of surgery and ‘other.’ The selection of the subscales was based on a review of risk factors used in multifactorial models in the medical literature.

7.4.6.3 Prediction using POSSUM
POSSUM morbidity score and the Portsmouth POSSUM (P-POSSUM) mortality score were used in this study as a ‘benchmark’ or ‘gold standard’ to compare against the surgical prediction of major complications. POSSUM was chosen as this is the most widely established risk model in use, it was able to predict complications across the range of operations performed in this study, and it assesses morbidity as well as mortality. The individual risk factors which were prospectively collected were entered into a Microsoft Access database which used the POSSUM morbidity equation and the P-POSSUM mortality equation to generate a percentage prediction for morbidity and mortality. Although POSSUM provides us with an excellent comparative prediction, there are important differences in the design of the POSSUM score and surgical prediction using a VAS, which means that it is not possible to conclude that one is statistically better than the other (section 7.7.5).

7.4.7 Clinical feedback
Feedback was given to the surgeons at the completion of Phase II. There was a pause in the study between Phases II and III to enable the results of Phases I and II to be analysed and then to be presented to the surgical department. Phase II patients that did not have final outcome data, as they had been operated on within the previous 30 days, were not included in the feedback analysis. The feedback included the results for all surgeons combined, and then the results for each individual surgeon.

Feedback for all surgeons included:

- A summary of operative procedures.
- A summary of complications.
- The overall accuracy of the clinician’s prediction of major complications and POSSUM’s prediction of complications. Discrimination statistics (area under the ROC curve) and calibration bar graphs were produced.
• A comparison between the POSSUM prediction of mortality and observed mortality. Following this each surgeon was given a printout of his/her individual results. The information in this document included:
  • The number of predictions made.
  • The mean, median and the distribution of preoperative and postoperative VAS scores.
  • The discrimination and goodness of fit of the individual’s preoperative and postoperative scores for major complications.
  • The calibration of preoperative and postoperative scores. This looked at the accuracy of prediction in low risk, medium risk and high risk cases.
  • A weighted ratio of predicted vs observed complications.
  • Some general comments about improvement in prediction. When there was a specific case where a useful ‘lesson’ about prediction could be learnt, this was discussed.
  • A comparison of the individual’s results to the overall surgical prediction and to POSSUM.

Following clinical feedback being presented and discussed the study was recommenced.

7.4.8 Statistics
7.4.8.1 Statistics for assessing sample size
Sample size calculation was based on the Wilcoxin statistic for the sample size required to compare areas under the ROC curve. Tables of the sample size required, with a one sided test of significance p=0.05 have been published (Hanley and McNeil 1982). The discrimination of prediction in phase I, based on study one was expected to be approximately 0.70. To test for an improvement in prediction up to 0.80 in phase II of the study, with an 80% power and a one sided test of significance p=0.05 would require 158 predictions in each group. To test for an additional improvement in the discrimination of prediction after feedback in phase III of the study, for example from 0.80 to 0.90, with a power of 80% and a one sided test of significance p=0.05 would require 110 predictions in each group. These figures were an estimation, as it was not known what the discrimination of prediction would be in the separate phases of the study. It was therefore decided to include more predictions than required by this sample size analysis. The actual comparisons made in this study [for example to compare a discrimination of 0.78 (from phase I and II) and 0.895 (from phase III) required a sample size of approximately 110 predictions in each group] ended up requiring a smaller sample size than the number of predictions included in the study. Our findings therefore maintained a power of >80% and a Type I error of <5%.
7.4.8.2 Statistical analysis of the results

Descriptive statistics were used to describe the type of surgery and the frequency of complications. The Mann Whitney U test was used to compare the VAS scores of patients with and without complications. The performance of clinical prediction was tested using discrimination and goodness of fit statistics, and the ratio of predicted to observed complications. The discrimination of prediction, which refers to the ability of prediction to separate patients who develop a complication from those who do not develop a complication, was calculated using the area under the ROC curve (c statistic). The goodness of fit, which refers to the ability of prediction to assign the correct probabilities of outcome to individual patients (low risk to low risk patients, high risk to high risk patients) was calculated using the Hosmer-Lemeshow test, and was also visually displayed using a ‘calibration bar graph,’ which divided the database into deciles, and for each decile compared the predicted and observed complications (Tekkis et al. 2002). Other statistics, such as those used for model development, are discussed in the appropriate sections.

7.4.9 Ethics committee approval

The Leeds Teaching Hospitals NHS Trust Research Ethics Committee, and the Research and Development Department, Leeds Teaching Hospitals NHS Trust approved the study. As this was looking at using ‘a new and possibly more helpful scoring system’ it was recommended that the study be incorporated into the audit program of the department of surgery.

7.5 Results

7.5.1 Number of patients and predictions

In Phase I, 357 patients were enrolled, with 565 surgical predictions being completed. In Phase II, 363 patients were enrolled with 537 surgical predictions being completed. In Phase III, 139 patients were enrolled with 193 surgical predictions being completed. A total of 1295 surgical predictions are therefore included in the study results.

7.5.2 Number of predictors

This study was open to all surgeons (see Table 7.13). Altogether, 1295 predictions were made by 58 people; 1127 of these predictions (87%) were completed by 24 people, who submitted a minimum of 13 predictions (Figure 7.7).
Figure 7.7 Number of predictions by different surgeons

7.5.3 Demographics and pathology
The demographics and pathology of the patients are presented in Table 7.4.

7.5.4 Details of surgery
The type of surgery is presented in Table 7.5. The majority of patients enrolled in the study were undergoing major complex procedures. Of these 646 cases were elective and 213 cases (25%) were acute.

Table 7.4 Demographics and pathology

<table>
<thead>
<tr>
<th>Demographics</th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>192</td>
<td>193</td>
<td>83</td>
<td>468</td>
</tr>
<tr>
<td>Female</td>
<td>165</td>
<td>170</td>
<td>56</td>
<td>391</td>
</tr>
<tr>
<td>Age (median, range)</td>
<td>61 (17-92)</td>
<td>63 (19-90)</td>
<td>64 (16-92)</td>
<td>63 (16-92)</td>
</tr>
<tr>
<td>Pathology</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Inflammatory</td>
<td>138</td>
<td>112</td>
<td>44</td>
<td>294</td>
</tr>
<tr>
<td>Malignant</td>
<td>132</td>
<td>113</td>
<td>35</td>
<td>280</td>
</tr>
<tr>
<td>‘Reconstructive’</td>
<td>58</td>
<td>56</td>
<td>25</td>
<td>139</td>
</tr>
<tr>
<td>Vascular</td>
<td>29</td>
<td>82</td>
<td>35</td>
<td>146</td>
</tr>
</tbody>
</table>

‘Reconstructive’ includes hernia, gastro-oesophageal reflux surgery, prolapse surgery, stoma closure and ileoanal pouch formation.
<table>
<thead>
<tr>
<th>Type of surgery</th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Upper GI Surgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cholecystectomy</td>
<td>24</td>
<td>20</td>
<td>15</td>
<td>59</td>
</tr>
<tr>
<td>Bile Duct</td>
<td>5</td>
<td>3</td>
<td></td>
<td>8</td>
</tr>
<tr>
<td>Gastrectomy</td>
<td>16</td>
<td>11</td>
<td>1</td>
<td>28</td>
</tr>
<tr>
<td>Gastric bypass</td>
<td>5</td>
<td>4</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>Gastro-oesophageal junction</td>
<td>6</td>
<td>6</td>
<td>2</td>
<td>14</td>
</tr>
<tr>
<td>Oesophagectomy</td>
<td>14</td>
<td>11</td>
<td>5</td>
<td>30</td>
</tr>
<tr>
<td>Pancreas</td>
<td>5</td>
<td>6</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>Spleen</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Colorectal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anterior resection</td>
<td>25</td>
<td>29</td>
<td>7</td>
<td>61</td>
</tr>
<tr>
<td>Pelvic clearance, Hartmann’s, abdominoperineal resection</td>
<td>25</td>
<td>18</td>
<td>5</td>
<td>48</td>
</tr>
<tr>
<td>Sacrectomy</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>Total colectomy</td>
<td>13</td>
<td>9</td>
<td>1</td>
<td>23</td>
</tr>
<tr>
<td>Segmental colectomy</td>
<td>45</td>
<td>37</td>
<td>17</td>
<td>99</td>
</tr>
<tr>
<td>Stoma formation</td>
<td>17</td>
<td>10</td>
<td></td>
<td>27</td>
</tr>
<tr>
<td>Stoma closure</td>
<td>34</td>
<td>26</td>
<td>12</td>
<td>72</td>
</tr>
<tr>
<td>Fistula closure</td>
<td>4</td>
<td>1</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Ileoanal Pouch</td>
<td>20</td>
<td>5</td>
<td>2</td>
<td>27</td>
</tr>
<tr>
<td>Prolapse</td>
<td>2</td>
<td>7</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Other General Surgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abdominal Hernia</td>
<td>11</td>
<td>12</td>
<td>7</td>
<td>30</td>
</tr>
<tr>
<td>Laparoscopy +/- appendicectomy</td>
<td>10</td>
<td>9</td>
<td>10</td>
<td>29</td>
</tr>
<tr>
<td>Laparotomy</td>
<td>18</td>
<td>20</td>
<td>3</td>
<td>41</td>
</tr>
<tr>
<td>Multiple organ resection</td>
<td>3</td>
<td>2</td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>Small bowel</td>
<td>16</td>
<td>21</td>
<td>7</td>
<td>44</td>
</tr>
<tr>
<td>Peptic ulcer disease</td>
<td>5</td>
<td>6</td>
<td>3</td>
<td>14</td>
</tr>
<tr>
<td>Other General</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Vascular Surgery</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aneurysm</td>
<td>14</td>
<td>24</td>
<td>15</td>
<td>53</td>
</tr>
<tr>
<td>Carotid</td>
<td>8</td>
<td>20</td>
<td>8</td>
<td>36</td>
</tr>
<tr>
<td>Limb endarterectomy</td>
<td>5</td>
<td>12</td>
<td>4</td>
<td>21</td>
</tr>
<tr>
<td>limb bypass</td>
<td>1</td>
<td>18</td>
<td>4</td>
<td>23</td>
</tr>
<tr>
<td>Other vascular</td>
<td>3</td>
<td>10</td>
<td>5</td>
<td>18</td>
</tr>
</tbody>
</table>

2 cases in Phase I and 3 in Phase II had more than one major procedure at the time of their operation.
7.5.5 Complications

The number of patients who developed a complication is presented in Table 7.6. As an adverse event may lead to more than one complication, there were a total of 264 major complications in 176 patients with a major complication. There were 988 complications in 509 patients who developed any complication. The breakdown of major complications is presented in Table 7.7.

Table 7.6 Number of patients with a complication

<table>
<thead>
<tr>
<th></th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>357</td>
<td>363</td>
<td>139</td>
<td>859</td>
</tr>
<tr>
<td>Mortality</td>
<td>15</td>
<td>23</td>
<td>5</td>
<td>43 (5.0)</td>
</tr>
<tr>
<td>Major morbidity</td>
<td>71</td>
<td>45</td>
<td>17</td>
<td>133</td>
</tr>
<tr>
<td>Major complications</td>
<td>86 (24.1)</td>
<td>68 (18.7)</td>
<td>22 (15.8)</td>
<td>176 (20.5)</td>
</tr>
<tr>
<td>Minor Morbidity</td>
<td>124</td>
<td>151</td>
<td>58</td>
<td>333</td>
</tr>
<tr>
<td>All complications</td>
<td>210 (58.8)</td>
<td>219 (60.3)</td>
<td>80 (57.6)</td>
<td>509 (59.3)</td>
</tr>
<tr>
<td>No complications</td>
<td>147</td>
<td>144</td>
<td>59</td>
<td>350 (40.7)</td>
</tr>
</tbody>
</table>

Major Complications is mortality plus major morbidity.

All complications is mortality plus major morbidity plus minor morbidity.

A patient with more than one complication is categorized according to the most major complication documented.

7.6 Feedback

The feedback given to all surgeons, and an example of feedback given to an individual surgeon, is shown in Appendix B. In addition to generating discrimination and goodness of fit statistics for each individual, calibration bar graphs were also produced. These were helpful as they demonstrated the accuracy of prediction in low, medium and high risk cases. The ratio of predicted to observed complications then identified (for all cases combined) if the surgeon was over predicting or under predicting. The combination of clinically relevant observations and statistically useful outcome data for both the individual and the overall group was appreciated.

The next two sections present the results (section 7.7) and discussion (section 7.8) for the questions: How accurate is surgical prediction? Can surgical prediction be improved?
### Table 7.7 Breakdown of major complications

<table>
<thead>
<tr>
<th>Complication</th>
<th>Phase I</th>
<th>Phase II</th>
<th>Phase III</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Myocardial infarction</td>
<td>10</td>
<td>4</td>
<td>5</td>
<td>19</td>
</tr>
<tr>
<td>Cardiac failure</td>
<td>3</td>
<td>6</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Arrhythmia</td>
<td>6</td>
<td>4</td>
<td>1</td>
<td>11</td>
</tr>
<tr>
<td>Respiratory failure</td>
<td>23</td>
<td>17</td>
<td>9</td>
<td>49</td>
</tr>
<tr>
<td>Respiratory distress syndrome</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Pneumonia</td>
<td>5</td>
<td>9</td>
<td>1</td>
<td>15</td>
</tr>
<tr>
<td>Renal failure</td>
<td>7</td>
<td>9</td>
<td>1</td>
<td>17</td>
</tr>
<tr>
<td>Urinary tract infection</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Anastomotic leak</td>
<td>9</td>
<td>9</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>Deep infection</td>
<td>5</td>
<td>3</td>
<td>1</td>
<td>9</td>
</tr>
<tr>
<td>Fistula</td>
<td>7</td>
<td>4</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Bowel obstruction</td>
<td>3</td>
<td>2</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>Gastrointestinal bleed</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Clostridium difficile</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Enterotomy</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Wrap migration</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Pancreatitis/fistula</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Stroke</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>8</td>
</tr>
<tr>
<td>Venous thrombosis</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Pulmonary embolus</td>
<td>1</td>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Major bleeding</td>
<td>6</td>
<td>6</td>
<td>3</td>
<td>15</td>
</tr>
<tr>
<td>Graft occlusion</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Vessel occlusion</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Graft infection</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Endoleak</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Fasciotomy</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Septicaemia</td>
<td>10</td>
<td>7</td>
<td>5</td>
<td>22</td>
</tr>
<tr>
<td>Multi-organ failure</td>
<td>3</td>
<td>8</td>
<td>2</td>
<td>13</td>
</tr>
<tr>
<td>Wound infection</td>
<td>2</td>
<td>2</td>
<td>0</td>
<td>4</td>
</tr>
<tr>
<td>Wound haematoma</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>118</td>
<td>103</td>
<td>43</td>
<td>264</td>
</tr>
</tbody>
</table>
7.7 Results: Accuracy of prediction of major complications by the surgeon

7.7.1 VAS scores with and without complications

The surgeon’s prediction (for all predictions made) discriminated between patients who did and did not develop a complication (Table 7.8). The difference in the median postoperative VAS scores for mortality was 38 mm, for major complications was 25mm and for any complication was 14mm.

Table 7.8 Differences in VAS scores in patients with and without complications

<table>
<thead>
<tr>
<th></th>
<th>Mortality</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Preoperative Scores</td>
<td>Postoperative Scores</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Mortality</td>
<td>No Mortality</td>
<td>Mortality</td>
</tr>
<tr>
<td>Median score (range)</td>
<td>61 (8-100)</td>
<td>24 (1-100)</td>
<td>61 (1-100)</td>
</tr>
<tr>
<td>p Value</td>
<td>P&lt;0.001</td>
<td></td>
<td>P&lt;0.001</td>
</tr>
</tbody>
</table>

|                      | Major Complication         |                      |                      |
|                      | Preoperative Scores        | Postoperative Scores |                      |
|                      | Complication               | No complication      | Complication         | No complication      |
| Number of predictions| 127                        | 886                  | 127                  | 884                  |
| Median (range)       | 40 (4-100)                 | 22 (1-91)            | 46 (4-100)           | 21 (0-90)            |
| p Value              | P<0.001                    |                      | P<0.001              |

|                      | All complications          |                      |                      |
|                      | Preoperative Scores        | Postoperative Scores |                      |
|                      | Complication               | No complication      | Complication         | No complication      |
| Number of predictions| 127                        | 886                  | 127                  | 884                  |
| Median (range)       | 30 (1-100)                 | 18 (2-65)            | 30 (0-100)           | 16 (1-78)            |
| p Value              | P<0.001                    |                      | P<0.001              |

The MWU test was used to compare scores of patients with and without complications.
7.7.2 Discrimination and goodness of fit of prediction of major complications

The discrimination of prediction, the goodness of fit statistics and the ratio of predicted to observed complications are summarised in Table 7.9. As major complications are more difficult to predict than mortality, a c statistic for the prediction of major complications of 0.80 shows good prediction and compares reasonably well to established models.

Table 7.9 Accuracy of prediction of major complications using the surgeon’s VAS prediction and the POSSUM score

<table>
<thead>
<tr>
<th></th>
<th>Global VAS Phase I</th>
<th>Multifactorial VAS Phase II</th>
<th>VAS After feedback Phase III</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preoperative VAS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c statistic</td>
<td>0.778 (0.73-0.82)</td>
<td>0.779 (0.73-0.83)</td>
<td>0.895 (0.825-0.97)</td>
</tr>
<tr>
<td>HL statistic</td>
<td>( \chi^2 = 11.04, p=0.199 )</td>
<td>( \chi^2 = 6.82, p=0.555 )</td>
<td>( \chi^2 = 10.68, p=0.221 )</td>
</tr>
<tr>
<td>P:O ratio</td>
<td>1.25</td>
<td>1.48</td>
<td>1.45</td>
</tr>
<tr>
<td><strong>Postoperative VAS</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c statistic</td>
<td>0.810 (0.77-0.85)</td>
<td>0.799 (0.75-0.85)</td>
<td>0.918 (0.86-0.98)</td>
</tr>
<tr>
<td>HL statistic</td>
<td>( \chi^2 = 5.48, p=0.705 )</td>
<td>( \chi^2 = 609, p=0.637 )</td>
<td>( \chi^2 = 4.38, p=0.882 )</td>
</tr>
<tr>
<td>P:O ratio</td>
<td>1.27</td>
<td>1.54</td>
<td>1.53</td>
</tr>
<tr>
<td><strong>POSSUM Morbidity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c statistic</td>
<td>0.760 (0.70-0.80)</td>
<td>0.722 (0.66-0.78)</td>
<td>0.791 (0.71-0.87)</td>
</tr>
<tr>
<td>HL statistic</td>
<td>( \chi^2 = 9.03, p=0.339 )</td>
<td>( \chi^2 = 22.95, p=0.003 )</td>
<td>( \chi^2 = 4.02, p=0.855 )</td>
</tr>
<tr>
<td>P:O ratio</td>
<td>1.81</td>
<td>2.24</td>
<td>2.38</td>
</tr>
<tr>
<td><strong>POSSUM Mortality</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c statistic</td>
<td>0.750 (0.70-0.80)</td>
<td>0.718 (0.66-0.78)</td>
<td>0.749 (0.68-0.83)</td>
</tr>
</tbody>
</table>

Discrimination of prediction as determined by the area under the ROC curve (c statistic).
The Goodness of Fit is calculated using the Hosmer Lemeshow (HL) statistic.
P:O is the ratio of predicted complications against the number of observed complications.

