QUALITY OF CARE IN SURGERY AND METHODS TO ASSESS THE VOLUME-OUTCOME RELATIONSHIP

Mr Erik K. Mayer BSc (Hons), MBBS, MRCS (Eng)

Division of Surgery

Department of Surgery & Cancer

Imperial College London

Doctor of Philosophy (PhD) 2010

Supervisors: Professor the Lord Darzi of Denham,
Mr Thanos Athanasiou & Mr Justin Vale

DEDICATION

To my wife Anne, who has allowed me to pursue my ambitions while standing by me and bringing up our two lovely daughters Olivia and Chloe

ABSTRACT

This thesis outlines the proposed attributes of healthcare which can define its quality and reviews existing research in this area. Specifically, the limitations of existing research into the volume-outcome relationship within surgery are highlighted, thereby addressing the 'health outcome' dimension of quality of care. The principles and existing application of funnel plots within surgery are reviewed alongside their ability to overcome the limitations, risks and implications of contemporary ranking of surgical performance. Methods by which future research should be conducted are proposed as a conceptual model. This model, along with the proposed 'quality' attributes, identifies the research themes of the empirical studies.

The application of a validated methodological scoring system formally explores the limitations of existing volume-outcome research within the uro-oncological field. A cross-sectional analysis of administrative data assesses the volume-outcome relationship for radical cystectomy in England, using an improved methodology that incorporates a multilevel model to account for the relationship and influence of both the surgeon and institution and adjustment for institutional structural and process of care confounders. Subsequently, risk-adjusted funnel plot methodology is applied to the dataset to further explore provider performance. A longitudinal analysis assesses compliance with healthcare policy for radical pelvic surgery in England by exploring the patterns of service provision, in response to the existing understanding of the volume-outcome relationship for uro-oncology.

This thesis demonstrates the limitations with existing volume-outcome methodology and the need for improved methodology for both the interpretation and presentation of volume-outcome research. Appropriate handling of the data in volume-outcome analysis, which recognises the hierarchical nature, is important to adequately inform future service reconfiguration. Volume-outcome research is just one component of a quality framework that will help align healthcare with quality improvement programmes and must incorporate a multidimensional

approach to measuring and presenting both the clinical and patient-orientated measures of quality of care.

SUMMARY OF KEY POINTS ARISING FROM THIS THESIS

1. **It is possible to describe the structured approach of a quality framework which is required to appraise the quality of care in surgery in order to enhance future quality improvement programmes** (Chapter 1), published as: Mayer EK, Chow A, Vale JA, Athanasiou T. Appraising the Quality of Care in Surgery. World J Surg. 2009;33(8):1584-93.

2. **Development of multilevel modelling based on a methodological framework is needed to direct future research and the clinical translation of evidence for volume-outcome relationship** (Chapter 1), published as: Mayer E, Chow A, Darzi A, Athanasiou T. How can we assess Quality of Care in Surgery? In: T. Athanasiou (eds.), Key Topics in Surgical Research and Methodology, Springer-Verlag Berlin Heidelberg 2010.

3. **Funnel plots should play an increasingly important role as a quality control tool in future surgical performance assessment and will help to overcome some of the existing difficulties in assessing and reporting performance** (Chapter 1), published as: Mayer EK, Bottle A, Rao C, Darzi AW, Athanasiou T. Funnel plots and their emerging application in surgery. Ann Surg. 2009;249(3):376-83.

4. **The methodological quality of existing volume-outcome research as applied to cystectomy, prostatectomy and nephrectomy is only modest at best. Pooled analysis confirms a higher volume, lower mortality relationship for cystectomy and nephrectomy** (Chapter 2), published as: Mayer EK, Purkayastha S, Athanasiou T, Darzi A, Vale JA. Assessing the quality of the volume-outcome relationship in uro-oncology. BJU Int. 2009;103(3):341-9.

5. **Appropriate interpretation of the volume-mortality relationship was only possible after adjusting for institutional and surgeon volume effects and also structural and process of care confounders** (Chapter 3), published as: Mayer EK, Bottle A, Darzi AW, Athanasiou T, Vale JA. The volume-mortality relation for radical cystectomy in England: retrospective analysis of hospital episode statistics. BMJ 2010 Mar 19;340:c1128. doi: 10.1136/bmj.c1128.