There was no change in the accuracy of prediction of complications between Phases I and II of the study. The discrimination for the POSSUM morbidity score for preoperative prediction for Phase I and II combined (with 95% confidence intervals) was 0.780 (0.746-0.810) and for postoperative prediction was 0.805 (0.772-0.84). There was however an improvement in the discrimination of the prediction of complications in Phase III of the study. In comparison the discrimination of the prediction made by the POSSUM morbidity score or the P-POSSUM mortality score did not change throughout the study. The similarity in the discrimination of the prediction of complications by the preoperative VAS, postoperative VAS, POSSUM morbidity and P-POSSUM mortality scores in Phase I of the study is illustrated by the very similar ROC curves in Figure 7.8.
Figure 7.8 ROC curves for Phase I of the study

Preoperative is preoperative VAS prediction, Postoperative is postoperative VAS prediction, POSMorbidity is POSSUM morbidity, POSMortality is P-POSSUM mortality,

Figure 7.9 ROC curves for Phase III of the study
In contrast to this, the differences in the discrimination of prediction in Phase III are contrasted by the separation of the ROC curves in Figure 7.9.

### 7.7.3 Calibration of prediction of complications against observed complications

The comparison of the predicted major complications versus observed major complications is shown in the calibration bar graph (derived from a calibration table) for the surgeons’ postoperative VAS prediction in Figure 7.10. The results for preoperative VAS prediction are very similar. The graph shows that surgeons overpredicted the rate of major complications in low risk and medium risk cases, but that the ability to predict high risk cases was excellent. This suggests that the surgical prediction of major complications may be improved if the surgeon decreases the estimation (the percentage score) of risk in low and medium risk cases. The 95% confidence intervals confirm that the overprediction of complications by the surgeon was statistically significant in deciles 1 to 8.

![Postoperative prediction graph](image)

**Figure 7.10** Calibration graph of surgeon’s postoperative VAS prediction of major complications

There was also a significant overprediction of major complications by POSSUM morbidity score (Figure 7.11). This applied to all categories of risk (low, medium and high risk cases). The overprediction by the POSSUM morbidity score was greater than the overprediction of complications by the surgeon. For example, the ratio of predicted to observed major complications for the preoperative VAS score was 1.34:1, for the postoperative VAS score
was 1.41:1 and for the POSSUM morbidity score was 2.05:1. The reason for this is presented in section 7.7.5.

![POSSUM Morbidity Prediction](image)

**Figure 7.11 Calibration graph of POSSUM morbidity score against major complications**

### 7.7.4 Mortality

The task given to the surgeon in this study was to predict major complications, which included both major morbidity and mortality. As there were 43 deaths in the 176 patients who had a major complication it was decided to assess if the prediction of major complications also had the ability to discriminate between (or to separate out) those who died against those who did not die. The overall discrimination of prediction for mortality for the preoperative VAS prediction, postoperative VAS prediction and the P-POSSUM mortality score were 0.835 (0.78-0.89), 0.836 (0.79-0.89) and 0.836 (0.79-0.855) respectively. These almost identical results which show very good discrimination confirm that P-POSSUM and the surgical prediction of major complications were able to predict those patients at the greatest risk of dying. For the P-POSSUM mortality score there was no evidence of a lack of fit with an HL statistic of $\chi^2 = 7.975$, 8 d.f, $p=0.436$. The calibration for P-POSSUM was accurate for low and medium risk patients, but over predicted mortality in the highest risk patients (Figure 7.12). The overall ratio of predicted mortality against observed mortality was 1.46:1.
7.7.5 Statistical issues pertaining to the comparison between surgical prediction and POSSUM

Although POSSUM provides a good general comparison to surgical prediction there are some important differences between prediction and POSSUM which mean that it is not possible to conclude that one is better than the other. Our primary goal, which was to assess the accuracy of the surgical prediction of major complications, meant that these differences were unavoidable. There are two main reasons for this (which also applies to other scoring systems).

Firstly, this study wanted to look at major events, or major complications, rather than all morbidity. This resulted in our definition of major complications (grades III, IV, V Clavien-Dindo classification complications) being different to the POSSUM definition of morbidity (which also included a range of grade I and II Clavien-Dindo classification complications). For example complications such as wound infection and urinary infection, which are included in POSSUM, were usually not severe enough to fall within our definition of a major complication. This is why the POSSUM morbidity score over predicted the number of major complications (section 7.7.3; Figure 7.11). To complete the analysis of how accurate the POSSUM morbidity score performed in predicting morbidity it is necessary to use the POSSUM definition of morbidity. Using these definitions 353 of 859 patients developed morbidity. The discrimination was only moderate [0.686 (0.650-0.722)], the goodness of fit
was very good (HL statistic $\chi^2 = 4.475$, 8 d.f, $p=0.812$), and the ratio of predicted to observed complications was very accurate with a ratio of 1.02:1. In comparison to this when the Clavien-Dindo classification of major complications was used the discrimination of the POSSUM morbidity score for the 859 cases was reasonable [0.754 (0.712-0.796)], and the goodness of fit was poor (HL statistic $\chi^2 = 14.471$, 8 d.f, $p=0.07$).

Secondly, in terms of prediction it is intuitively better to include mortality in the prediction of a major event, rather than to attempt to predict major complications excluding mortality. As POSSUM uses different equations to predict mortality and morbidity, and has no equation that combines these endpoints, it means a combined mortality and morbidity endpoint cannot be directly compared to POSSUM. To give a more complete comparison between POSSUM and surgical prediction it was therefore necessary to compare the VAS prediction against both the POSSUM morbidity score (sections 7.7.2 and 7.7.3) and the P-POSSUM mortality score (section 7.7.4). Surgical prediction was shown to have a comparable, and therefore very good, discrimination in both of these comparisons.

A third observation is that most patients who die also develop another complication. In this study 31 of 43 patients who died developed another major complication. These cases therefore should also have been captured by the POSSUM morbidity score. The POSSUM morbidity score can therefore be used to give a close, but not exact, comparison to the prediction of major complications.

7.7.6 Sub analysis looking at surgical speciality and the experience of the predictor

7.7.6.1 The impact of different surgical specialities

The discrimination of prediction was similar for the different types of surgery: with a postoperative c statistic of 0.794 (0.754-0.835) for colorectal/general surgery, 0.854 (0.798-0.909) for upper gastrointestinal surgery and 0.869 (0.810-0.929) for vascular surgery. The improvement in prediction in phase III was spread across all of the surgical subspecialties.

7.7.6.2 The impact of different levels of experience

Table 7.10 compares the discrimination of the surgeon’s prediction of major complications relative to their level of experience or seniority. The discrimination of prediction for all groups compared was remarkably similar.
Table 7.10 Discrimination of prediction for surgeons of different levels of experience

<table>
<thead>
<tr>
<th>Numbers</th>
<th>Most Senior Operator</th>
<th>Consultant Prediction</th>
<th>Registrar Prediction</th>
<th>Junior Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictions</td>
<td>859</td>
<td>615</td>
<td>565</td>
<td>115</td>
</tr>
<tr>
<td>Deaths</td>
<td>43</td>
<td>22</td>
<td>36</td>
<td>11</td>
</tr>
<tr>
<td>Complications</td>
<td>176</td>
<td>116</td>
<td>113 (77+36)</td>
<td>40 (29+11)</td>
</tr>
</tbody>
</table>

**Preoperative Prediction**

<table>
<thead>
<tr>
<th></th>
<th>Death</th>
<th>Complications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td>0.831 (0.76-0.90)</td>
<td>0.826 (0.75-0.91)</td>
</tr>
<tr>
<td>Complications</td>
<td>0.808 (0.77-0.84)</td>
<td>0.757 (0.71-0.85)</td>
</tr>
</tbody>
</table>

**Postoperative Prediction**

<table>
<thead>
<tr>
<th></th>
<th>Death</th>
<th>Complications</th>
</tr>
</thead>
<tbody>
<tr>
<td>Death</td>
<td>0.831 (0.77-0.90)</td>
<td>0.835 (0.75-0.92)</td>
</tr>
<tr>
<td>Complications</td>
<td>0.832 (0.80-0.87)</td>
<td>0.801 (0.76-0.85)</td>
</tr>
</tbody>
</table>

Most senior operator is the most senior surgeon scrubbed. Sometimes this was a registrar.
Registrar is a specialist registrar with at least two years of registrar experience
Junior prediction includes a junior registrar or a senior house officer
The discrimination of prediction as determined by the area under the ROC curve (c statistic) is presented with 95% confidence intervals.

**7.8 Discussion: Accuracy of prediction of major complications by the surgeon**

**7.8.1 Phase I: Initial assessment. Initial prediction ‘as good as’ POSSUM**

POSSUM was the earliest scoring system using multiple risk factors to become widely established (Copeland et al. 1991). This combines a 12-factor physiology score and a 6-factor operative severity score, using four grades of severity for each of these risk factors. Equations are then used to calculate the predicted rates of morbidity and mortality. At the time it represented a significant advance by providing a risk adjusted audit of complications between different surgical units and surgeons (Copeland et al. 1995). Advantages of POSSUM include its potential application across a broad range of elective and acute general surgical procedures, and its ability to be adapted for specific applications – including orthopaedic surgery (Copeland 2002), vascular surgery (Pythierech et al. 2001), colorectal surgery (Tekkis et al. 2004) and oesophago-gastric surgery (Tekkis et al. 2004). An adjustment of the equation was
also required to improve the prediction of mortality in low risk surgery, and is called the Portsmouth POSSUM or P-POSSUM score (Prytherech et al. 1998). POSSUM has been widely used in many different countries, and is still considered to be one of the best risk prediction models.

In Phases I and II of this study (Study Two) the performance of the surgeon’s VAS score in predicting major complications was comparable to the POSSUM morbidity score, in terms of both discrimination and goodness of fit. The ability of surgical prediction to perform as well as a well-established model which takes into account 18 risk factors is a very good result. A discrimination for preoperative prediction of 0.778 and for postoperative prediction of 0.810 is also comparable to the discrimination of other models.

7.8.2 Phase II: Structured assessment. No improvement in prediction.

In Phase II the objective was to test if a more considered clinical assessment, which takes into account the main domains of surgical risk, improved surgical prediction. The introduction of a multifactorial VAS made no difference to the accuracy of the surgeon’s prediction.

Although it can be concluded that there was no advantage in using a multifactorial VAS, the tool used (the multifactorial VAS) may not have successfully tested “if a more structured approach, which consciously takes into account the different components of risk,” can improve surgical prediction. This was suggested by the way in which the multifactorial VAS scales were completed, and is examined in section 8.3.

7.8.3 Phase III: Impact of feedback

In contrast, high quality outcome feedback did result in an improvement in surgical prediction. This improvement occurred even though there was no change in the discrimination of the POSSUM morbidity score or the P-POSSUM mortality score. As well as enabling us to benchmark the usefulness of clinical prediction, POSSUM was also used as a ‘control’ to determine if the improvement of prediction was due to better prediction rather than the more recent cases being easier to predict. This is an important finding, not only because of the immediate improvement in prediction, but because it also provides a potential mechanism for ongoing improvement. If the improvement in prediction had been due to the learning curve for using the VAS then we would have expected an improvement in prediction between Phases I and II of the study. The observation that there was no improvement between Phases I
and II therefore makes it more likely that the significant improvement in Phase III is related to the clinical feedback that was given (that is, the intervention at the end of Phase II).

There are examples in medicine where outcome feedback has resulted in an improvement in performance. The reduction in infection as a result of feedback of wound infection rates to surgeons was first reported by Cruse et al. in the 1970s in a meticulous study involving over 60,000 patients (Cruse and Foord 1973; Cruse and Foord 1980). Subsequent studies have suggested that wound infection surveillance with feedback to surgeons decreases the rate of infection by approximately 30% (Widmer and Battegay 2010; Hayley 1995). Perhaps the most comprehensive demonstration of the benefits of high quality feedback has been the National Surgical Quality Improvement Program (NSQIP) run by the American College of Surgeons. The program provides hospitals and surgeons with rigorous risk-adjusted clinical mortality and morbidity data, which includes benchmarking the performance of other hospitals (Ingraham et al. 2010). The initial impact in the Veteran Affairs Hospitals was a 27% reduction in postoperative mortality after major surgery and a 45% reduction in morbidity (Khuri et al. 2002). After expanding the program into private American hospitals, an analysis of 118 institutions demonstrated that 66% of hospitals improved their risk-adjusted mortality (mean O: E improvement: 0.174; P < 0.05) and 82% improved their risk adjusted complication rates (improvement: 0.114; P < 0.05) (Hall et al. 2009).

These studies have emphasized that a number of prerequisites need to be followed for effective feedback to be given. A surveillance program needs to be in place, with data being collected in an accurate, efficient and confidential manner. This requires written definitions of infection, regular clinical case-finding, post-discharge follow up for short-staying patients, computer storage of data, and analysis and reporting of the data in coded form that does not publicly identify individuals. The variation in the intrinsic risk of the different patients or hospitals being compared also needs to be controlled for. This requires time, money and commitment (Hayley 1995). Our study reflected a number of these principles. This included precise definitions, comparison of results with a risk-adjusted model, staff dedicated to identifying and documenting complications, and the feedback of clinically relevant and useful information.
7.8.4 Literature comparing POSSUM and surgical prediction

At the time of writing three other studies have compared surgical prediction, using a VAS scale, to POSSUM. The first (Hartley and Sagar 1994), used a postoperative VAS with three categories of risk. They compared the high risk category (which identified 24% of patients) against a POSSUM morbidity score of 50% or more (which identified 31% of patients). Surgical prediction was more sensitive and specific at identifying patients that went on to have complications. The second study assessed the accuracy of six scoring systems for elderly patients undergoing hip fracture surgery (Burgos et al. 2008). These included the preoperative prediction of an anaesthetist using a 10 scale VAS and the POSSUM morbidity score. The VAS was the best system for predicting major complications, with a discrimination of 0.83 compared to 0.73 for POSSUM, although there was an overlap of the 95% confidence interval. This study however did not include the six-factor Operative severity score in its POSSUM score. The third study (Markus et al. 2005) of patients undergoing upper abdominal surgery, assessed the accuracy of postoperative prediction using the surgeon’s gut feeling scored on a 100mm VAS with 10 anchors, and compared this to the POSSUM morbidity score. The observed morbidity rate was 29.5%. The predicted rates were 32% with VAS and 46% with POSSUM. In this study, which used the POSSUM definitions of morbidity, clinical prediction more closely matched the observed incidence of morbidity than the POSSUM morbidity score. Our study, by demonstrating that prediction had a comparable discrimination and goodness of fit to POSSUM, was consistent with and complemented these other studies. The improvement in prediction relative to POSSUM after clinical feedback was a new finding.

As the prediction of complications by the clinician is subjective, there is no guarantee that different groups of surgeons or clinicians will predict with equal accuracy. However on the four occasions where surgical prediction of major complications using a VAS has been compared to POSSUM, with different doctors, in different countries and in different specialties, the performance of the VAS score has been comparable to POSSUM. These studies suggest that clinical prediction, when well used, is as useful as multifactorial models for predicting surgical risk. The potential advantages, weaknesses and possible applications of clinical prediction using a VAS are addressed in section 7.11.

7.9 Implications of the change in prediction before and after surgery

This section addresses the question, what is the importance of the surgeon changing the risk of complications postoperatively?
7.9.1 Previous work assessing a change in VAS score

Change in prediction had previously been assessed in three studies. The first, by Pettigrew et al. (1987), noted that 14 of 25 (56%) patients with an increased prediction of complications developed a major complication. The main reason given for increasing the VAS score was the technical difficulty of the procedure. They concluded that operative performance was the main factor in the development of postoperative complications. The other two studies are described in this thesis. One assessed major complications (Study One, section 7.2.1) and the other assessed wound infection (Chapter 4). The wound infection study showed that there was a significant increase in wound infection when a patient with a low risk of infection preoperatively had the risk increased postoperatively. The main reason given for increasing the risk was the extent of contamination. The major complication study also demonstrated that an increase in the prediction of a major complication was associated with significantly more complications, especially in patients with a low preoperative risk. The majority of reasons given for increasing the risk of prediction were technical in nature. However the contribution of the surgeon to these technical events (or the inevitability of them) was not identified. The issues that were addressed in this study (the Leeds study, study two) was whether these results were repeatable, and whether the increase in prediction of complications due to ‘technical reasons’ were inherent to the difficulty of the situation as it presented at the time of surgery, or if this was contributed to by technical performance. The methods are presented in section 7.4.

7.9.2 Results

Of the 1295 risk predictions, 490 (37.8%) had a change in the prediction of a major complication. There was an increase in risk in 262/1295 (20.2%) and a decrease in risk in 228/1295 (17.6%). This was similar to an increase of 17.2% and a decrease of 12.7% in study one. The significance of the surgeon changing his/her prediction of a major complication postoperatively was confirmed by using the Wilcoxin signed rank test. For patients with no complications there were more negative changes in ranking than positive changes, with a reduction in mean ranking giving a Z value of -4.802, p<0.001. For patients with major complications there were more positive than negative changes in ranking, with an increase in mean ranking, giving a Z value of -4.562, p<0.001. The postoperative ranking of patients was therefore significantly more accurate than the preoperative ranking.
The association between these changes in risk assessment and subsequent major complications is illustrated in Figure 7.13.

Low risk: up to and including the median score of 25mm,
High risk: above the median score of 25mm. The \( \chi^2 \) test was used.

**Figure 7.13 Flow chart for major complications, illustrating changes in the VAS score**

The main finding was in the high risk preoperative patients, with an increase in prediction being associated with a doubling of the rate of major complications, and a decrease being associated with a significant decrease in the rate of major complications. In the Dunedin study an increase in risk was also associated with a significant increase in complications; however this increase was mainly in the preoperative low risk patients.

The reasons given for changing the VAS prediction are shown in Table 7.11. Of the 434 reasons given 54.5% were technical, 38% disease related and 7.4% anaesthetic or medical.
The most common reasons for increased technical problems were the overall difficulty of surgery, the magnitude of the operation, bleeding and anastomotic issues. For disease related issues these included the extent of contamination, the extent of malignancy and more severe vascular disease than expected. The main medical problem causing an unexpected increase in risk was cardiac disease. The main reasons for decreasing the VAS were the operation going well, performing a lesser operation than expected, sepsis being less than predicted and the tumour being less advanced than expected. The technical and pathological reasons given were similar to study one. The main difference was a greater proportion of reasons being related to pathology, including most cases of contamination being due to pathology rather than technical problems. More medical reasons, including cardiac events and anaesthetic difficulties were also identified. In the 18.6% of cases when there was a change in score no reason was documented. In these cases the magnitude of the change was less than when a reason was stated (p<0.001) and was more likely to be in a downward rather than an upward direction (p<0.001).

To help to decide if a change in the VAS prediction, and the reasons given for changing the VAS score, were due to the findings at surgery or related to surgical technique the surgeon was asked to identify if the change in prediction was unavoidable, related to technical proficiency or due to surgical error (Figure 7.5). This was categorized in 423 cases, and was unavoidable in 222 cases (53%), related to proficiency in 166 cases (39%), due to error in 13 cases (3%) and placed in greater than one category on 22 occasions (5%). In Table 7.12 the written reasons for changing the risk of complications are compared to the categorical answers (unavoidable, technical proficiency, error) given by the surgeon. There was a reasonable match between the categorical answers and the written reasons given for changing the VAS prediction. For ‘unavoidable cases’ most of the stated reasons were related to pathology, for ‘technical proficiency cases’ most of the stated reasons were technical in nature and for ‘error cases’ the reasons given were technical in nature (with sometimes more than one reason being given).

However it was difficult to be certain about the relationship between technical performance and a change in prediction. This was because the information received was too imprecise. For example when a technical reason was written, 5% were categorized as error, 72% were categorized as technical proficiency, 21% were categorized as being unavoidable and 2% were categorized as technical and unavoidable! The high proportion of cases (79%)
categorized as least in part as technical does highlight the importance of surgical technique. However what proportion of the technical proficiency category could have been ‘improved on’, or to what extent a lack of technical proficiency may have contributed to complications remained uncertain.

Table 7.11 Reasons given for changing the VAS score for major complications

<table>
<thead>
<tr>
<th>Category</th>
<th>Increase in risk</th>
<th>Decrease in risk</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change in risk</td>
<td>262</td>
<td>228</td>
<td>490</td>
</tr>
<tr>
<td>Reason stated</td>
<td>234</td>
<td>165</td>
<td>399</td>
</tr>
<tr>
<td>Reason not stated</td>
<td>28</td>
<td>63</td>
<td>91</td>
</tr>
<tr>
<td>When a reason is given….</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt;1 reason given</td>
<td>20</td>
<td>15</td>
<td>35</td>
</tr>
</tbody>
</table>

| TECHNICAL                                     |                  |                  |       |
| Duration of surgery                           | 8                | 0                | 8     |
| Difficulty of surgery                         | 37               | 92               | 129   |
| Extent of operation                           | 21               | 16               | 37    |
| Anastomotic                                   | 17               | 0                | 17    |
| Gastrointestinal and biliary contamination    | 7                | 0                | 7     |
| Bleeding                                      | 23               | 2                | 25    |
| Injury to other organs                        | 6                | 0                | 6     |
| Decision making                               | 0                | 3                | 3     |
| Other                                         | 2                | 3                | 5     |

| PATHOLOGY                                     |                  |                  |       |
| Contamination/sepsis                          | 31               | 15               | 46    |
| Gastrointestinal ischemia                     | 12               | 1                | 13    |
| Malignancy                                    | 24               | 16               | 40    |
| Adhesions                                     | 15               | 5                | 20    |
| Vascular                                      | 11               | 4                | 15    |
| Anatomy                                       | 2                | 2                | 4     |
| Other                                         | 22               | 5                | 27    |

| MEDICAL                                        |                  |                  |       |
| Cardiac                                       | 7                | 5                | 12    |
Table 7.12 Relationship between categorical classification and the written reason for a change in risk assessment

<table>
<thead>
<tr>
<th>Category</th>
<th>Pathological reason</th>
<th>Technical reason</th>
<th>Medical reason</th>
<th>More than one reason</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unavoidable</td>
<td>154</td>
<td>44</td>
<td>13</td>
<td>11</td>
</tr>
<tr>
<td>Technical Proficiency</td>
<td>2</td>
<td>150</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Unavoidable &amp; technical</td>
<td>2</td>
<td>4</td>
<td>0</td>
<td>16</td>
</tr>
<tr>
<td>Error</td>
<td>0</td>
<td>10</td>
<td>0</td>
<td>3</td>
</tr>
</tbody>
</table>

A second and important reason for caution in interpreting these results was that some surgeons were uncomfortable about using the word error. Our impression that some errors were placed in the category of technical proficiency, resulting in the ‘error category’ being under reported was supported by our observation that the majority of errors were reported by the most ‘technically excellent’ surgeons.

7.9.3 Discussion
7.9.3.1 Comparison of results with previous study
Although there was some variation in results, the findings of this study supported the general observations and trends noted in Study One. The importance of the surgeon changing the prediction of major complications was again confirmed. In particular, an increase in score was able to identify a high risk group of patients who developed twice the number of major complications. The reasons given for changing the score were also very similar, with the same technical and pathological issues being highlighted. There were some differences. In study two a decrease in the risk in preoperative high risk patients was associated with a significant decrease in complications. Also in Study Two a greater proportion of the reasons given for changing prediction were related to pathology and medical issues.
New information obtained in this study was the finding that in 53% of cases the surgeon considered the change in risk to be unavoidable (mainly related to pathology), in 39% to be related to technical proficiency and in 3% to error. Although this study does confirm that what the surgeon sees and does intraoperatively is important, it also indicates that a greater contribution to the change in risk was caused by ‘unavoidable’ factors, or factors beyond the technical skills of the surgeon, especially the surgical pathology, and to a lesser extent to changes in medical co-morbidity.

7.9.3.2 Potential applications
Although this study has not been able to clearly define the role of error or technical performance, it has demonstrated how a change in VAS score can identify important events and collect useful data about these events. One benefit is the ability to identify a group of high risk patients. While it is almost certain that intraoperative interventions are happening to negate problems as they are identified by the surgical team, the question remains whether this process can be optimized. In section 4.4.7 a question was asked about increasing intraoperative prophylaxis and at which threshold should an extra dose of prophylactic antibiotic be given or therapeutic antibiotics commenced? In terms of postoperative intervention, the identification of a high risk group of patients who could benefit from more intensive management postoperatively is also an area that could be investigated further.

Another possible application for a change in VAS score may be in the area of highlighting technical issues in surgical training. For example if a VAS form was completed independently by both surgeon and trainee, a change in the prediction of complications would identify significant technical events, and provide opportunity to explore reasons why this may have happened.

7.10 Inclusion of VAS in a model for predicting complications
7.10.1 Introduction
Section 7.10 addresses the question; can surgical prediction be used to improve a model assessing the risk of complications? This had been demonstrated in Study One. We wanted to test the reliability of this finding.
7.10.2 Methods

7.10.2.1 Model design

To examine the impact of the surgeon’s VAS score on a model, it is necessary to examine the interactions between the risk factors used in the model and the VAS score. The best way to achieve this was to develop a risk model from the data collected during the risk prediction study, and then to assess the impact of the VAS score on this model.

In terms of model development I wanted to test the feasibility of developing a simple model for the prediction of major complications, which could be used in a variety of clinical settings. I was also interested in assessing whether a preoperative model which could help with clinical decision making was feasible. The steps used in model development were:

1) Literature search with an emphasis on the selection and performance of risk factors in different models.
2) Selection of risk factors which would cover the main domains of risk.
3) Statistical analysis using patient data collected in the Leeds study (Study two).
4) Repeating the statistical analysis with the addition of the surgeon’s VAS score.

7.10.2.2 Statistical analysis

Logistic regression analysis was performed using SAS version 6.12 (SAS Institute Inc., 100 SAS Campus Drive, Cary, North Carolina, USA). The analysis was performed for both major complications and mortality. The selected risk factors were entered in appropriately stratified categories of risk (Table 7.13). The univariate odds ratio and the adjusted odds ratio were reported. A correlation matrix was used to check for possible interactions between the variables included in the logistic regression analysis, and the presence of multi-collinearity was checked using variance inflation factors. Logistic regression analysis for major complications was then repeated using the preoperative and postoperative VAS scores.

7.10.3 Results

7.10.3.1 Selection of risk factors

For each of the main areas contributing to the risk of complications two variables which identified risk and which were readily available were selected. As the main purpose of this section is to examine the impact of the VAS score on a model, the process of the selection of the individual risk factors has not been discussed.

The risk factors selected were:
a) Functional status and cardiovascular and pulmonary morbidity: ASA classification of physical status and age.

b) Severity of pathology: for infection - albumin and white cell count; for cancer - the presence of metastatic disease.

c) Urgency of surgery: timing/urgency of surgery and the number of operations in the last 30 days (which also overlaps with the complexity of surgery).

d) Complexity of surgery: category of surgery (complexity or magnitude of the operation) and the duration of surgery.

e) Other: No risk factor was chosen in the general model, although the surgeon’s VAS prediction may help to identify risk factors in this area.

7.10.3.2 Logistic regression analysis

7.10.3.2.1 Major complications

Results for logistic regression analysis are presented in Table 7.13.

Table 7.13 Summary of univariate odds ratios and adjusted odds ratios for major complications using logistic regression analysis

<table>
<thead>
<tr>
<th>Variable</th>
<th>Categorisation</th>
<th>Number (% +ve)</th>
<th>Univariate OR</th>
<th>Adjusted OR</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>VAS excluded</td>
</tr>
<tr>
<td>Age</td>
<td>&lt;56</td>
<td>302 (14.9)</td>
<td>1</td>
<td>0.94 (0.57-1.57)</td>
</tr>
<tr>
<td></td>
<td>56-75</td>
<td>359 (22.6)</td>
<td>1.66 (1.11-2.49)</td>
<td>0.94 (0.57-1.57)</td>
</tr>
<tr>
<td></td>
<td>&gt;75</td>
<td>198 (25.3)</td>
<td>1.93 (1.23-3.03)</td>
<td>1.03 (0.57-1.86)</td>
</tr>
<tr>
<td>ASA</td>
<td>I</td>
<td>145 (9.0)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>II</td>
<td>384 (12.8)</td>
<td>1.48 (0.78-2.84)</td>
<td>1.13 (0.56-2.30)</td>
</tr>
<tr>
<td></td>
<td>III</td>
<td>256 (26.2)</td>
<td>3.60 (1.91-6.79)</td>
<td>2.41 (1.15-5.03)</td>
</tr>
<tr>
<td></td>
<td>IV/V</td>
<td>74 (63.5)</td>
<td>17.67 (8.43-37.1)</td>
<td>7.77 (3.09-19.6)</td>
</tr>
<tr>
<td>Albumin</td>
<td>&gt;35</td>
<td>736 (15.8)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>25-34</td>
<td>90 (38.9)</td>
<td>3.40 (2.12-5.41)</td>
<td>1.60 (0.85-3.01)</td>
</tr>
<tr>
<td></td>
<td>&lt;25</td>
<td>33 (75.8)</td>
<td>16.7 (7.66-40.4)</td>
<td>6.06 (2.04-18.0)</td>
</tr>
<tr>
<td>----------------------</td>
<td>-----------</td>
<td>-----------</td>
<td>------------------</td>
<td>------------------</td>
</tr>
<tr>
<td>WCC</td>
<td>3-20</td>
<td>826 (19.7)</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&lt;3, &gt;20</td>
<td>33 (39.4)</td>
<td>2.64 (1.29-5.43)</td>
<td>0.866 (0.31-2.41)</td>
</tr>
<tr>
<td>Cancer</td>
<td>None, local or in LN</td>
<td>806 (19.9)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Recurrent or metastatic</td>
<td>53 (30.2)</td>
<td>1.75 (0.95-3.22)</td>
<td>1.77 (0.87-3.61)</td>
</tr>
<tr>
<td>Urgency</td>
<td>Elective</td>
<td>285 (10.9)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Urgent</td>
<td>360 (18.6)</td>
<td>1.87 (1.19-2.96)</td>
<td>0.942 (0.56-1.58)</td>
</tr>
<tr>
<td></td>
<td>Acute</td>
<td>197 (31.5)</td>
<td>3.76 (2.33-6.07)</td>
<td>2.18 (1.16-4.06)</td>
</tr>
<tr>
<td></td>
<td>Emergency</td>
<td>17 (94.1)</td>
<td>131.09 (16.8-999)</td>
<td>36.70 (3.2-413.9)</td>
</tr>
<tr>
<td>Number of Operations</td>
<td>1</td>
<td>819 (18.4)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>&gt;1</td>
<td>40 (62.5)</td>
<td>7.37 (3.84-14.6)</td>
<td>2.31 (0.93-5.75)</td>
</tr>
<tr>
<td>Complexity of surgery</td>
<td>Minor, Intermediate</td>
<td>56 (5.4)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Major</td>
<td>534 (11.3)</td>
<td>4.22 (1.52-17.6)</td>
<td>2.0 (0.52-7.69)</td>
</tr>
<tr>
<td></td>
<td>Major complex</td>
<td>269 (26.0)</td>
<td>6.21 (2.20-26.1)</td>
<td>2.79 (0.68-11.5)</td>
</tr>
<tr>
<td>Duration</td>
<td>0-89</td>
<td>227 (9.7)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>90-179</td>
<td>383 (19.5)</td>
<td>2.12 (1.27-3.53)</td>
<td>2.67 (1.37-5.23)</td>
</tr>
<tr>
<td></td>
<td>&gt;180</td>
<td>249 (33.3)</td>
<td>4.66 (2.79-7.78)</td>
<td>8.28 (4.0-17.16)</td>
</tr>
<tr>
<td>VAS</td>
<td>0-19</td>
<td>355 (5.4)</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>20-39</td>
<td>335 (20.0)</td>
<td>4.42 (2.59-7.54)</td>
<td>3.02 (1.70-5.38)</td>
</tr>
<tr>
<td></td>
<td>40-59</td>
<td>98 (33.7)</td>
<td>8.98 (4.8-16.75)</td>
<td>3.88 (1.88-8.00)</td>
</tr>
<tr>
<td></td>
<td>60 and up</td>
<td>71 (80.0)</td>
<td>71.97</td>
<td>20.18</td>
</tr>
</tbody>
</table>
Summary

<table>
<thead>
<tr>
<th></th>
<th>(34.2-151.6)</th>
<th>(7.9-51.3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>c statistic</td>
<td>0.818</td>
<td>0.847</td>
</tr>
<tr>
<td>HL statistic</td>
<td>0.333</td>
<td>0.998</td>
</tr>
</tbody>
</table>

The preoperative VAS score was used.

This is for 859 patients, with 176 documented major complications. All variance inflation factors were less than 2, indicating that there were no significant interactions between the different variables in the model. The discrimination for the model was 0.818. The addition of the preoperative VAS significantly improved the model, with the discrimination improving to 0.847 ($\chi^2$ improvement of 46.5, p<0.0001), and with an improvement in the goodness of the fit also being demonstrated.

Table 7.14 summarises discrimination and goodness of fit statistics for different options.

**Table 7.14 Model results for different options**

<table>
<thead>
<tr>
<th>Model</th>
<th>Area under ROC curve</th>
<th>Hosmer and Lemeshow test</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAS prediction only</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preoperative VAS</td>
<td>0.798</td>
<td>0.993</td>
</tr>
<tr>
<td>Postoperative VAS</td>
<td>0.822</td>
<td>0.181</td>
</tr>
<tr>
<td>Model with and without VAS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model only</td>
<td>0.818</td>
<td>0.333</td>
</tr>
<tr>
<td>Model + Preoperative VAS</td>
<td>0.847</td>
<td>0.998</td>
</tr>
<tr>
<td>Model + Postoperative VAS</td>
<td>0.857</td>
<td>0.875</td>
</tr>
<tr>
<td>Stepwise Entry</td>
<td>0.839</td>
<td>0.944</td>
</tr>
<tr>
<td>Preoperative prediction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model – Duration of surgery</td>
<td>0.783</td>
<td>0.080</td>
</tr>
<tr>
<td>Model – Duration of surgery + Preoperative VAS</td>
<td>0.824</td>
<td>0.722</td>
</tr>
</tbody>
</table>

Using a stepwise entry for logistic regression analysis four risk factors were identified. In order of importance these were the VAS score, duration of surgery, albumin and ASA. The discrimination was 0.839, and the HL goodness of fit excellent at 0.944. These cover the domains of complexity of surgery, sepsis, medical comorbidities and ‘other’. In terms of
thinking about the possibility of a preoperative model the only risk factor not available preoperatively was the duration of surgery. Although the model was not as good when duration of surgery was removed, when the preoperative VAS was included both the discrimination (0.824) and goodness of fit were good. This supports the concept of developing a preoperative model which includes a preoperative VAS score.

7.10.3.2.2 Mortality
Using the nine risk factors outlined above, the model had an excellent discrimination of 0.905, and a goodness of fit of 0.8477. Using stepwise logistic regression analysis, the four most important risk factors for predicting mortality were the ASA score, albumin, cancer and age (c statistic 0.895, H-L statistic p=0.780). The surgeon’s VAS score was not included as the surgeon was not asked to predict mortality.

The univariate odds ratios for ASA were: I -1, II-0.94 (0.18-4.92), III-3.52 (0.78-15.94), IV/V-34.32 (7.83-150.5). The odds ratios for albumin were: albumin >35 – 1, albumin 25-34 – 1.76, albumin <25 7.79 (2.11-28.7). The odds ratios for age were: 0-55 1, age 56-75 1.71 (0.68-4.29) and for age >75 5.27 (2.21-12.58). The odds ratios for cancer were: none, local or lymph node only 1, recurrent or metastatic disease 2.65 (1.07-6.60). Factors that were not helpful in predicting mortality were the white cell count, duration of surgery and complexity of surgery. These results were very similar to the NSQIP results where the six main predictors of mortality (in order) were ASA, functional health status, age, sepsis, disseminated cancer and albumin.

7.10.3.2.3 Limitations of logistic regression analysis
Although it has been demonstrated that surgical prediction using a VAS improves a model of risk prediction, a much larger sample size is needed for a robust model to be developed. One problem related to the small sample size is that bootstrapping within the study, or splitting the database for reasons of internal validation, was not possible. A second problem was that there were insufficient cases to assess the predictive ability of all the selected risk factors, such as the number of operations.

7.10.3.2.4 Summary of findings for different risk factors
Most of the selected variables were significant in one of the two models. Preoperative VAS (morbidity), duration of surgery (morbidity), albumin (both), ASA (both), age (mortality) and
cancer (mortality). Approaching significance, but limited by small numbers in the high risk categories, were the number of operations and emergency surgery. The category complexity of surgery was not significant, but this is an important factor in many larger models. White cell count was not significant and should be removed.

7.10.4 Discussion

7.10.4.1 Why does the VAS score improve risk models for predicting major complications?
The surgeon’s prediction of major complications significantly improved our model for predicting major complications, both in terms of discrimination and goodness of fit. At stepwise entry the VAS score was also the strongest individual predictor of major complications. These results confirm the findings of study one.

There are probably four reasons why the prediction of complications by the surgeon can improve a model for predicting complications. The first is that surgical prediction in this study was the strongest predictor of complications. The more accurate the risk factors, the better the model. Secondly, clinical prediction allows the surgeon to apply his/her experience and understanding to the specific clinical scenario, and to integrate the information available in a way that may transcend the predictive ability of the individual risk factors. Thirdly, surgical prediction may incorporate a number of ‘difficult to define’ variables: such as the surgeon’s experience, a global assessment of the patient and the quality of hospital care.

The fourth, and perhaps the main, reason is a statistical one. Risk models are based on variables that most patients have a ‘meaningful score’ for. Although these risk factors may not be critical to an individual patient, over a population their importance is reinforced. Examples include ASA, age and the urgency of surgery. In contrast, important but low frequency clinical events, such as liver cirrhosis, morbid obesity, aortic valve stenosis or a recent myocardial infarction, which are important risk factors for the individual patient, are usually not included in established models as they occur at ‘too low’ a frequency to be statistically significant. This results in important risk factors which occur at a low frequency not being captured by risk models. The surgeon’s risk assessment can identify and appropriately weigh these critical but infrequent clinical events. This can therefore add greater flexibility and accuracy to risk models.
7.10.4.2 Observations about models

7.10.4.2.1 Simplicity

The most accurate and widely used models, such as POSSUM, NSQIP and APACHE, all use multiple risk factors. Many of the risk factors in these models ‘overlap’. That is, more than one risk factor may measure (at least in part) the same underlying problem. This can result in some significant risk factors making a very small contribution to the model. It has been demonstrated that models can be simplified and still perform well. For example the NSQIP has assessed simplifying its data collection and has demonstrated that the number of risk factors can be reduced with only a small reduction in the accuracy of their models (Dimick et al. 2010).

An alternative approach, which was taken in this study, was to identify the main domains of risk, and to select one or two risk factors for each domain. Such a model may be less complex and still give good prediction. The challenge is to have a sufficiently simple model which does not lose too much discrimination. An example is the Surgical Risk Scale (Sutton et al. 2002). This model uses three risk factors: the ASA score, the timing of surgery and the grade or complexity of the operation for predicting mortality. Although there is a small drop in the discrimination of prediction when compared to P-POSSUM (Brooks et al. 2005; Neary et al. 2007; Tran ba Loc et al. 2010), the surgical risk scale performs remarkably well. In this study I have demonstrated that it is possible to develop a simple model, using nine risk factors, for the prediction of major complications. Further work would require a larger population, adjustment of the risk factors, development of the use of the VAS score, and the generation of predictive equations.