6. **A volume-outcome relationship for re-intervention after radical cystectomy existed at the institutional and surgeon level** (Chapter 4). Submitted as: Mayer EK, Bottle A, Aylin P, Darzi AW, Athanasiou T, Vale JA. The volume-outcome relationship for radical cystectomy in England: An analysis of outcomes other than mortality. 2010.

7. **In the risk-adjusted funnel plot, using the final complex statistical model, no English NHS Trust demonstrated a truly divergent higher rate of mortality or re-intervention within 30 days following radical cystectomy** (Chapter 5). Submitted as: Mayer EK, Bottle A, Aylin P, Darzi AW, Vale JA, Athanasiou T. What is the role of risk-adjusted funnel plots in the analysis of radical cystectomy volume-outcome relationships? 2010.

8. **Four years after the introduction of Improving Outcomes Guidance for radical pelvic surgery, only a third of English NHS Trusts achieved the minimum standard of 50 procedures** (Chapter 6), published as: Mayer EK, Bottle A, Darzi AW, Athanasiou T, Vale JA. Provision of radical pelvic urological surgery in England, and compliance with improving outcomes guidance. BJU Int. 2009;104(10):1446-51.

STATEMENT OF ORIGINALITY

All the work presented in this thesis is my own and it is the work upon which I expect to be examined.

ACKNOWLEDGEMENTS

I am indebted to my supervisors Ara Darzi, Thanos Athanasiou and Justin Vale for their guidance and support which has allowed me to complete this period of research towards a higher research degree. I am thankful for their mentorship which has helped me develop both professionally, as an academic and clinician, and personally. I am also extremely grateful for the fantastic array of opportunities and experiences which have been made available to me by Ara Darzi since I started in his Department in 2006.

I must give special thanks to Dr Alex Bottle, from the Dr Foster Unit, with whom I have worked closely during periods of this research. Alex's access to, and experience with, HES was crucial in allowing me to explore the research questions as outlined in this thesis. I hope that our working relationship will continue to be fruitful in the years to come. I also thank Dr Paul Aylin (assistant director of the Dr Foster Unit) for agreeing to let me work in collaboration with them.

I would not be where I am today without the continued support of my family and for this I am, and will always be, grateful.

EXACT DETAILS OF CONTRIBUTIONS

All the work in this thesis was carried out by me, however in the process of completing all of the studies other individuals were involved; the details are outlined below. The whole thesis was overseen by my supervisors Professor the Lord Ara Darzi, Mr Thanos Athanasiou and Mr Justin Vale.

Chapter 1: Introduction concept by Erik Mayer (EM) and Thanos Athanasiou (TA). Written by EM. For the following reference that was incorporated into the introduction, Andre Chow helped research and write the draft manuscript versions:

Mayer E, Chow A, Darzi A, Athanasiou T. How can we assess Quality of Care in Surgery? In: T. Athanasiou (eds.), Key Topics in Surgical Research and Methodology, Springer-Verlag Berlin Heidelberg 2009.

For the following reference that was incorporated into the introduction, Christopher Rao kindly edited the funnel plot illustrations for visual aesthetics:

Mayer EK, Bottle A, Rao C, Darzi AW, Athanasiou T. Funnel plots and their emerging application in surgery. Ann Surg. 2009;249(3):376-83

Chapter 2: Study concept and design by EM. Data extracted by EM and TA. Data analysis by EM and checked by TA. Study written by EM.

Chapter 3: Study concept by EM. Design of the study by EM, TA and Justin Vale (JV). Data extracted from HES by Alex Bottle (AB). Data analysis by EM and AB (multilevel modelling in SAS). Study written by EM and AB (methods).

Chapter 4: Study concept by EM. Design of the study by EM, TA and JV. Data extracted from HES by AB. Data analysis by EM and AB (multilevel modelling in SAS). Study written by EM and AB (methods).

Chapter 5: Study concept and design by EM. Data extracted from HES by AB. Data analysis by EM and AB (predicted probabilities from SAS). Study written by EM.

Chapter 6: Study concept and design by EM. Data extracted from HES by AB. Data analysis by EM and checked by AB. Study written by EM.