7.10.4.2.2 Differences in the prediction of mortality and morbidity

Some models, including POSSUM and the NSQIP, score both mortality and major complications. POSSUM assesses the risk of mortality and morbidity using the same risk factors, with the same categories of severity, but gives different results for each endpoint by adjusting the final equations. The NSQIP assesses mortality and morbidity separately, with the risk factors being determined independently for each endpoint. Our study confirmed that different risk factors are important for predicting mortality and morbidity.

Two risk factors in our study which predicted mortality but not major complications were age and cancer status. The importance of age and cancer status in predicting mortality is also
emphasized by their inclusion in the majority of models which predict mortality, such as the ACPGBI colorectal cancer model and colorectal POSSUM. The difference in endpoints that predict mortality and morbidity can be illustrated by age. Although age is a minor factor when it comes to developing a complication, once you have a complication your reserve for managing it reduces as you get older, resulting in age being predictive of mortality. In patients aged older than 75 it has been shown that there is an increase in perioperative morbidity of 1.2 to 2 times, but the mortality rate increases by 2.9 to 6.7 times, even after adjusting for differences in comorbidities (Bentrem et al. 2009). One factor that contributes to age ‘predicting’ death is the association of age with major (and sometimes unrecognized) cardiac disease (Bentrem et al. 2009). In our study cardiac disease contributed to 8 of 43 deaths.

7.10.5 Summary on models and the surgeon’s VAS score
This study demonstrated that it is possible to develop a relatively simple model for predicting major complications and that the surgeon’s prediction of complications using a VAS can significantly improve the prediction of the model. The combination of objective and subjective data can make a model more accurate and may also make it more applicable to the individual patient. It was also noted the that the risk factors for morbidity and mortality are different and that models that look at predicting both morbidity and mortality should assess the risk factors for these endpoints independently. However a much larger sample size is needed to validate this and develop a more robust model.

7.11 Potential applications of surgical prediction using a VAS score
In this section the strengths and limitations of surgical prediction using a VAS score are explored. This is followed by an assessment of the results of this study to help address the question of whether surgical prediction of major complications can be used as a tool to help influence decision making.

7.11.1 Advantages of surgical prediction using a VAS score
7.11.1.1 Versatility of VAS prediction when compared to other risk models
All models for predicting mortality and morbidity are developed for use in a specific setting. This may be in a certain place such as in the intensive care unit (APACHE), or for a specific type of surgery, such as all general surgery (POSSUM, SRS) or for colorectal surgery (colorectal POSSUM), or more specifically still for a specific disease process such as colorectal cancer (ACPGBI model). As the performance of the model falls away when it is
used in a different setting, there are advantages of having a scoring system that can be applied in a variety of settings. A scoring system based on the prediction of the surgeon, which can be adjusted according to the assessment and experience of the surgeon, has the advantage of being uniquely versatile. This was confirmed in our comparison of the VAS prediction with a number of scoring systems. These included POSSUM, colorectal POSSUM (Tekkis et al. 2004), the ACPGBI colorectal cancer model (Tekkis et al. 2002), the multifactorial index of preoperative risk factors in colorectal surgery (Ondrula et al. 1992) and the Preoperative Risk Assessment in elective surgery (Klotz et al. 1996). The details of this analysis have not been included in the thesis. However we can report that the surgeon’s VAS score showed credible versatility in that it was able to perform as well as a range of risk models designed for different settings.

7.11.1.2 Importance of major complications
Most risk models predict mortality rather than major complications (for example ACPGBI model, colorectal POSSUM, APACHE). Mortality, when appropriately risk adjusted, has been shown to be a reliable endpoint for comparing the performance of different institutions. It is therefore good for comparative audit. However, it may be that major complications are ‘under rated’ as an endpoint to be used in risk models. Major complications occur more frequently, can be identified earlier in the process of care (potentially allowing earlier intervention), impact on the quality of life and are of great importance to the patient. The VAS score is one way of measuring the risk of major complications.

7.11.1.3 The ability of the VAS score to adapt to changes in medical practice
The finding that feedback can improve prediction emphasizes the ability of the VAS score to improve and adapt. One criticism of POSSUM is that it has not been adjusted for improvements in practice over the last 20 years and this may be one reason why the model tends to over predict mortality and morbidity (Chandra et al. 2009). Whereas a defined scoring system cannot adapt and improve, the evidence of this study suggests that the surgeon, and therefore prediction using the VAS score, can.

7.11.1.4 Prediction of complications can be used by a range of surgeons
Two observations in this study support the wide applicability of surgical prediction using a VAS. The first is that 58 clinicians made predictions. The second is that the accuracy of prediction was similar for consultants, specialist registrars and even for some more junior
doctors (Table 7.13). One reason for this is that the predictions were made for operations appropriate to the doctor’s level of experience. The accuracy of prediction with such a diverse group of contributors was an encouraging finding, and suggests that the results may be generalised to other groups of doctors.

7.11.2 Limitations of surgical prediction using a VAS score
Whereas the flexibility and adjustability of the VAS is a strength, it is also an inherent weakness. The surgeon’s judgement, with its potential advantages, also introduces an element of subjectivity and unpredictability to the assessment of risk. Even when the prediction is excellent in one setting, it does not mean that this will always be the case. Three areas which are discussed, and which all require further work are: the way in which the surgeon uses the VAS score, the repeatability (reliability) of prediction, and if the VAS score measures what it is intended to measure (validity).

7.11.2.1 Variable use of the visual analogue scale
The variable use of the visual analogue scale by different surgeons is a major concern, and is an important reason that explains, at least in part, the differences in the results of this study compared to Study One. Issues pertaining to the use of the VAS are examined in section 8.2.

7.11.2.2 Reliability
Reliability refers to the consistency or repeatability of prediction. Reliability is a central concern when it comes to the prediction of risk as there will always be some variation when different surgeons use clinical judgement to predict complications. A major limitation of this study is that no formal reliability testing has yet been performed.

In general terms the excellent level of prediction in this study, where the prediction technique using a VAS has been repeated over 1000 times, suggests that the reliability of predictions for this group of surgeons was reasonable. The ‘general reliability’ of surgical prediction is also supported by good prediction in the studies using a labelled or anchored VAS, which also report a favourable performance when compared to POSSUM (section 7.8.4).

However to properly establish the reliability of surgical prediction of major complications using a VAS would require the whole study to be repeated a number of times, and/or specific studies to be set up to assess different aspects of reliability. Intra-assessor reliability looks at
how consistently an individual predicts risk. Given the same clinical problem does he make the same assessment? Cognitive feedback (Chapter 8) gives us a mechanism for exploring this question. Inter-assessor reliability looks at how consistently different surgeons predict the risk of complications for the same patient/same clinical scenario. To examine this question the performance of different surgeons given the same case vignettes needs to be studied.

7.11.2.3 Validity
Validity is related to whether a tool, in this case the prediction of major complications using a VAS, actually measures what it is meant to measure. The accuracy of the prediction suggests that the validity (including the predictive validity and the construct validity) of the VAS is very good. However, one area where further work would be useful would be to assess the best subscales to use in the multifactorial VAS. Although no modifications of the multifactorial VAS were recommended by any of the participating surgeons, a formal study has not been performed. An approach that could lead to improvements in the design of the multifactorial VAS involved interviewing surgeons using a repertory grid technique (Fransella and Bannister 1977) to identify the main themes used by surgeons in the prediction of complications. The content validity of each axis (which looks at whether each axis measures what it should measure, or if there is a significant overlap between axes), can also be assessed by using collinearity statistics. This is examined in section 8.4.2.2 and in Table 8.3 where it was demonstrated that there was good content validity for the group of surgeons who scored each of the subscales of risk analytically.

7.11.2.4 Feedback
Although outcome feedback has been shown to improve the prediction of complications, the ongoing ability of feedback to improve prediction has not been assessed. Would further outcome feedback improve prediction? Would the addition of cognitive feedback (section 8.4.3) further improve prediction?

7.11.3 Possible applications
Is there a role for using the surgeon's prediction of major complications, a 'subjective' score, as a tool either to assist clinical decision making or to determine a certain course of action?
7.11.3.1 Using the VAS alone

The potential for using the VAS score as a ‘test’ for a certain course of action can be assessed by converting our prediction of complications into likelihood ratios (Table 7.15). The best cut off point would vary according to the primary objective of the task. If the objective was to identify high risk patients, then a score of >50mm would identify a group of patients with a likelihood ratio of 8.3, or a positive predictive value (PPV) of a major complication of approximately 70%. High risk patients could be recommended for a range of interventions. Some potential applications are suggested.

Table 7.15 Likelihood ratios for surgical prediction of major complications

<table>
<thead>
<tr>
<th>Preoperative Prediction</th>
<th>Postoperative Prediction</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>Sens</td>
</tr>
<tr>
<td>20mm</td>
<td>1.6</td>
</tr>
<tr>
<td>30mm</td>
<td>2.4</td>
</tr>
<tr>
<td>40mm</td>
<td>4.2</td>
</tr>
<tr>
<td>50mm</td>
<td>8.2</td>
</tr>
<tr>
<td>60mm</td>
<td>17.9</td>
</tr>
<tr>
<td>70mm</td>
<td>59.5</td>
</tr>
</tbody>
</table>

LR is likelihood ratio, Sens is sensitivity, Sp is specificity, PPV is positive predictive value and NPV is negative predictive value

Preoperative assessment: patients with a high VAS score could be referred to a consultant preassessment clinic. Preoperative intervention: patients with a high VAS score could be admitted to a high dependency unit (HDU) preoperatively for optimization. This approach was taken in the hip fracture study (Burgos et al. 2008) where they recommended that any patient with an acute hip fracture with a VAS score of 40mm or higher be admitted to a HDU for more aggressive perioperative care. Postoperative intervention: high risk patients could all be admitted to a HDU postoperatively, or alternatively a ‘high risk postoperative protocol’ could be put in place for more intensive monitoring and clinical assessments in the ward.

But is the VAS prediction score reliable enough and robust enough to be used in this way? The issue of reliability has been discussed previously (section 7.11.2.2). Our study and other studies suggest that in an appropriate environment the VAS prediction may be reasonably reliable. My concern is not primarily related to the reliability of the VAS score, or to the
ability of the surgeon’s prediction to identify the high risk patient. The main problem is the potential for a ‘change in scoring behaviour’ by the surgeon. That is, if there is a potential ‘gain,’ or a perceived benefit, for giving a patient a score of >50mm, then the surgeon may increase his score to obtain the benefit for his patient. This phenomenon is called ‘gaming’ (Langer 1977). Previous studies looking at surgical prediction have not been influenced by gaming because there has been no advantage (or perceived benefit) from having a certain score.

7.11.3.2 Combining the VAS score with other objective criteria

Another approach would be to combine subjective and objective criteria. That is, to use the VAS prediction as a component of the decision making process, and to combine this with a range of objective criteria. This approach is mirrored in the development of a multifactorial risk model for predicting complications, which incorporates the VAS prediction of the surgeon with other objective clinical risk factors. In this scenario, pre-assessment to a consultant’s clinic or admission to an HDU could be determined by a VAS score of greater than an agreed value and some agreed objective ‘minimum’ criteria. Using this approach the benefits of both objective criteria and clinical assessment could be combined into a tool which is workable, reproducible and reasonably transparent.

7.12 Conclusions to Study Two

1) Surgical prediction using a global VAS score is ‘as good as’ POSSUM.

2) Surgical prediction using a multifactorial VAS is the same as when using a global VAS. The null hypothesis was supported by this finding.

3) Meaningful and clinically relevant outcome feedback can significantly improve the prediction of major complications by the surgeon. The null hypothesis was rejected, and the alternative hypothesis, that prediction could be improved, was supported.

4) A change in the prediction of complications after surgery increases the accuracy of prediction and can identify a group of high risk patients.

5) The surgeon’s VAS score can improve multifactorial risk models for predicting major complications. The null hypothesis was rejected. The surgeon’s prediction may help with designing a model to help with preoperative decision making in the individual patient.

6) In this study the prediction of the surgeon using a VAS score was accurate enough to justify it being used in interventional studies.
7.13 Further questions for study

This study has helped to ‘build a foundation’ for further work on surgical prediction. Suggestions of work arising from study two are as follows:

1) More robust studies looking at validation and reliability of prediction using a VAS.
2) Assessment of ongoing improvement: can further feedback result in an ongoing improvement in prediction?
3) Linking a change in VAS score to an intervention: can a change in prediction be used in an interventional study?
4) Development of a broadly applicable preoperative risk model, which incorporates the preoperative VAS score and can be used to help with preoperative decision making and counselling.
5) Using the VAS score to change patient management: can the VAS score be used to help in decisions about patient management? This may include decisions about a certain intervention or about entry into a more intensive treatment regimen.
CHAPTER 8
8. Factors to consider in the design and implementation of a visual analogue scale

8.1 Introduction
This chapter examines the issues identified when using a visual analogue scale to quantify the surgeon’s predictions of major complications. The three main themes addressed are:
1) Variable use of the ‘global’ visual analogue scale,
2) Design and variable use of the ‘multifactorial’ visual analogue scale,
3) Additional information gained by using a multifactorial visual analogue scale.

8.2 Variable use of the ‘global’ visual analogue scale
8.2.1 Introduction
The VAS must be appropriately designed and implemented if the surgeon’s prediction is to be correctly interpreted. Uncertainty about the correct use of the VAS, or inconsistent use of the VAS, would result in incorrect predictions being made. Not only would this (at best) result in the VAS being a ‘very blunt tool,’ but of greater concern would be its ability to give misleading results. Section 8.2 reports on problems discovered with the use of the VAS in phase I of our study (in chapter 7), and describes the steps taken to address these problems.

In our study of major complications (The Leeds study in Chapter 7) clear instructions were given about the linear use of the VAS (Figure 7.5a). The correct use of the VAS was also discussed in a departmental meeting, and with the majority of surgeons in the operating theatre. In order to check that the VAS was being used in a linear and reproducible a questionnaire which was to be administered at the completion of the first phase of the study was designed.

8.2.2 Methods
The questionnaire administered is shown in Figure 8.1. This aim of the questionnaire was to confirm how the individual surgeon was using the VAS. The secondary aim was to see if the surgeon’s prediction could be accurately converted into a percentage score for the estimation of risk. The form approached this problem in three ways. Firstly, it asked the surgeon if they had been using the VAS as a linear scale. Secondly, the surgeon was asked to place a mark on a 100mm VAS for where he would place a low risk, medium risk and high risk patient, and was then asked to give a percentage score for these different areas of risk.
Assessment of the Clinician's Prediction of the Risk of Complications

Clinician

Use of the Visual Analogue Scale....
Have you used the Visual analogue scale as a linear scale (1-10 or 1-100)?
If not how would you summarise your use of the VAS (categories, logarithmic, other)

Have you used the visual analogue scale the same way for all patients, or has your use of the scale varied according to the procedure?

| Low Risk | High Risk |

Risk of Major Complication

1) Place an “X” on the scale for a very low risk, low risk, average risk, high risk and very high risk patient
2) Place a vertical line on the scale where you would change from very low risk to low risk, from low risk to average risk, form average risk to high risk and from high risk to very high risk
3) Estimate a % likelihood for major complications that you would assign to the vertical lines which separate the risk categories

Clinical examples
Please mark the following visual analogue scales for how you would score a patient with average risk (Age 75, controlled hypertension, no other co-morbidity) and an increased risk (Diabetes, obesity, angina on walking up two flights of steps, previous abdominal surgery) for the following procedures....

- Laparoscopic cholecystectomy
- Acute laparotomy with small bowel resection
- Total gastrectomy/low anterior resection

Figure 8.1 Questionnaire
Thirdly, the surgeon was given six clinical scenarios and was asked to score the risk of these on a VAS. The operations used in the clinical scenarios were adjusted for vascular, colorectal and upper gastrointestinal surgeons and also for registrars. There were therefore four different forms – Figure 8.1 is the form used for registrars.

Although these clinical scenarios could be described as case vignette studies (section 7.11.2.2) the questionnaire was not designed to examine inter-assessor reliability. The intention of this part of the questionnaire was to give the surgeon an opportunity to assess how he/she was using the VAS by giving them a familiar case scenario to score. It was felt than inter-assessor variability could be more meaningfully addressed once the VAS forms were being used correctly.

If the surgeon was using the VAS as a linear scale then completion of the questionnaire was straight forward. For those using a non-linear scale the results of the questionnaire were discussed with the surgeon. Using the blank VAS (on uppermost box on the form), an individualized ‘conversion scale’ for the surgeon was discussed. The surgeon’s predictions from Phase I of the study were available for review if the surgeon wished to refer to them. The conversion scale was discussed until the surgeon was confident that it accurately reflected his/her percentage prediction of complications. If there was uncertainty about this, then the surgeon’s predictions were to be removed from the study database.

8.2.3 Results
8.2.3.1 Results of questionnaire
Forty-six surgeons/registrars had submitted a VAS prediction in the first phase of the study (because of rotation in clinical staff 14 new surgeons contributed to Phase II of the study). Of these 46 surgeons, one had left the hospital, and three could not adequately explain how they were using the scale. The predictions of these doctors were therefore removed from the database. Of the remaining 42 surgeons, 34 used the scale in a linear way, 6 in a logarithmic way (see Figure 8.2) and 2 used an approach which optimized the space of the VAS.

A second area of variation was that three people used the same type of scale for all patients (for example a logarithmic scale), but redefined the scale according to the complexity of the operation, so that the patient with an average risk for operations of differing complexity were always placed in the same area of the VAS scale. Therefore patients at average risk of a
complication after a laparoscopic cholecystectomy, a right hemicolectomy or a total gastrectomy were scored in the same position on the VAS scale. An example of using this approach is illustrated in Table 8.1

**Clinical examples**

Please mark the following visual analogue scales for how you would score a patient with average risk (Age 75, controlled hypertension, no other co-morbidity) and an increased risk (Diabetes, obesity, angina on walking up two flights of steps, previous abdominal surgery) for the following procedures:

- Laparoscopic Nissen Fundoplication
- Sub-total gastrectomy
- Ivor-Lewis

![Risk of Major Complication](image)

1. Place an “X” on the scale for a very low risk, low risk, average risk, high risk and very high risk patient.
2. Place a vertical line on the scale where you would change from very low risk to low risk, from low risk to average risk, form average risk to high risk and from high risk to very high risk.
3. Estimate a % likelihood for major complications that you would assign to the vertical lines which separate the risk categories.

**Figure 8.2 Example of logarithmic use of the VAS**

In almost all cases the discussion about how the VAS was used was clearly understood and the establishment of an individualized conversion scale was not difficult (Figure 8.1). In a minority of cases the surgeon thought that he/she used the VAS in a linear way, and it was only during the case scenario exercises that it was realized that this was not the case. On three occasions a second meeting was arranged to clarify points and to decide if an accurate
conversion scale could be identified. Once a conversion scale was agreed on, the surgeon was given the option of continuing to use the VAS in the same way, or of changing across to using the VAS as a linear scale. Six surgeons elected not to change, and two changed across to a linear scale.

Table 8.1 Percentage prediction of major complications for a surgeon redefining the VAS scale according to the ‘complexity of surgery’

<table>
<thead>
<tr>
<th>Complexity of operative procedure</th>
<th>Percentage risk of a major complication</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low risk patient marked at 15mm</td>
</tr>
<tr>
<td>Low risk operation (cholecystectomy)</td>
<td>5%</td>
</tr>
<tr>
<td>Medium risk operation (hemicolecotomy)</td>
<td>8%</td>
</tr>
<tr>
<td>High risk operation (total gastrectomy)</td>
<td>10%</td>
</tr>
</tbody>
</table>

8.2.3.2 Impact of using adjusted percentage scores on the prediction of complications
For the eight surgeons who used the VAS score in a consistent but non-linear way both the measured VAS score and a percentage VAS score were documented. The comparative discrimination of the prediction of major complications before and after conversion to a percentage score is shown in Table 8.2.

Table 8.2 Comparison of the discrimination of prediction before and after adjustment to a percentage scale

<table>
<thead>
<tr>
<th></th>
<th>Unadjusted Scores</th>
<th>Percentage Scores</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preoperative Prediction</td>
<td>0.753 (0.632-0.873)</td>
<td>0.778 (0.653-0.902)</td>
</tr>
<tr>
<td>Postoperative Prediction</td>
<td>0.768 (0.654-0.882)</td>
<td>0.800 (0.679-0.921)</td>
</tr>
</tbody>
</table>

c statistic, or area under the ROC curve

This analysis included 145 cases, with 22 major complications. The use of the converted scores resulted in a definite, but relatively small, improvement in the discrimination of
prediction. The discrimination of prediction for these eight surgeons after adjustment was still not as good as it was for the surgeons who used the VAS in a linear way. This suggests that an ‘under correction’ was more likely than an ‘over correction’.

8.2.4 Discussion

8.2.4.1 Implications of the variable use of the VAS

8.2.4.1.1 Impact on the results of Study Two (Leeds study)

Although the variable use of the VAS had the potential to undermine our study, the impact on the results was very minimal. The reasons for this are:

1) Clear instructions that had been given about how to use the VAS. This had resulted in only 8 of 46 surgeons using the VAS in a non-linear way.

2) Only 145 predictions were involved, and these were all in Phase I of the study. Because the problem was addressed and fixed at the end of Phase I, there was no impact on the Phase II and Phase III results.

3) The problem was also ‘corrected for’ by using the adjusted percentage prediction in our analysis of the Phase I results.

4) In terms of discrimination, the correction resulted in only a relatively minor improvement in the Phase I results.

8.2.4.1.2 Implication on the results of Study One (Dunedin study)

One major question not addressed in Chapter 7 is the improvement in the surgeon’s prediction of major complications in study two compared to study one. The most likely explanation for this difference is that the VAS was used in an inconsistent manner in study one, resulting in the effectiveness of the surgeon’s predictions being underestimated. In study one, no specific instructions were given on the VAS being used as a linear scale (a greater percentage of surgeons may have used the VAS in a non-linear way). Any difference in using the VAS would have been present throughout the whole study (not just at the beginning of the study), and there was no ‘correction’ for different uses of the VAS form. When the surgeons involved in study one were contacted some of them confirmed that they had not used the VAS as a linear scale.
8.2.4.2 Different ways of using an unlabelled VAS scale

It is not surprising, when presented with an unanchored VAS, that people are ‘naturally inclined’ to fill out the VAS in different ways. This is probably a reflection of the different approaches people have to solving problems.

Three different approaches to using the VAS scale were identified. The linear use of the scale was used by 80% of predictors because of the instructions given to use the VAS in this manner. The logarithmic use of the scale is an intuitive way of scoring. It has been shown that children, and also adults in cultures where there is no formal mathematical education, do not distribute numbers in a linear manner (Dehane et al. 2008). Instead they devote more space to small numbers and use a compressed or logarithmic mapping for larger numbers. A common logarithmic use of the VAS was to place the number 30 near the middle of the 100mm line. In the third method, placing the patient with an average risk of complications one third of the way along the VAS creates the most space to demonstrate four categories of risk – low risk, average risk, high risk and very high risk. Spatially this is an optimal use of the scale, and is an example of spatial thinking (Mathewson 1999; Hsu 2009).

8.2.4.3 Using visual analogue scales to document subjective and objective measurements

The visual analogue scale has been introduced in section 1.14. This has been shown to be reliable and sensitive for rating subjective experiences and is used to measure endpoints such as quality of life, health status and the severity of symptoms (Aitken 1969). Because of the subjective nature of these measurements they are easier to measure visually or spatially rather than to be given a number. However the biggest problem, as demonstrated in this study, is that the scale can be used differently by different people. When this happens there is no empirical basis for assigning a value to any point along the VAS. Depending on how it has been used, the VAS may therefore be relatively imprecise, and when used inconsistently it cannot be used to compare small differences.

The visual analogue scale can be further defined, by using anchors. The anchors define the line, and can dramatically improve how consistently a visual scale is used by different people. Using a VAS with numerical anchors can therefore combine the advantages of a visual scale with the advantages of a numerical scale. Therefore while an unanchored VAS is probably best for identifying patients at a lower or higher risk than average, if the objective of a study is to measure risk in a numerically meaningful way, then it is recommend that an anchored
VAS should be used. This can maintain the advantages of visual input (it may be easier to express a judgement visually than to than write a number between 0 and 100 in a box) and would still help the judgement to be expressed numerically.

8.2.4.4 Use of an anchored VAS in predicting major complications
The importance of using an anchored VAS helps us to understand the differences in results in the medical literature. The three studies that did not use an anchored VAS all showed a relatively weak prediction of complications (Pettigrew and Hill 1986; Pettigrew et al. 1987 and the Dunedin study which is discussed above). In contrast the studies that used an anchored VAS all showed very good prediction of complications. There are four studies in this category. Three showed results that were comparable to POSSUM and are discussed in section 7.8.4. The fourth (Arvidsson et al. 1996) showed that preoperative prediction by an anaesthetist was better than other individual predictors of complications such as the ASA score. These studies suggest that the way the VAS scale is used is important, and that an anchored VAS gives a more consistent, and therefore more accurate, prediction of major complications.

Our study (Study Two) is the exception. The reason for this is that measures were put in place to ensure that the VAS was used in a consistent and transparent way. This included giving clear instructions on how to use the VAS, administering the above questionnaire and identifying eight surgeons who did not use the VAS in a linear way. This extra work could have been avoided if we had chosen to use an anchored numerical VAS!

8.2.5 Conclusion
In order for VAS to give accurate and repeatable results for surgical prediction it must be used in a transparent and consistent way. The questionnaire showed that 19% of surgeons had not used the VAS as a linear scale. If the objective of a study is to combine the predictions of different surgeons and to compare surgical prediction to other systems for predicting major complications, then an anchored VAS with a numerical scale is required.
8.3 Variable use of the multifactorial visual analogue scale

8.3.1. Introduction
The multifactorial VAS was introduced in Phase II of our study to help us answer the question, “Can clinical prediction be improved by providing prompts to encourage the clinician to think about specific areas of risk before making a global assessment?”

8.3.2 Method

8.3.2.1 Design of the multifactorial VAS
The number of cues used in judgment varies widely and is said to range between three and 64. Based on Miller’s work (Miller 1956), which demonstrated that human capacity for information processing can cognitively cope with seven ± two chunks of information, it is generally recommended to use seven ± two cues. The multifactorial VAS (Figure 7.6) was designed to cover the main domains of surgical risk. The use of six VAS scales was in line with Miller’s recommendation. The six VAS scales chosen were functional status, cardiac and respiratory morbidity, severity of pathology, complexity of surgery, urgency of surgery and ‘other.’ After considering these six domains of risk the surgeon was asked to score his prediction of a major complication.

8.3.2.2 Use of the multifactorial VAS
Specific instructions on how the six subscales of the multifactorial VAS should be used were not given. The intention was for these subscales to be a ‘working area’ to help the surgeon to consider the different domains of risk before coming to an overall prediction of major complications.

8.3.3 Results
Analysis showed that the use of the additional six visual analogue scales fell into two groups. The first group used each domain of risk independently, with a wide range of scores on the different axes, which were appropriate for the patient’s known risk factors. This was best represented by a group of 13 contributors (Consultants 8, SpR 4 and SHO 1), who contributed 292 predictions. This group took what can be described as an ‘analytical approach’. The second group ‘scored around the mean’ of their global prediction. In these cases all the domains of risk were scored within 10 to 15mm of the global prediction. Obvious and important risk factors when present were not ‘independently measured’ on the appropriate domain of risk. This group was represented by 17 contributors (Consultants 5, SpR 10, and
SHO 2) and contributed 347 predictions. There was a continuum between these two extremes with 91 predictions not clearly fitting into either category. There was no difference in the accuracy of the global prediction of risk between the two groups.

8.3.4 Discussion
As no instructions had been given on how to use the six additional visual analogue scales it was not surprising that the scales were filled out differently. What was unexpected was that two distinct ‘methods’ for completing the VAS scales emerged, and that these represented two different, but equally effective, approaches to thinking about risk prediction. The group that scored each scale independently used an analytical approach. In contrast it appears that the second group had decided about their global prediction of risk before completing the six VAS subscales. They then scored each of the scales based on the global score, but slightly adjusted the score upwards for a ‘bad risk factor’ and downwards for ‘a low risk factor.’ These surgeons had ‘intuitively decided’ what the global score was before looking at the different domains of risk.

The cognitive continuum is a theory developed by Hammond as a framework for research into judgment and decision making (Hamm 1988). Using this framework, analysis moves along a continuum from intuitive to analytical thinking. Intuition is characterised by rapid information processing, low cognitive effort, the perceptual evaluation of cues, weighted averaging of information and difficulties in articulating the judgement process. In contrast analysis involves a slow and effortful processing of information according to logical rules and quantitative cues. A quasi-rational approach is midway along the continuum. The different ways in which the multidimensional VAS scales was completed in this study is an excellent illustration of the differences in intuitive and analytical thinking.

In our study the introduction of the multifactorial VAS (in phase II) made no difference to the accuracy of the prediction of major complications. This is probably because the surgeons didn’t actually change their approach to predicting complications. Those who used the scales independently were analytical in their approach from the beginning. They tended to think about the risk factors before making a decision about global risk, and the multifactorial form in phase II of our study gave then a convenient space to help them work through this process. In contrast the intuitive thinkers made an intuitive decision in both Phases I and II of the study. The separate domains of risk were essentially ‘ignored,’ and were marked at about the same
position as the global prediction to ‘fill out the form.’ If this is correct then our study did not succeed in testing if prediction using a more structured approach would result in an improvement in the prediction of major complications. The way that someone approaches risk assessment is likely to be fairly fundamental to how they think, and this was not altered by simply providing a working area for the different domains of risk!

8.4 Additional information gained when using a multifactorial visual analogue scale

8.4.1 Introduction: Potential advantages in using a multifactorial VAS

The three potential advantages of using a multifactorial VAS are: to improve surgical prediction (It did not improve prediction – see discussion in section 8.3), to gain a better understanding of how the surgeon predicts major complications, and to use the additional information to provide cognitive feedback to the surgeon.

The multifactorial VAS can improve our understanding about how the surgeon predicts major complications by enabling us to see how the surgeon weighted different risk factors into making a decision about the risk of a major complication. This is perhaps the ‘first step’ to looking inside the ‘black box’ of surgical decision making. This extra information can also be fed back to the surgeons. Whereas outcome feedback concentrates on results, the additional information on how a decision is reached is called cognitive feedback. Examples of the type of information that can be provided, and possible benefits of this are presented in the next section.

8.4.2 Methods and results

8.4.2.1 Methods: Interpreting information provided in the different domains of risk

The methods for using the VAS are described in section 7.4.6. In terms of interpreting the information provided by the multifactorial VAS, this can help us to assess how the surgeon predicts major complications by comparing the global prediction of risk against how the surgeon scored each individual domain of risk. It can also enable us to assess the relationship between the different domains of risk and the outcome (the incidence of a major complication).

For this process to be useful, each of the domains of risk would have to be scored in a transparent and defined way and each of the domains of risk would have to be scored independently (using an analytical rather than an intuitive approach). The following analysis
has therefore been performed using the thirteen surgeons who used the VAS scales independently. As no specific instructions were given on the use of the individual visual analogue scales, the results are to illustrate the type of information that a multifactorial VAS can provide, and have not been used to come to any conclusions about surgical prediction.

The relationship between each domain of risk and the prediction of major complications was assessed by using Pearson correlations and multiple regression analysis. The ability of different domains of risk to predict complications (outcome) was assessed by using logistic regression analysis.

### 8.4.2.2 Results

The Pearson Correlations between each variable and the global score (the prediction of a major complication) were: pathology 0.786, cardiorespiratory 0.704, function reserve 0.678, timing/urgency of surgery 0.640 and complexity of surgery 0.604. In the 189 cases when ‘other’ was scored (scoring ‘other’ was optional, depending on if there was a relevant risk factor in this domain) the correlation to the global score was 0.730.

The multiple linear regression analysis, looking at the contribution of the different domains of risk to the overall prediction of risk is reported in Table 8.3. This shows that the strongest contributor to the global prediction of risk was pathology, followed by cardiorespiratory function. When ‘other’ was considered important enough to be scored as a category it was the third most important contributor to the global prediction of risk. The domain of risk that contributed the least to the surgeon’s prediction of risk was the timing (urgency) of surgery.

<table>
<thead>
<tr>
<th>Domain of Risk</th>
<th>Beta coefficient</th>
<th>p Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Functional status</td>
<td>0.124</td>
<td>0.003</td>
</tr>
<tr>
<td>Pathology</td>
<td>0.407</td>
<td>0.000</td>
</tr>
<tr>
<td>Cardiorespiratory</td>
<td>0.298</td>
<td>0.000</td>
</tr>
<tr>
<td>Complexity</td>
<td>0.177</td>
<td>0.000</td>
</tr>
<tr>
<td>Timing</td>
<td>0.073</td>
<td>0.038</td>
</tr>
<tr>
<td>Other</td>
<td>0.181</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Logistic regression analysis was then performed between the six domains of risk and the outcome (complication or not). The results, for each domain of risk entered as a continuous variable, showed that the surgeon’s scoring of pathology was the only domain of risk significantly correlated to complications.

The surgeon’s pathology score was therefore the domain of risk that contributed the most to the surgeon’s global prediction of complications, and was also the only one of the surgeon’s scores that was significantly associated with the actual outcome (major complications).

8.4.3 Discussion: The role of cognitive feedback

Cognitive feedback is underpinned by the theory of judgement analysis and the Brunswik Lens Model (Appendix C). Whereas outcome feedback compares the surgeon’s prediction to the results, cognitive feedback looks at how the ‘cues’ (the domains of risk) have been used to come to a decision (risk of a major complication). It concentrates on how the decision is reached, rather than on how accurate the final decision is. Although it has been argued by cognitive psychologists that cognitive feedback is more effective in improving prediction than outcome feedback (Balzer et al. 1989), it should be noted that these two types of feedback are not necessarily in competition. Both can be used. Cognitive feedback has been reported to be particularly helpful when relationships between the cues (the domains of risk/risk factors) and the outcome being predicted are complex, especially when the best way to combine the cues is not known (Tape et al. 1992). Because of this Cognitive feedback is believed to be ideal for teaching complex judgement tasks.

There are examples of cognitive feedback improving judgement in the medical literature. This includes improving clinical prescribing (Denig et al. 2002) and improving the accuracy for predicting urinary tract infection (Wigton 1987). In a study using patient vignettes, cognitive feedback has also been shown to improve the prediction of major complications (Jacklin et al. 2009).

In summary the six cues provided by the multifactorial VAS can provide us with a descriptive method for ‘explaining’ the clinical summation of risk by the surgeon. This may provide information about the prediction of all surgeons (as in section 8.4.2.2) or about an individual surgeon. Using the information in our above example it was demonstrated that when the surgeon uses pathology as the strongest cue the prediction of major complications was very
good, but when the timing of surgery was used as the strongest cue the prediction of complications was less accurate. Whether this type of information could result in a further improvement in surgical prediction is a question that needs to be studied. Unfortunately the study on the prediction of complications was limited by having to be completed within the time frame of one year. This meant that it was only possible to provide feedback once. If we had been able to continue the study for a longer period of time it would have been possible to assess if cognitive feedback would have resulted in an incremental improvement in the prediction of major complications.

8.5 Conclusions

1) An anchored numerical VAS should be used in studies predicting major complications when the objective of the study is to measure risk in a numerically meaningful way.
2) The introduction of a multifactorial VAS did not improve surgical prediction, as providing a working area for the different domains of risk did not alter the approach the surgeon took to predicting major complications. It is likely that the study did not succeed in assessing if a more structured approach would improve the prediction of complications.
3) The multifactorial VAS can provide useful additional information about surgical prediction. Further work is required to determine if this can contribute to an incremental improvement in the prediction of major complications.
CHAPTER 9

9. Conclusions

9.1 Part A

Part A of this thesis is not just about ceftriaxone. It is primarily about the optimal use of prophylactic antibiotics to prevent infection. The impact of prophylactic antibiotics in reducing postoperative infection when first introduced was significant. Prophylactic antibiotics have now been used routinely for over 30 years, have been well studied, and the principles behind their use are well understood. Further improvements and additional steps taken to prevent infection are in comparison ‘fine tuning.’ It is my hope that this work will contribute to ongoing small improvements in our use of prophylactic antibiotics, and in the prevention of nosocomial infection.

9.2 Wound infection

9.2.1 The mechanism of action of ceftriaxone

The superior performance of ceftriaxone over other antibiotics in preventing wound infection (both in the RCT when compared to cefotaxime and in the meta-analysis when compared to a range of antibiotics) leads to the following questions: Why did ceftriaxone perform better than other antibiotics which had a similar in-vitro performance against most wound pathogens? Is there an optimal pharmacokinetic profile for an ideal prophylactic antibiotic?

Ceftriaxone combines a long half-life of 8 hours, high protein binding of over 90% and an excellent tissue penetration index of 92%. This combination of pharmacokinetic properties is rare, and is certainly unique among the cephalosporins. This results in a concentration of antibiotic well above the MIC 90 of most pathogens, which is maintained for 24 hours. Ceftriaxone is therefore present in the wound at a level of activity (relative to the MIC or MBC for most pathogens) which is greater than the level of activity of most other antibiotics. This enables ceftriaxone to suppress bacterial activity more effectively and is likely to account for its improved performance.

9.2.2 Ceftriaxone and stratification of the risk of wound infection

The meta-analysis showed no advantage for ceftriaxone when clean surgery was being performed, but there was a significant advantage in clean-contaminated, contaminated and
dirty surgical procedures. In the RCT the better protection of ceftriaxone for appendicectomy was in the context of polymicrobial infection in contaminated and dirty wounds. These observations support the idea of risk stratification, with different antibiotics being used for different levels of risk or contamination. Although there is no advantage in using ceftriaxone in clean surgery to prevent wound infection, ceftriaxone is a good prophylactic antibiotic to use at the ‘more severe end of the spectrum.’

9.3 Chest infection and urinary infection

9.3.1 Mechanism of prophylactic antibiotics in preventing chest infection and urinary infection

The question, “Is there any evidence for prophylactic antibiotics decreasing infections distant to the wound?” was examined by focusing on chest and urinary tract infection. The requirements for an antibiotic to prevent urinary or chest infection may be slightly different than those to prevent a wound infection. This is because in the respiratory and urinary systems, in addition to the contamination that occurs during surgery, there may be contamination after surgery as well. An antibiotic which is effective for up to 24 hours may therefore have an advantage in decreasing these infections. The RCT was not powered to test for differences in these infective endpoints, but did follow the trend of other RCT where ceftriaxone was used, with a lower frequency of these infections being noted. As a result of this the meta-analysis was performed. This confirmed that ceftriaxone had the ability to decrease both urinary and chest infections. It was initially thought this was because of the longer duration of action of ceftriaxone. However in the meta-analysis the difference in performance was the same when a single dose or multiple doses of the comparative antibiotic was given. This would suggest that reduction in chest and urinary infection with ceftriaxone is at least in part due to its better performance during the operative period.