PUBLICATIONS & PRESENTATIONS FROM THIS THESIS

PUBLICATIONS DIRECTLY FROM THIS THESIS

Mayer EK, Purkayastha S, Athanasiou T, Darzi AW. Redefining quality of care. J R Soc Med. 2007;100(3):122-4.

Mayer EK, Bottle A, Athanasiou T, Darzi AW, Vale JA. Volume—outcome relationship in surgical urology: Myth or reality? British Journal of Medical and Surgical Urology 2008;1:50-7.

Mayer EK, Purkayastha S, Athanasiou T, Darzi A, Vale JA. Assessing the quality of the volume-outcome relationship in uro-oncology. BJU Int. 2009;103(3):341-9.

Mayer EK, Bottle A, Rao C, Darzi AW, Athanasiou T. Funnel plots and their emerging application in surgery. Ann Surg. 2009;249(3):376-83.

Mayer EK, Bottle A, Darzi AW, Athanasiou T, Vale JA. Provision of radical pelvic urological surgery in England, and compliance with improving outcomes guidance. BJU Int. 2009;104(10):1446-51.

Mayer EK, Chow A, Vale JA, Athanasiou T. Appraising the Quality of Care in Surgery. World J Surg. 2009;33(8):1584-93.

Mayer EK, Bottle A, Darzi AW, Athanasiou T, Vale JA. The volume-mortality relation for radical cystectomy in England: retrospective analysis of hospital episode statistics. BMJ 2010 Mar 19;340:c1128. doi: 10.1136/bmj.c1128.

Mayer E, Chow A, Darzi A, Athanasiou T. How can we assess Quality of Care in Surgery? In: T. Athanasiou (eds.), Key Topics in Surgical Research and Methodology, Springer-Verlag Berlin Heidelberg 2009.

Mayer E, Darzi A, Athanasiou T. The role of volume-outcome relationship in surgery. In: T. Athanasiou (eds.), Key Topics in Surgical Research and Methodology, Springer-Verlag Berlin Heidelberg 2009.

Mayer EK, Bottle A, Aylin P, Darzi AW, Athanasiou T, Vale JA. The volume-outcome relationship for radical cystectomy in England: An analysis of outcomes other than mortality. Submitted 2010.

Mayer EK, Bottle A, Aylin P, Darzi AW, Vale JA, Athanasiou T. What is the role of risk-adjusted funnel plots in the analysis of radical cystectomy volume-outcome relationships? Submitted 2010.

OTHER PUBLICATIONS ARISING FROM WORK ASSOCIATED WITH THE WORK IN THIS THESIS

E Mayer, Faiz O, Athanasiou T, Vincent C. Measuring and enhancing elective service performance in NHS operating theatres: an overview. J R Soc Med 2008;101:273–277.

EM Burns, **EK Mayer**, Faiz O. Surgeon volume does not predict outcomes in the setting of technical credentialing: results from a randomized trial of colon cancer. Ann Surg. 2009 May;249(5):866

A Chow, **EK Mayer**, AW Darzi, T Athanasiou. Patient Reported Outcome Measures: The Importance of Patient Satisfaction in Surgery. Surgery 2009;146(3):435-43.

Mayer E & Flowers J. How to Measure Inequality in Health Care Delivery. In: T. Athanasiou (eds.), Key Topics in Surgical Research and Methodology, Springer-Verlag Berlin Heidelberg 2009.

Chow A, **Mayer E**, Darzi A, Athanasiou T. Patient Satisfaction in Surgery. In: T. Athanasiou (eds.), Key Topics in Surgical Research and Methodology, Springer-Verlag Berlin Heidelberg 2009.

PRESENTATIONS

Mayer EK. Assessing the quality of the volume-outcome relationship in uro-oncology. Department of Biosurgery & Surgical Technology, Imperial College London, November 2007. (Research meeting oral presentation)

EK Mayer, A Bottle, AW Darzi, T Athanasiou, JA Vale. The UK's volume-outcome relationship for radical cystectomy. Division of Surgery, Imperial College London, annual research afternoon, December 2007 (Prize – Highly Commended). (Short paper oral presentation)

EK Mayer, A Bottle, AW Darzi, JA Vale, T Athanasiou. The UK's volume-outcome relationship for radical cystectomy. British Association of Urological Surgeons Annual Meeting, Manchester, June 2008 (Best of BAUS). (Short paper oral presentation). Abstract: BJU International 2008;101(s5):15.