9.3.2 Ceftriaxone and stratification of risk to prevent urinary and chest infection

The decision to use prophylactic antibiotics to decrease urinary or chest infection depends on the incidence of these infections. For example if the infection rate is less than 5% (in the meta-analysis chest infection was 3.3% and urinary infection 3.8%) then the number needed to treat (NNT) may be sufficiently high that it is not justified to provide prophylaxis aimed at preventing these infections (Section 6.4.6). However if the rate of these infections is 10% then
the numbers are more likely to support active prophylaxis. One approach would be to select out high risk groups. In the meta-analysis for chest infection this was in patients undergoing upper gastrointestinal and hepatobiliary surgery (NNT 33) and for urinary infection this was in patients having pelvic surgery (NNT 19).

Once a decision is made to give prophylactic antibiotics against these infections, then ceftriaxone would currently be recommended, as the meta-analysis shows that it has a consistent advantage over most other prophylactic antibiotics in decreasing chest infection and/or urinary infection.

9.4 Cost effectiveness of prophylactic antibiotics

Previous work assessing the cost effectiveness of prophylactic antibiotics tended to look at the cost of the antibiotic and its administration and/or assumed that the cost of all wound infections not prevented were the same, regardless of what antibiotic was used. The RCT demonstrated that the cost of infections not prevented is different for different antibiotics. Using cost as a measure of the severity of infection had the following benefits:

1) It allowed the true cost effectiveness of prophylactic antibiotics to be measured.
2) It provided us with a tool to compare antibiotic efficacy. The total cost of infection was able to discriminate between antibiotics as the cost of infection was sometimes significantly different even when the frequency of infection was the same. Wound scoring systems, although helpful in identifying infections and in giving a severity of infection score, have not been able to demonstrate a significant difference in the severity of infection between antibiotics. Although this result needs to be validated, it can be concluded that the total cost of infection has the potential to be a useful research tool. It is recommended that this should be included in future prospective trials comparing the efficacy of different prophylactic antibiotics. Work should also be done to try to identify less labour intensive (and cheaper!) ways to measure differences in the severity of postoperative infections.

9.5 Lessons learned from the surgeon’s prediction of wound infection

9.5.1 Protocols may be better than prediction

At the beginning of the thesis the following question was asked, “How good is the clinician at predicting the likelihood of a wound infection in the individual patient?” The answer is that
the prediction of the surgeon was poor. Prediction of wound infection in medical literature (using both multiple risk factors and logistic regression analysis) was also poor.

As it is difficult to predict who will develop a wound infection it is important that robust evidence based protocols are put in place. As well as adherence to appropriate guidelines for giving prophylactic antibiotics, protocols may include measures such as controlling blood glucose and the eradication of MRSA colonization before surgery. There is also evidence that a surveillance program which provides feedback on infection rates to individual surgeons can make a difference.

9.5.2 The importance of the surgeon changing the prediction of wound infection during the operation

The second question asked about the prediction of wound infection was, “How important is it when the clinician changes his/her estimation of the risk of a wound infection during the operation?” This was important when there was an increase in the prediction of wound infection. The main reasons given for this were intraoperative contamination, or the extent of sepsis being greater than anticipated. Intraoperative findings can therefore provide an opportunity to adjust the prophylactic antibiotic regimen.

9.6 The importance of the ASA score

Another question assessed was, “How important is the global health of the patient, as measured by the ASA classification of physical status, in the development of wound infection?” The answer is that the ASA score is more important than initially appreciated. It is in the context of effective prophylactic antibiotic use that the wound classification and duration of surgery become less significant, and that the severity of the illness and other host factors, which are indirectly measured by the ASA classification of physical status, become more important in determining if a wound infection will occur.

The following points are therefore relevant when trying to prevent wound infection.

1) It needs to be recognised that patients with a higher ASA score are at a higher risk of wound infection.
2) Greater efforts need to be made to optimize the host’s defences against bacterial infection. Current research looking at supplemental oxygenation, control of glucose, and maintenance of
body temperature are assessing different aspects of the physiology of the wound and the host’s defence against infection. It may be that a multifactorial approach which optimizes patient co-morbidities, the physiology of the wound and nutritional status will help.

9.7 Recommendations arising from Part A

Based on the above findings, recommendations for prophylactic antibiotic use and questions for further study are as follows.

9.7.1 Recommendations for the use of ceftriaxone as a prophylactic antibiotic

The RCT and meta-analysis have shown ceftriaxone to be in many ways an ‘ideal’ prophylactic antibiotic. It is very effective in reducing wound infection, and can also prevent other infections such as chest and urinary infection. It is cost effective, able to reduce the severity of infections that occur, and has few side effects when given as a single dose at the induction of anaesthesia. The meta-analysis supports the following recommendations being made:

1) Prevention of wound infection. There is no benefit in using ceftriaxone for clean surgery. However ceftriaxone is very useful as prophylaxis in clean-contaminated, contaminated and dirty surgery.

2) Prevention of pneumonia. When there is a high risk of postoperative pneumonia, such as following upper gastrointestinal and hepatobiliary surgery, then it is reasonable to use ceftriaxone prophylaxis.

3) Prevention of urinary tract infection. When there is a high risk of postoperative urinary infection, such as following O&G and colorectal surgery, then it is reasonable to use ceftriaxone prophylaxis.

9.7.2 Pathophysiology of postoperative infections

Although the pathophysiology of wound infection is well understood, three issues about the pathophysiology of wound infection where further research may help are:

1) Differences in wound infection in patients with a high ASA score. Does infection develop after a smaller inoculation of bacteria in a patient with a high ASA score?

2) The level of antibiotic cover required when there is contamination of the wound. Does giving more antibiotic when there is more contamination make a difference?
3) The critical time frame for the development of chest and urinary infection. Work to determine the usual sequence of events for the development of urinary infection and chest infection will help to clarify if antibiotic prophylaxis aimed at reducing these infections is required for just the intraoperative period, or for a longer period of time.

9.7.3 Clinical aspects of postoperative infections
Clinical interventions which may reduce postoperative infections that could be examined include:
1) Is there a role for increasing prophylactic cover intraoperatively when the surgeon believes the risk of infection has increased?
2) Is there a role for increasing prophylactic cover for a patient with an ASA score of III, IV or V? Studies looking at stratifying risk according to wound classification and the ASA score may identify a high risk group of patients who would benefit from extra interventions.
3) Are all interventions that have been shown to prevent infection being used? As prediction in the individual patient is poor the most reliable way to improve outcomes is to put in place protocols based on proven interventions.
4) More work needs to be done looking at host factors that contribute to wound infection. This is consistent with current research looking at the homeostasis of the wound.

9.8 Limitations of this study
Two important issues have been beyond the scope of this thesis. These are Clostridium difficile infection and antibiotic resistance. Although this thesis supports the use of ceftriaxone as a first line prophylactic antibiotic, the applicability of the findings will partly depend on,
a) the level of bacterial resistance to ceftriaxone in an institution,
b) evidence which addresses if single dose antibiotic prophylaxis has an adverse effect on patterns of bacterial resistance and/or on Clostridium difficile infection. Reasons why it has been difficult to address these issues are discussed in Appendix A.

9.9 Part B
Part B of this thesis assesses the ability of the surgeon to predict major complications. Although prediction is fundamental to clinical practice relatively little scientific work assessing the accuracy of surgical prediction of major complications had been performed. Our
work on surgical prediction involved two major prospective studies (the thesis concentrates on the second study), and covered three main themes. These were the accuracy of prediction of complications, the use of a VAS to capture the prediction of complications and the interface between prediction and scoring systems.

9.10 The context of performing a study assessing the prediction of major complications
Although there are a number of models used to give a risk adjusted estimation of mortality and morbidity following surgery, they are not regularly used by most practicing surgeons. Because of our interest in a patient centred approach to predicting and preventing complications, we wanted to assess how accurate the surgeon was in predicting major complications. In our first study surgical prediction was not as accurate as expected, but an increase in the VAS score was associated with an increase in major complications and the VAS prediction improved the accuracy of a multivariate model for predicting postoperative complications. The second study was designed to help answer a series of questions arising from the first study.

9.11 Accuracy of surgical prediction using a VAS
9.11.1 How accurate is surgical prediction?
In the second study surgical prediction was very good, being comparable to POSSUM. One reason for the improvement in prediction was the identification and elimination of problems caused by an inconsistent use of the VAS scale. A review of the literature identified three other studies, which all used a clearly defined VAS, where clinical prediction was comparable to POSSUM. Therefore surgical prediction has the potential to be as good as multifactorial models at predicting major complications when the VAS is used consistently.

9.11.2 Can surgical prediction be improved?
This question was answered by our three part study. In Phase I the ability of the surgeon to predict major complications was very good, with discrimination for preoperative prediction of 0.78, and 0.81 for postoperative prediction. In Phase II surgical prediction was not improved by using a multifactorial VAS. In Phase III accurate and clinically meaningful outcome feedback resulted in a significant improvement in prediction. The discrimination for preoperative prediction was 0.895, and for postoperative prediction was 0.92. The discrimination of the POSSUM morbidity score and P-POSSUM mortality score were similar
throughout the study. The finding that surgical prediction can be improved is important, not only because of the immediate improvement in prediction, but because it also provides a potential mechanism for ongoing improvement.

9.12 What lessons can be learned about the optimal use of the VAS for scoring the prediction of major complications?
The main lesson was that a clearly labelled or anchored VAS should be used. This applies equally to the global VAS and to the different domains of risk in the multifactorial VAS. As well as maintaining some of the advantages of visual input, and enabling the judgment to be expressed numerically, an anchored VAS makes it possible for different doctors to use the scale consistently.

9.13 Clinical feedback
Providing detailed and clinically relevant outcome feedback resulted in an improvement in the prediction of complications. Clinical feedback had previously been shown to influence behaviour (change of action) and result in an improvement in clinical performance. However in this study it was demonstrated that outcome feedback can also alter how a surgeon thinks (change of cognition) and result in an improvement in prediction. It was also demonstrated how the multifactorial VAS can provide us with additional information and that this information can be used to give cognitive feedback. The combination of outcome feedback and cognitive feedback may have the potential to further improve the accuracy of prediction.

9.14 What is the importance of the surgeon changing the risk of complications postoperatively?
The second study confirmed that the surgeon changing the risk of complications postoperatively was important, with an increase in the VAS score identifying a group of patients with a significantly higher risk of complications. Although technical factors were highlighted, a greater contribution to the change in risk was caused by ‘unavoidable’ factors, especially the surgical pathology.
9.15 Can surgical prediction be used to improve a model assessing the risk of complications?

Using the risk factors prospectively collected, it was demonstrated that a reasonably simple and practical model, with good discrimination, could be developed. The addition of the surgeon’s VAS prediction resulted in a significant improvement in the model. The preoperative VAS score made a major contribution to the development of a potentially useful preoperative model. One reason why the surgeon’s prediction is helpful is that it can identify and incorporate clinically important but ‘low frequency’ risk factors. This combination of objective and subjective data can make a model more accurate and may also make it more applicable to the individual patient.

9.16 Can surgical prediction be used to impact on patient management?

One objective of this study was to collect data to help decide if the surgeon’s prediction of complications using a VAS could be used to influence decision making. In the second study surgical prediction was shown to be ‘reasonably accurate,’ to be comparable to POSSUM, and also to be versatile and ‘generalisable’ enough to perform well against a number of other risk models. In this context an examination of the likelihood ratios of the surgeon’s prediction showed it would have the ability to ‘act as a test,’ or to ‘be used to help determine a course of action.’ As well as using the VAS score alone to determine a course of action, an alternative would be to use the VAS prediction as a component of the decision making process. Using this approach the benefits of both objective criteria and clinical assessment could be combined into a tool which is workable, reproducible and reasonably transparent.

9.17 Conclusions to Part B

The results of Part B have provided a database which takes clinical prediction from being an interesting concept, to being a potentially useful clinical tool. Part B began with surgical prediction using an unanchored VAS which was neither specific enough nor sensitive enough to be used to influence perioperative decision making. In contrast the prediction in study two was comparable to POSSUM and was shown to have a number of potential applications. Whereas the flexibility and adjustability of the VAS is a strength, it is also an inherent weakness. The need for an objective scoring system is not questioned, especially for comparative audit. However, it appears that the knowledge, experience and intuition of the clinician, factors lacking in an objective system, give an accurate prediction of postoperative
outcome. Although this prediction is subjective, the findings in study two and in the studies where a clearly defined VAS has been used suggest that prediction using a VAS has the potential to be reliable, and may have a number of useful applications. The findings of this study suggest that further work exploring the accuracy of prediction, and its potential applications, is justified.

Based on the above findings, five areas for further work are suggested.

1) Reliability. As prediction using a VAS is ‘subjective’ there is always the potential for the tool to be poorly used (making it less reliable than objective measures) or to be misused (because of a potential gain – the problem of gaming). More rigorous testing of the methodology should be performed. Problems with the repeatability of prediction may potentially be addressed with appropriate outcome and cognitive feedback.

2) Assessment of ongoing improvement. Is ongoing outcome feedback able to improve prediction further? Can cognitive feedback provide an incremental improvement in prediction?

3) Development of a broadly applicable preoperative risk model, incorporating the preoperative VAS score, which can be used in preoperative decision making and counselling.

4) Can a change in prediction be used as a tool for an interventional study?

5) Can the VAS score be used in an interventional study? Can this be used alone, or in combination with objective risk factors, to help to decide about a certain intervention or about entry into a more intensive regimen of treatment?

9.18 Closing summary

This thesis has examined the themes of prediction and prevention by reviewing the literature, and by using established techniques (RCT, meta-analysis, VAS) in both a traditional and innovative manner.

Our RCT and meta-analysis emphasised the potential benefits of ceftriaxone and examined the question of using prophylactic antibiotics to prevent chest and urinary infection. This raised important questions about the optimal pharmacokinetic properties of a prophylactic antibiotic. Innovations were also used. The first RCT that incorporated an assessment of the cost of infection into its comparison of antibiotic performance was performed. The cost of infection was shown to be able to demonstrate significant differences in the severity of infection. A second innovation was to assess the ability of the surgeon to predict wound
infection, using a VAS. The prediction of wound infection was poor, and emphasised the importance of protocols derived from evidence based medicine.

The second theme of the thesis, the prediction of major complications, was something that little work had been done on, and our work illustrates some of the potential difficulties that arise when working in ‘relatively uncharted’ waters. Identifying the variable use of the VAS in the second study was an important finding. This partly explained why the prediction of complications and wound infection in the first study had been ‘underestimated’ and also allowed us to correct for (and more importantly to prevent) this problem in the second study. The second study, which was designed to assess if surgical prediction could be improved, was unique in at least two ways. It introduced the use of a multifactorial VAS and assessed the impact of accurate and clinically relevant outcome feedback. This demonstrated (for the first time) that feedback can improve surgical prediction of major complications. The use of the multifactorial VAS also opens up the possibility of providing cognitive feedback, which is ideally suited to improving complex decision making. This work has therefore helped to lay a foundation for further research.
References


Bochner F, Burgess NG, Martin D. 1996. Approaches to rationing drugs in hospitals; an Australian perspective. *Pharmacoeconomics* 10:467-474


Cheadle WG. 2006 Risk factors for surgical site infection. *Surgical Infections* 7 Suppl 1:S7-11


Copeland GP. 2002. The POSSUM system of surgical audit. *Archives of Surgery* 137:15-20


Department of Surgery Clinical Audit Research Unit. 1996. *The Otago Audit System: Codes for Surgical Audit (General/Vascular)* 1996 Revision. Dunedin: University of Otago


gastrostomy significantly reduces local and systemic infective complications. *American Journal of Gastroenterology* 94:3220-3224


Hanley JA, McNeil BJ. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology 143:29-36


Hayley RW. 1995. The scientific basis for using surveillance and risk factor data to reduce nosocomial infection rates. Journal of Hospital Infection 30(Suppl):3-14


Horan TC, Gaynes RP. 2004. Surveillance of nosocomial infections. Lippincott Williams & Wilkins


Jones RN, Barry AL, Aldridge KE, Gerlach EH. 1987. Comparative antimicrobial activity of aminothiazolyl methoxyimino cephalosporins against anaerobic bacteria, including 100 cefoxitin-resistant isolates. *Diagnostic Microbiology & Infectious Disease* 8:157-168


Kracht M. 1993. Comparison of single-dose ceftriaxone and multiple-dose cefazolin or cefotaxime in abdominal surgery prophylaxis. *International Congress for Infectious Diseases* 149


Miller G. 1956. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review* 63:81-97


Mishriki SF, Law DJ, Jeffery PJ. 1990. Factors affecting the incidence of postoperative wound infection. *Journal of Hospital Infection* 16:222-230


Murphy KA, Omo GL. 1992. $198,000 cost saving antibiotic substitution plan. *ASHP-Midyear-Clinical-Meeting* 27:P-5D


Stockley JM, Allen RM, Thomlinson DF, Constantine CE. 2001. A district general hospital's method of post-operative infection surveillance including post-discharge follow-up, developed over a five-year period. *Journal of Hospital Infection* 49:48-54


produce variants of Staphylococcal b-Lactamase. *Diagnostic Microbiology and Infectious Diseases* 12:57-65


APPENDIX A:

A. Reasons why it has been difficult to assess the impact of single dose ceftriaxone prophylaxis on Clostridium difficile infection and antimicrobial resistance

Introduction

Although this thesis supports the use of ceftriaxone as a first line prophylactic antibiotic, the applicability of the findings will partly depend on the impact of ceftriaxone on Clostridium difficile infection and on antimicrobial resistance. If this occurs then the benefits and disadvantages of using prophylactic ceftriaxone need to be weighed before any recommendations can be made.

Four reasons why it has been difficult to assess the impact of ceftriaxone prophylaxis on Clostridium difficile infection and antimicrobial resistance.

1) The different patterns of infection and antibiotic resistance between different countries and institutions.
   Antimicrobial resistance and Clostridium difficile infection are the cause of major morbidity and mortality in some countries, including the United States of America and the United Kingdom. However these problems are currently relatively uncommon in some other countries, such as New Zealand and Switzerland. In the New Zealand hospital where the main RCT of this thesis was performed, there was minimal resistance to third generation cephalosporins (which did not change over the time course of this study) and MRSA infection was also very rare. We were therefore unable to collect useful data about resistance as part of our RCT. Clostridium difficile infection was assessed as part of our protocol, but in the four cases where it was documented (on an intention to treat basis) therapeutic antibiotics had also been started for another postoperative infection. At the Basel University Hospital in Switzerland, which also has a low incidence of antimicrobial resistance, the use of single dose antibiotic prophylaxis did not cause any clinically significant antimicrobial resistance (Misteli et al. 2011). However it cannot be assumed that prophylactic antibiotics will not contribute to clinical problems in an environment where significant antimicrobial resistance is already present.

2) The changing patterns of hospital infection and resistance over time.
A meta-analysis should have been an ideal opportunity to assess the impact of prophylactic ceftriaxone on *Clostridium difficile* infection. Unfortunately most of the RCTs looking at the prophylactic use of ceftriaxone were performed before 2000. In these studies the incidence of *Clostridium difficile* infection was very rare (often no cases were documented). If the same studies were performed in 2010 then the incidence of *Clostridium difficile* infection is likely to have been be higher, and more rigorous protocols would be used to identify as many cases as possible. In this context it is interesting to note that a number of other meta-analyses of prophylactic antibiotic use that attempted to address this question also found that they were unable to do so.

3) The difficulty in distinguishing between the impact of therapeutic and prophylactic antibiotics.

Although the emergence of resistant pathogens and *Clostridium difficile* infection has been widely documented with the use of therapeutic antibiotics, we have been unable to find a study that assesses the impact of the use of single dose prophylactic antibiotics. To answer this question, the impact of therapeutic antibiotics, multiple dose prophylactic antibiotics and single dose prophylactic antibiotics needs to be separated out. Observations that demonstrate changes in hospital resistance over time cannot be used to conclude that single dose antibiotic prophylaxis will cause or contribute to antibiotic resistance. Wittmann and Condon (1991) claim, “selection of resistant bacteria has not been significant and is unlikely to become so with single dose prophylaxis.” Others have claimed that antibiotic exposure greater than 48 hours is responsible for a modification of the susceptibility of antimicrobial agents (Conte et al. 1972). However in the context of dramatic changes in the profile of antibiotic resistance in many hospitals, the possibility of single dose antibiotics contributing to resistance, rather than just selecting out resistant bacteria that are already present, is a question that needs to be clarified. This question can only be answered with carefully designed prospective trials.

4) The impact of ‘additional antibiotics’ given to treat infection

In this thesis there were a number of scenarios where ceftriaxone prophylaxis resulted in fewer infections. If patients receiving another prophylactic antibiotic have more infections, and this results in additional therapeutic antibiotics being given (or an increased total exposure to antibiotics), then the impact on additional therapeutic antibiotics given on the development of resistance and on the development of *Clostridium difficile* infection needs to
be compared to the impact of single dose prophylactic ceftriaxone. In this context it may be that the impact of ceftriaxone is much less than that of additional therapeutic antibiotics that have needed to be given.
APPENDIX B:

B. Does giving accurate feedback help to improve the surgical prediction of major complications?

Introduction
This appendix presents the actual feedback given to clinicians at the end of the second phase of the Leeds study. Two documents are included in the appendix. The first is a general report given to all clinicians. This presents the overall results for Phases I and II of the study for all surgical predictions, and is presented on pages 2-6. The second is an example of the feedback given to an individual clinician, and is presented on pages 7 & 8.
Surgical audit of the ability of the clinician to predict perioperative complications

Progress so far
We have collected data predicting the risk of a major complication on 724 patients. Data collection of risk factors and complications is complete in 665 patients. These 665 patients are used in this interim analysis.

Our objective is to feedback the overall results of the study and your own results. We then hope to continue data collection for another 350 patients and will assess if the accurate feedback of results will result in a subsequent improvement in the prediction of complications. We hope that this will be a clinically relevant and a worthwhile exercise for all those who have helped by participating in the study.

Interim Analysis of Results

Overview of Procedures

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Leg endarterectomy or bypass</td>
<td>26</td>
</tr>
<tr>
<td>Aortic surgery and iliac aneurysm</td>
<td>40</td>
</tr>
<tr>
<td>Carotid endarterectomy</td>
<td>23</td>
</tr>
<tr>
<td>Major amputation</td>
<td>5</td>
</tr>
<tr>
<td>Other vascular</td>
<td>2</td>
</tr>
<tr>
<td>UGI</td>
<td></td>
</tr>
<tr>
<td>Oesophageal surgery</td>
<td>22</td>
</tr>
<tr>
<td>Gastric surgery</td>
<td>31</td>
</tr>
<tr>
<td>Obesity surgery</td>
<td>5</td>
</tr>
<tr>
<td>Duodenal surgery</td>
<td>10</td>
</tr>
<tr>
<td>Biliary, spleen and pancreatic</td>
<td>24</td>
</tr>
<tr>
<td>Cholecystectomy</td>
<td>37</td>
</tr>
<tr>
<td>Lap Nissan, Heller’s, gastropexy</td>
<td>12</td>
</tr>
</tbody>
</table>

Colorectal & GI acutes

- Laparotomy (other)       33
- Small bowel              28
- Laparoscopy, appendix, gyne 19
- Colonic resection       191
- Stoma formation         32
- Stoma closure, restore GI continuity 53
- Pouch surgery           27
- Recurrent pelvic surgery 13
- Other colorectal        15
- Other General           27
- Incisional hernia       19

Completion of forms

Surgical: 1059 forms in 665 patients, 423 by consultants, 563 by registrars and 73 by SHO’s.
Forms were completed by 58 people

Elective and Acute

Of the 665 patients the surgery was elective in 495 and was acute in 170.

Complications

665 patients.
If a patient has a complication in more than one category, only the most significant complication is included.…….

- Mortality 32
- Major complications 106
  - Reoperation
  - Organ failure requiring support (mainly ventilation, but occasionally CPAP, prolong inotropes or dialysis) [the frequencies of this group overlap with the reoperation category]
  - Major non operative interventions without reoperation or organ failure in 14 (Percutaneous abscess drainage, ERCP with biliary stent, nephrostomies, cardioversion)
  - Life threatening medical complications without reoperation or organ failure in 10 patients (MI, PE, CVA)
  - Other major surgical complications without any of the above in 4 (fistulae, graft infection)
- Other complications not included in the definition of a major complication, but included in the definition of complications in the POSSUM morbidity score 219.
This included wound complications (93), urinary sepsis or retention (25), chest infection (18), other infections (29), CLD diarrhoea (9), thrush (2), unstable angina (4), arrhythmias (16), cardiac failure (3), TIA (2), GI obstruction (9), prolonged ileus or obstruction requiring TPN (8) and a number of other problems.

- ‘Minor’ problems not included in POSSUM definition of morbidity 43. These included ileus, atelectasis, febrile morbidity, significant stoma problems, and ongoing pain problems.
- No complications in 265

Comments on complications

At first glance the complication rate is very high [mortality 4.8%, mortality and major complications 20.7%, all complications recognised in POSSUM morbidity papers 54% and any adverse event 60%]. The contributing factors to this are…

1) The complexity of the work that is referred to the LGI
2) Inclusion of acute cases
3) Deliberate inclusion of high risk patients in the study
4) The comprehensive nature of the follow-up. A large number of the complications outside the major complication category, such as wound problems and urinary infection have been diagnosed in the time period between discharge from hospital and one month from the time of surgery.

Outline of statistics used for assessment of a test

Discrimination: Receiver operating characteristic (ROC) curve analysis. This is a measurement of how accurately a test discriminates between those with a positive and negative outcome. When the area under the ROC curve has a value between 0.7-0.8 (out of 1) there is reasonable discrimination and when >0.8 this represents very good discrimination

Calibration and Goodness of fit: The Hosmer-Lemeshow test assesses the overall goodness of fit of the predictive model. An excellent fit approximates a p value of 1.0. If the data and the model are statistically different (p<0.05) then the model does not fit the data. The goodness of fit can be visually assessed by model calibration. This is done by splitting the prediction of complications into groups/deciles and assessing the accuracy of the model for each decile.

Results of clinical prediction of complications and comparison with POSSUM

Summary table of prediction complications

<table>
<thead>
<tr>
<th>Prediction method</th>
<th>Major Complications</th>
<th>POSSUM morbidity complications</th>
<th>Death</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ROC</td>
<td>H-L test</td>
<td>ROC</td>
</tr>
<tr>
<td>Surgical preoperative</td>
<td>0.77 (0.73-0.82)</td>
<td>0.81</td>
<td>0.825 (0.75-0.90)</td>
</tr>
<tr>
<td>Surgical postoperative</td>
<td>0.80 (0.76-0.84)</td>
<td>0.98</td>
<td>0.71 (0.67-0.75)</td>
</tr>
<tr>
<td>POSSUM morbidity</td>
<td>0.75 (0.70-0.80)</td>
<td>0.04</td>
<td>0.70 (0.66-0.74)</td>
</tr>
<tr>
<td>POSSUM mortality</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ROC: the area under the receiver operating characteristic curve
H-L: p value of the Hosmer-Lemeshow test, with 8 degrees of freedom
( ) 95% confidence intervals
Surgical preoperative and postoperative prediction of major complications

1) There was a reasonable preoperative discrimination and a good postoperative discrimination, see table above and ROC curve below

2) There was a good fit (see table) between the prediction and the results

Surgical Assessment

Pre and Postoperative

<table>
<thead>
<tr>
<th>1 - Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
</tr>
<tr>
<td>Sensitivity</td>
</tr>
<tr>
<td>1.00</td>
</tr>
</tbody>
</table>

Source of the Curve

- Reference Line
- SRISPOS
- SRISPRE

1) There was a reasonable preoperative discrimination and a good postoperative discrimination, see table above and ROC curve below

2) There was a good fit (see table) between the prediction and the results

Surgical Assessment

Pre and Postoperative

<table>
<thead>
<tr>
<th>1 - Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
</tr>
<tr>
<td>Sensitivity</td>
</tr>
<tr>
<td>1.00</td>
</tr>
</tbody>
</table>

Source of the Curve

- Reference Line
- SRISPOS
- SRISPRE

3) Model calibration

There was a close association between an increase in the predicted complication rate and an increase in the observed major complication rate. The surgeon however consistently overestimated the likelihood of a major complication across the full range of clinical prediction. This resulted in significantly more complications being predicted that actually occurred. 196 major complications were predicted (29.5% overall) and 132 major complications occurred (20.7%, 95% CI 17.7-23.8%). The ratio of predicted to observed complications was 1.5:1.
Comparison of clinical prediction and POSSUM prediction of major complications

The discrimination of the different tests was 0.8 for postoperative surgical prediction and 0.75 for POSSUM. This is illustrated in the following ROC curve.

SRISPOS is postoperative prediction and POSMORB is POSSUM morbidity prediction

Calibration demonstrates that the POSSUM morbidity score consistently overestimates major complications with a ratio of predicted to observed of 2.1:1.

The Hosmer-Lemeshow test shows that the POSSUM morbidity model does not fit the data, $p<0.05$. This is because the POSSUM morbidity score is designed to predict more than major complications. This includes all wound complications, all postoperative infective events and a range of postoperative medical problems. This illustrates one problem of comparing scoring systems. They all use slightly different definitions of complications. The definition of major complications in this study (enclosed) is in-line with recommendations to base the grading of a complication on the severity of the intervention required to fix the problem. There are no models for predicting major complications that match this definition. We used the POSSUM morbidity score as this was the best and the most established comparative scoring system for measuring complications in the literature.
The above results compare the prediction of the surgeon, and the POSSUM morbidity score against both the definition of major complications and the definition or morbidity used in POSSUM. The discrimination of prediction is the same using both predictions against both definitions of complications. The discrimination is less for both predictions using the POSSUM morbidity definition of complications as this includes a range of complications which are more minor, more difficult to predict and which incorporate a greater proportion of the patients being studied. As we are interested in major complications we have compared both scores against our definition of major complications. The over prediction of complications confirms that the POSSUM morbidity score does over predict major complications, but this is expected as it was designed to include a range of minor complications in its prediction of morbidity.

Comparison of clinical prediction and POSSUM mortality score in predicting death

For assessment of mortality, clinical prediction and the POSSUM prediction have a very similar discrimination (between 0.81 to 0.825), see Table. Although the clinical prediction of major complications is not designed to quantify the likelihood (% chance) of death, it can discriminate between patients who do and do not die as effectively as the POSSUM mortality score.

The discrimination and goodness-of-fit of the POSSUM mortality score was not as good as expected. POSSUM significantly over predicted the mortality (see calibration table below: the units for death is %), with a ratio of predicted to observed mortality was 2.9:1.

The observation that POSSUM over predicts mortality has been made previously. Attempts to correct this have been made by introducing modifications of the POSSUM score, including the Portsmouth POSSUM score for mortality. The results using this mortality score will be presented later.

Summary

- Clinical prediction of complications is as good as POSSUM.
- Clinical prediction overestimated the incidence of major complications.
- The POSSUM morbidity overestimation of major complications is due to the definitions of complications used.
- All methods had similar discrimination at predicting mortality, although there is room for improving both the discrimination and goodness of fit of this prediction.

SEE NEXT SECTION....YOUR PREDICTION of complications and how this compares to the above.....
**Individual report on the ability of the clinician to predict complications of surgery**

Name: XXXXXXXXXXXX
Number of reports: 163
Number of major complications: 28
Contribution to study: Most frequent contributor

**Preoperative score**
Mean score 30 (95% CI 28-33).
Median 28

**Postoperative score**
Mean 29 (26-33)
Median 25

**Distribution of preoperative score**

<table>
<thead>
<tr>
<th>Frequency</th>
<th>40</th>
<th>30</th>
<th>20</th>
<th>10</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std. Dev</td>
<td>16.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>30.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>163.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Distribution of postoperative score**

<table>
<thead>
<tr>
<th>Frequency</th>
<th>40</th>
<th>30</th>
<th>20</th>
<th>10</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std. Dev</td>
<td>20.59</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>29.3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>163.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Discrimination of prediction**

Preoperative Area under the ROC curve
0.76 (0.66-0.86)
Postoperative 0.83 (0.74-0.92)
The preoperative prediction is similar to the overall preoperative surgical prediction. The postoperative prediction is slightly better (0.83 v 0.80)

Overall good ability to distinguish between patients who do and do not develop complications

**Preoperative Goodness of fit**
Hosmer-Lemeshow statistic, p=0.78.
Excellent

**Postoperative Goodness of fit**
Hosmer-Lemeshow statistic, p=0.90
Excellent
Complications were over predicted by a ratio of 1.8:1
This was most notable in the low (17% v 5%) to average (31% v 18%) risk patients. The prediction became more accurate in moderate high risk (44% v 28%) and high risk (66% v 50%) patients.

Complications were over predicted by a ratio of 1.7:1
The main difference was an excellent prediction of complication in high risk patients. The over prediction of complications in low and moderate risk patients was unaltered.

Specific Comments
- Mean risk of a major complication was 17.2% and the mean prediction of a complication was approximately 30%. These results would suggest that the accuracy of prediction of complications could be improved by decreasing the % prediction of complications in low and medium risk patients, while maintaining the same scoring for high risk patients
- The prediction of complications in high risk patients was excellent.

General Comments
- One possible explanation for differences in predicted and observed complications may be due to the definition of a major complication. The definitions we are using for a major complication in this study are enclosed.

Thank you for contributing to this study. We hope that these comments are clinically relevant and useful. If you have any queries or questions you may contact us at the following email addresses.
John Woodfield  j.r.woodfield@clear.net.nz
Peter Sagar  peter.sagar@leedsth.nhs.uk
Dinesh Thekkinkattil  drdiku@gmail.com
Praveen Gogu  gogupraveeni@googlemail.com
APPENDIX C:

C. The Brunswik Lens Model statistics

Introduction

Underpinning the use of cognitive feedback is the theory of Judgement Analysis and the Brunswik Lens Model. The theory behind cognitive feedback, and the potential of the Lens Model to improve clinical feedback to surgeons, is discussed in this appendix.

How the Lens Model Works

Cognitive feedback looks primarily at providing information about how a decision is reached. To do this it examines how cues are used to make a judgment. The cue weights are simply the $\beta$ weights of the relationship of each cue to the judgment as determined by multiple regression analysis. These $\beta$ weights are fed back to the predictors (judge) in order to help explain how they reached their judgments. This Lens Model is illustrated in Figure 1.

![Figure C.1 The Lens Model](image)

When applied to this study, the judgement is the global prediction of major complications by the surgeon. The cues are the six domains of risk ($X_1$ to $X_6$). Each of these cues (pathology, cardiopulmonary morbidity, functional status, complexity of surgery, urgency of surgery and...
other) is correlated to the judgement and to the outcome. The criterion is the outcome, which is the presence or absence of a major complication; re represents the correlation between the cue and the outcome; rs represents the correlation between the cue and the prediction/judgment; achievement is the relationship between prediction (judgment) and the occurrence of major complications (criterion).

The Lens Model uses multiple regression analysis on both sides of the lens to correlate the prediction to the cues, and the outcome to the cues. From this the Brunswik lens model statistics can be generated. This helps to give additional useful information about surgical prediction. I have attempted to describe these statistics rather than to present these as mathematical formulas.

\( r_a \) is the achievement, or the relationship between the criterion (major complication) and the judgement (prediction). This is the overall accuracy of the prediction.

\( r_S \) is the ‘cognitive control,’ it is based on the correlation between the judgment (the global prediction of risk) and the cues (the prediction entered onto each axis or domain of risk). It looks at how consistently the surgeon applies the multiple regression equation to each case they predict. This is how consistently the cues are used in making the prediction. Good control is close to 1.0.

\( G \) is the ‘matching index.’ This looks at how well the model on the right of the lens (the model of the judge, or the model of the surgeon’s global prediction of complications) matches with the model on the left of the lens (the model relating the cues and the observed outcome). In other words it measures how well the ‘forecast model’ (how well the surgeon used the cues) matches the ‘environmental model’ (how the cues were correlated to the actual outcome). It tells us how good the current level of prediction is if the left side of the lens is/was ‘perfectly predictable.’ A good matching index is close to 1.0.

\( C \) is the correlation between the residuals (or the errors) of the regression equations. This tells us how much of the prediction is determined by factors other than the cues (the unmodelled component). In a good model this is close to 0. If \( C \) is high (0.5 would be very high) then the model is not predictive, and/or some important cues have not been included in the model.
Importantly the lens model also has the ability to deconstruct the process of prediction, and by doing this can identify what part of the prediction process should be improved.

**Potential applications of the Brunswik Lens Model to surgical prediction**

In this thesis the benefit of outcome feedback has been demonstrated. In section 8.4 it has also been explained how cognitive feedback may be used to further improve surgical prediction. However this can also be taken a step further by using the Brunswik Lens Model statistics. For example using the Lens Model for the predictions of individual surgeons would give us the ‘achievement’ and ‘cognitive control’ (consistency of how the cues are used in coming to a prediction) of each surgeon. It would also give us the matching index and the unmodelled components of the surgeon’s prediction. Deconstruction of the prediction process may also identify which part of the prediction process the surgeon needs to work on (Stewart and Lusk 1994).

The addition of outcome feedback, cognitive feedback and the lens model statistics provides the surgeon with a significant body of information about the prediction process. For the Lens Model to be used it is important that each of the cues are as meaningful as possible. This means that the scales in the multifactorial VAS must be scored independently and that each scale needs to be scored in a transparent and consistent (linear being the easiest) manner by each contributing surgeon. The potential impact of cognitive feedback and the Lens Model statistics on improving the prediction of complications has yet to be assessed.
APPENDIX D:

D. How important is it to follow complications for 30 days following surgery?

Introduction

Accurate documentation of complications is fundamental to both clinical audit and research. Although complication rates for a range of clinical indicators can be found for a patient’s stay in hospital, for many of these there is almost no documentation on the importance of 30 day follow up (Maina et al. 2002). The exception to this is wound infection, where the importance of 30 day follow has been well demonstrated. A number of studies have now demonstrated that the majority of wound infections may develop after discharge from hospital (Davey et al. 1988; Law et al. 1990; Lynch et al. 1992a; Mitchell et al. 1999, Hall et al. 2006; Taylor et al. 2003).

The purpose of this study was to accurately document the incidence of complications that develop between discharge from hospital and 30 days following surgery. In this study we wanted to identify which complications occur (or do not occur) during this time period, and to assess the frequency and severity of these complications.

Methods

A prospective clinical audit study was performed, which assessed all complications in hospital. Patients were then contacted by the use of a structured clinical telephone survey approximately 30 days following surgery.

Patients from a separate audit study in the department of General Surgery at the Leeds General Infirmary, over a 10 month period were included. Patients who died in hospital, who were still in hospital 30 days after surgery and those who were unable to be contacted by telephone were excluded.

Complications included all adverse patient events, whether they were primarily caused by the surgical procedure or not (Veen et al. 2005). Inpatient complications were documented by daily review of the clinical team and by twice weekly review of the clinical notes and patient by a clinical researcher. While the severity of inpatient complications was scored in five categories according to the scoring system proposed by Dindo et al. (2004), the severity of
outpatient complications was also scored in six categories depending on the severity of the clinical intervention required. These were: I – the patient was managed by general practitioner; II – the patient required extra support in the community; III – the patient had to attend hospital, including for investigations, but was not admitted; IV – the patient was admitted to hospital and managed by the use of antibiotics and intravenous fluids; V – the patient required a major intervention: such as an invasive radiology procedure, admission to intensive care unit or return to the operating theatre; VI the patient died. In terms of severity of complication I, II and III were considered minor, IV moderate and V and VI major.

Figure D. 1 Flow diagram for inclusion and exclusion of patients

A structured telephone survey was performed by a clinical researcher 30 days following the date of surgery. The researcher asked the following questions: 1) Have you had any problems since discharge from hospital? If yes, a description of the problem and of any treatment given for this was obtained. 2) Have there been any problems with the wound? If yes, a description of the problem and any treatment given for this was obtained. 3) Have you visited your doctor since discharge from hospital? If yes, was this (visit to your doctor) related to your surgery, and if so were you commenced on any new treatment? Three attempts were made to contact a
patient. Problems identified through the telephone survey were compared with complications documented prior to discharge from hospital to see if the problem was a continuation of a previously diagnosed problem or a new problem.

The Leeds Teaching Hospitals NHS Trust Research Ethics Committee, and the Research and Development Department, Leeds Teaching Hospitals NHS Trust approved the study.

Results

Of 578 patients undergoing surgery, 496 were discharged from hospital before one month, and of these 388 were successfully contacted by phone approximately 30 days following surgery (see Figure D.1), 107 patients were not able to be contacted by phone.

There were no significant differences noted between patients unable to be contacted by phone and patients successfully contacted by phone. In particular the type of operations and the frequency and severity of complications at the time of discharge from hospital in both groups were very similar. There was a non-significant trend towards patients younger than 40 (p=0.21) and older than 75 (p=0.14) unable to be contacted by phone.

The case mix included mainly colorectal and upper gastrointestinal surgery, followed by elective vascular and acute general surgical operations. Over 90% of the operations qualified as major surgery (see Table D.1). Acute surgery accounted for 21.3% of the cases. The complexity of the surgery, and the inclusion of acute cases contributed to the high number of cases (82) that died or were still inpatients on day 30. The median length of stay was 8 days (25% percentile 5, 50th 8, 75th 13) and the mean stay was 9.45 days (standard deviation of 5.99 days).
### Table D.1 Summary of operative procedures

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Procedure in all patients (578)</th>
<th>Procedure in patients contacted by phone (388)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Segmental colectomy</td>
<td>74</td>
<td>42</td>
</tr>
<tr>
<td>Anterior resection (including ultra-low)</td>
<td>41</td>
<td>30</td>
</tr>
<tr>
<td>Panproctocolectomy, total colectomy</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>Pelvic clearance/ pelvic exenteration/ Hartmann’s/APR/proctectomy</td>
<td>24</td>
<td>22</td>
</tr>
<tr>
<td>Partial sacrectomy</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>Pouch formation and removal</td>
<td>21</td>
<td>17</td>
</tr>
<tr>
<td>Stoma formation</td>
<td>26</td>
<td>24</td>
</tr>
<tr>
<td>Stoma closure</td>
<td>51</td>
<td>27</td>
</tr>
<tr>
<td>Fistula repair</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>En bloc excision multiple organs</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Rectal prolapse surgery</td>
<td>12</td>
<td>8</td>
</tr>
<tr>
<td>Oesophagectomy</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>Fundoplication/Heller’s myotomy</td>
<td>11</td>
<td>6</td>
</tr>
<tr>
<td>Gastrectomy</td>
<td>22</td>
<td>13</td>
</tr>
<tr>
<td>Roux-en-Y bypass/other gastric</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>CBD exploration</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Cholecystectomy</td>
<td>42</td>
<td>27</td>
</tr>
<tr>
<td>Whipples, distal pancreatectomy, debridement</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>PUD: Bleeding and perforation</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Small bowel resection, adhesionolysis</td>
<td>37</td>
<td>18</td>
</tr>
<tr>
<td>Laparotomy including trauma, abscess, gastrointestinal bypass</td>
<td>22</td>
<td>12</td>
</tr>
<tr>
<td>Laparoscopy, appendicectomy</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>Abdominal hernia repairs</td>
<td>20</td>
<td>11</td>
</tr>
<tr>
<td>Gynaecology and urology: partial cystectomy, TAH</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Aneurysm repair: Abdominal, thoracoabdominal, popliteal</td>
<td>27</td>
<td>23</td>
</tr>
<tr>
<td>Limb bypass surgery</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>Abdominal and limb endarterectomy and thrombectomy</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>Carotid endarterectomy</td>
<td>19</td>
<td>15</td>
</tr>
<tr>
<td>Other vascular: aortobifemoral graft, major amputations, mesenteric revascularization</td>
<td>10</td>
<td>6</td>
</tr>
</tbody>
</table>

Laparoscopic and open surgery is combined together

17 patients had more than one operative procedure at their initial operation.
<table>
<thead>
<tr>
<th>Complication</th>
<th>Total complications</th>
<th>After discharge from hospital</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Cardiovascular</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MI</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Unstable angina</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Cardiac failure</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td>Atrial fibrillation &amp; SVT</td>
<td>13</td>
<td>0</td>
</tr>
<tr>
<td>Other arrhythmias</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Respiratory</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pneumonia</td>
<td>16</td>
<td>2 (12.5%)</td>
</tr>
<tr>
<td>ARDS</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Respiratory failure</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Atelectasis</td>
<td>16</td>
<td>0</td>
</tr>
<tr>
<td>Pneumothorax</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Other</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td><strong>Renal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Renal failure</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Renal impairment</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>UTI</td>
<td>15</td>
<td>8 (53%)</td>
</tr>
<tr>
<td>Urinary retention</td>
<td>6</td>
<td>1 (16.6%)</td>
</tr>
<tr>
<td>Other: Catheter problems, haematuria</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td><strong>Haematological</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DVT</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>PE</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Bleeding, severe anaemia, haematoma</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Unexpected blood transfusion</td>
<td>6</td>
<td>0</td>
</tr>
<tr>
<td><strong>Neurological</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CVA</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>TIA</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Confusion</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Nerve palsy/numbness</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td><strong>Gastrointestinal</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anastomotic leak</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Fistula</td>
<td>3</td>
<td>2 (67%)</td>
</tr>
<tr>
<td>Ileus</td>
<td>25</td>
<td>0</td>
</tr>
<tr>
<td>Space SSI (Deep collection without anastomotic leak)</td>
<td>11</td>
<td>4 (36.4%)</td>
</tr>
<tr>
<td>Small bowel obstruction</td>
<td>12</td>
<td>6 (50%)</td>
</tr>
<tr>
<td>Condition</td>
<td>Count</td>
<td>Rate</td>
</tr>
<tr>
<td>-----------------------------------------------</td>
<td>-------</td>
<td>------</td>
</tr>
<tr>
<td>Clostridium difficile infection</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Functional: vomiting, diarrhoea, incontinence, constipation, other</td>
<td>21</td>
<td>15 (71%)</td>
</tr>
<tr>
<td>PUD/Haematemesis</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Dumping</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Dysphagia</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td><strong>Sepsis</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Septic shock without other documented cause</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Line sepsis</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Drain site infection</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Febrile morbidity</td>
<td>24</td>
<td>1 (4.2%)</td>
</tr>
<tr>
<td>Thrombophlebitis</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td><strong>Wound Problems</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wound infection</td>
<td>63</td>
<td>36 (57%)</td>
</tr>
<tr>
<td>Superficial wound dehiscence</td>
<td>19</td>
<td>16 (84%)</td>
</tr>
<tr>
<td>Haematoma,</td>
<td>7</td>
<td>1 (14%)</td>
</tr>
<tr>
<td>Lymphatic collection</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Vascular Surgery</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Graft/endarterectomy thrombosis or stenosis</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Graft infection</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td><strong>Other</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bile leak</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Drain/central line/feeding tube falling out</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Electrolyte problems</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Epidural problems</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>Fall</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Hernia</td>
<td>4</td>
<td>3 (75%)</td>
</tr>
<tr>
<td>Inotrope support</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>Lymphatic leak</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Medication – error in administration</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Medication – opioid overdose</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Medication – reaction/anaphylaxis</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Nutrition – failure to thrive</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Nutrition – Additional TPN</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Pain (wound, chest, abdominal)</td>
<td>15</td>
<td>10 (67%)</td>
</tr>
<tr>
<td>Pancreatitis</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Shingles</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Stoma - care problems</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Stoma - High output</td>
<td>4</td>
<td>3</td>
</tr>
</tbody>
</table>
Of 405 complications, 135 developed after discharge from hospital (33.3%). Overall 237 of 388 patients developed a complication (61.1%). A number of patients had more than one complication. Some patients had a complication in hospital and an additional complication after discharge from hospital; 77 (19.8%) developed a complication only after discharge from hospital. The type of complications that were identified in hospital and after discharge from hospital are summarised in Table D.2. A significant proportion of wound infections, episodes of ‘non infective’ wound morbidity (separation of the edges of the wound), urinary infection, early small bowel obstruction and thromboembolic events were documented after discharge from hospital. A number of patients also experienced major management issues with pain and with functional gastrointestinal problems after discharge from hospital. In contrast problems related to the physiological stress of surgery occurred in the early postoperative period.

The severity of the complications diagnosed after discharge from hospital are presented in Table D.3.

**Table D.3 Severity of complications diagnosed after discharge from hospital**

<table>
<thead>
<tr>
<th>Severity of complication</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managed by GP in the community</td>
<td>55</td>
</tr>
<tr>
<td>Required extra assistance in the community: Dressing changes</td>
<td>32</td>
</tr>
<tr>
<td>Further hospital procedures as an outpatient</td>
<td>2</td>
</tr>
<tr>
<td>Required readmission to hospital, IV fluids, antibiotics, but not other major interventions</td>
<td>38</td>
</tr>
<tr>
<td>Re-operation, TPN, Percutaneous drainage procedures, ICU admission</td>
<td>7</td>
</tr>
<tr>
<td>Death</td>
<td>1</td>
</tr>
</tbody>
</table>

**Discussion**

In this study one third of documented postoperative complications occurred after discharge from hospital. A similar result has been noted on two other occasions (Maina et al. 2002; Regenbogen et al. 2010). Although these complications were not as major as those that occurred in hospital, they were numerically significant and did include a number of complications of moderate and major severity. Follow-up after discharge from hospital was particularly important when assessing particular types of problems such as wound infection,
other infectious complications, thromboembolic events, small bowel obstruction and a range of functional problems. This study therefore reinforces the importance of following complications for 30 days after surgery.

The importance of 30 day follow-up in identifying wound infection has been well documented (Byrne et al. 1994; Davey et al. 1988; Law et al. 1990; Lynch et al. 1992a; Mitchell et al. 1999; Hall et al. 2006). Many wound infections are not clinically apparent at the time the patient is discharged from hospital. Descriptive studies have identified that approximately 50-80% of infections are apparent by the 16th postoperative day (Stockley et al. 2001; Ferraz et al. 1995; Hall et al. 2006). Other wound problems, including wound dehiscence, may also develop after discharge from hospital (Hall and Hall 2004). Our study confirms these observations with 57% of wound infections and 84% of episodes of superficial wound dehiscence occurring after discharge from hospital. In terms of severity of wound infection, almost all of these infections were managed in the community. This is consistent with the observation that wound infections that develop in the community are usually less severe (or not as expensive) as those that develop prior to discharge from hospital (Davey et al. 1992; Hall 1999; Woodfield et al. 2005). All studies assessing wound infection should therefore follow the patient for at least 30 days.

This study also made some interesting observations about other infections. In particular, 36% (4 of 11) of space surgical site infections (SSI) (in this study these were deep peritoneal collections without an anastomotic leak); 12.5% (2 of 16) of pneumonia; and 53% (8 of 15) of urinary sepsis was diagnosed following discharge from hospital. There was no other data on the frequency of these complications after discharge from hospital that we could find in the surgical literature. Urinary sepsis has previously been identified as a cause for readmission to hospital, but usually in studies that assess the use of urinary catheters (Wald et al. 2008). It is therefore recommend that accurate documentation of a range of infections, especially urinary infection and space SSI, also requires follow-up for at least 30 days.

Clinical small bowel obstruction (SBO) can occur at any time following surgery. However six of 12 episodes of obstruction occurred soon after discharge and required readmission to hospital. This is an important time period for clinically important SBO occurring. This is also an important time period for identifying major clinical thromboembolic events such as clinical
DVT and PE. With appropriate prophylactic measures these complications occur in only a few patients. However in 3 of the 4 occasions this was diagnosed after discharge from hospital. Studies assessing these complications should therefore be for at least one month, and probably longer.

It was not surprising that a number of surgically ‘less severe’ but nevertheless quite major management problems from the patient’s perspective, were identified after discharge from hospital. These problems included persistent pain, functional gastrointestinal symptoms and problems with the management of stoma devices (often contributed to by high stoma output). A number of these problems which are actively managed in hospital become much more difficult to manage in the community, especially if adequate support is not in place.

In contrast there was a group of complications that occurred in hospital, and were infrequent after discharge from hospital. These problems were mainly related to the immediate stress of surgery. They included cardiac problems, respiratory failure, atelectasis, renal failure, confusion, febrile morbidity, electrolyte disturbance and problems with epidural management and central lines. The majority of these complications occur shortly after surgery and were captured during the hospital stay.

Thirty day follow up was also important for identifying some low frequency but clinically important events. The example of thromboembolic events has already been discussed. However, 30 day follow up also identified an additional death (3.3% of deaths) and three patients who were readmitted and required additional surgery (15% of the reoperations).

A recent publication on complications after colorectal surgery (Regenbogen et al. 2010) has also reported on complications after discharge from hospital. Their findings had a number of similarities to the results of this study. Twenty-four per cent of complications were diagnosed after discharge from hospital. The most common causes were wound infection, sepsis and thromboembolic events. The percentage of different complications diagnosed after discharge from hospital were: wound infection 56%, sepsis from other causes 10% and thromboembolic events 54%. There were no diagnoses of myocardial infarction or renal failure after discharge from hospital. A number of problems identified in our study were not mentioned: including admission for bowel obstruction, re-operation and ‘functional problems.’
One issue in this study was the high rate of complications. Four factors account for this. The case mix shows that 90% of operations were major, with a significant proportion of these being complex major procedures. Secondly, all adverse postoperative events were recorded, not just complications that were directly related to the surgery. This ‘patient centred’ approach to documenting complications significantly increases the number of complications registered (Veen et al. 2005; Feldman et al. 1997; Wanzel et al. 2000). That the inclusion of medical, nursing and hospital system adverse events can result in high rates of complications is also confirmed in the SCOUT study, which had a mortality rate of 1.8% but also a morbidity rate of 51% in patients undergoing vascular surgery (Pomposelli et al. 1997).

Thirdly, some complications diagnosed following discharge were ‘managed’ while in hospital. After discharge their impact on the patient increased significantly resulting in them being considered as an adverse event or complication. An equally important reason is the vigilance of assessment and documentation of complications. Documentation of a large number of ‘minor’ complications is more complete when there is someone who does daily ward rounds to identify such events, be it a surgical nurse or a clinical researcher (Pomposelli et al. 1997).

Methods for following-up patients after discharge include the use of routine outpatient clinics, special clinics run by clinical or research staff, telephone surveys and questionnaires. An ideal method would be easy to administer, not too expensive, be acceptable to patients and give accurate results. No method easily fulfils all of these criteria. Routine clinics are often overbooked and effort is often made to avoid unnecessary consultations. Special clinics have some cost associated with them and may require an additional visit to hospital. Phone surveys are relatively cheap, acceptable to patients and can give a lot of additional information. However, capturing all patients is difficult and there are legitimate concerns about the accuracy of the information received. Questionnaires directed to all patients may have a low response rate, often in the range of 48-80% (Maina et al. 2002). Self-diagnosis by patients, especially of wound infection, has been shown to often be inaccurate (Zoutman et al. 1991).

We chose to perform a telephone survey, as this study was part of our clinical audit and there was no special budget put aside. Our response rate of 78% was within the reported range (Willsher et al. 2008; Manian and Meyer 1993; Taylor et al. 2003; Stockley et al. 2001; Reimer et al. 1987; Reilly et al. 2005). It is likely that the uptake could have been improved.
by routinely recording and using mobile phone numbers, by contacting patients during the evening as well as during the day, and by giving dedicated time to a member of hospital staff to complete the task. As the group of patients not contacted and those contacted were well matched it was felt that those contacted were likely to be representative of the whole group. It was also felt that there were a sufficient number of patients to assess the value of reviewing complications one month following surgery.

On reviewing the medical literature three issues were identified about telephone surveys: the ability to contact patients, the quality of the data received, and the effectiveness of combining telephone surveys with other methods of follow-up. In the ideal setting, with dedicated staff and protected time for contacting patients (Taylor et al. 2003) and a reasonably stable and/or affluent population it has been possible to obtain a follow-up rate of 93% (Taylor et al. 2003; Reimer et al. 1987; Stockley et al. 2001). The addition of other methods to contact non responders, including the selective use of questionnaires to patients who cannot be contacted by telephone has also been shown to be effective (Stockley et al. 2001; Willsher et al. 2008). A response rate as low as 38% has been reported when the ‘ideal setting’ is not in place (Manian and Meyer 1993).

In terms of data quality, the most information is available about follow-up of wound infection. There is good evidence that it is possible to identify significant numbers of patients who have developed infection after discharge from hospital. For example, a Canadian study (Reimer et al. 1987) documented an increase in SSI from 1.5% using the ‘standard hospital program’ to 5.4% following telephone survey. Quality however also depends on the accuracy of the information being received. In this context it has been demonstrated that a clinical review of a patient who reports a problem is beneficial. This may be performed by any appropriately trained health care worker (HCW) such as a research nurse or by the General Practitioner or by a review in a surgical clinic. The Scottish survey of wound infection in patients after inguinal hernia repair (Taylor et al. 2003) combined a telephone survey with a visit by a health worker. The confirmed rate of SSI was 5.3%, however an additional 2.1% of patients thought they had a SSI, but had no infection when reviewed by the HCW. The telephone survey had therefore overestimated the rate of SSI by 40%. In this context, an orthopaedic study (Reilly et al. 2005), which assessed SSI after joint replacement, diagnosed SSI in 3.1% as an inpatient and an additional 2.1% following discharge from hospital. They concluded that
telephone interview with rapid follow-up by a trained HCW is an effective method of identifying infection after discharge from hospital. Most studies found that telephone surveys were very acceptable to patients (Taylor et al. 2003; Stockley et al. 2001), and reported that it took on average ten to fifteen minutes work for each patient who was successfully contacted (Stockley et al. 2001; Manian and Meyer 1993). Our study also confirms these observations.

In conclusion patients should be followed for at least 30 days if an accurate picture of the complications of surgery is to be obtained. This is especially important when assessing wound infection, other infections, thromboembolic events and small bowel obstruction. It is less important when documenting cardiac events, organ failure, metabolic problems and problems related to perioperative care. Following up patients uses resources and can be expensive and the most effective method of performing patient follow up will vary in different areas. However the combination of a telephone survey (including the use of mobile phone numbers) with the patient being assessed by a HCW when a problem is documented has the advantage of being acceptable to patients and also accurately identifying most complications.