EK Mayer, A Bottle, AW Darzi, T Athanasiou, JA Vale. Provision of radical pelvic urological surgery in England and compliance with improving outcomes guidance? Division of Surgery, Imperial College London, annual research afternoon, December 2008. (Poster presentation)

Mayer EK, Bottle A, Darzi A, Athanasiou T, Vale JA. Risk-adjusted funnel plot analysis of radical cystectomy outcomes across English NHS Trusts. Accepted for poster presentation at British Association of Urological Surgeons Annual Meeting, Manchester, June 2010.

CONTENTS

DEDICATION 2

ABSTRACT 3

SUMMARY OF KEY POINTS ARISING FROM THIS THESIS 5

STATEMENT OF ORIGINALITY 7

ACKNOWLEDGEMENTS 7

EXACT DETAILS OF CONTRIBUTIONS 8

PUBLICATIONS & PRESENTATIONS FROM THIS THESIS 10

CONTENTS 13

LIST OF FIGURES 16

LIST OF TABLES 19

LIST OF ACRONYMS 21

CHAPTER 1 22

LITERATURE REVIEW - INTRODUCTION 22

1.1 APPRAISING THE QUALITY OF CARE IN SURGERY 23
1.1.1 The need to measure quality of care 24
1.1.2 Defining quality of care 25
1.1.3 Assessing quality of care and previous attempts at quality appraisal 26
1.1.4 Benchmarking quality of care 31
1.1.5 The need for a contemporary quality framework in surgery 35
1.1.6 Measuring quality of care 36
1.1.7 Future direction 50
1.1.8 Public health implications 51
1.1.9 How to design healthcare quality reforms 52
1.1.10 How to achieve healthcare quality improvement 54
1.1.11 Conclusions 55
1.1.12 Key points for appraising the quality of care in surgery 57
1.2 THE ROLE OF VOLUME-OUTCOME RELATIONSHIP IN SURGERY 58
1.2.2 Methodological framework for assessing volume-outcome relationship 62
1.2.3 Limitations of case-mix adjustment 70
1.2.4 Quality of life measures 71
1.2.5 The influence of the Surgeon and/or Institution 72
1.2.6 Existing literature in urology 76
1.2.7 Public health implications 79
1.2.8 Research & Ethical implications 81
1.2.9 Conclusion 83
1.2.10 Key points for volume-outcome relationship in surgery 84
1.3 FUNNEL PLOTS AND THEIR EMERGING APPLICATION IN SURGERY 85
1.3.1 Principles of funnel plots 86
1.3.2 Funnel plots and institutional or surgeon's performance 90
1.3.3 Over-dispersion and how to deal with it 98
1.3.4 Volume-outcome relationship 98
1.3.5 Assessment of publication bias in the literature and funnel plots 101
1.3.6 Public health issues related to surgery and application of funnel plots 104
1.3.7 Conclusions 109
1.3.8 Key points for funnel plots and their emerging application in surgery 110

CHAPTER 2 111

ASSESSING THE QUALITY OF THE VOLUME-OUTCOME RELATIONSHIP IN URO-ONCOLOGY 111

2.1 INTRODUCTION 112
2.2 MATERIALS AND METHODS 113
2.2.1 Study Selection 113
2.2.2 Primary inclusion criteria 113
2.2.3 Other considerations 114
2.2.4 Data Extraction 114
2.2.5 Endpoints 119
2.2.6 Statistical Analysis 119
2.3 RESULTS 121
2.3.1 Results of the random effects model 122
2.4 DISCUSSION 129
2.5 CONCLUSIONS 134
2.6 KEY POINTS ARISING FROM STUDY ONE 135

CHAPTER 3 136

THE VOLUME-MORTALITY RELATIONSHIP FOR RADICAL CYSTECTOMY IN ENGLAND 136

3.1 INTRODUCTION 137
3.2 METHODS 137
3.2.1 Data extraction 137
3.2.2 Assignment of institutional and surgeon volume bands 138
3.2.3 Model development 141
3.2.4 Outcome measure 143
3.2.5 Statistical Analysis 143
3.3 RESULTS 145
3.3.1 Patient demographics and institutional characteristics 145
3.3.2 Sensitivity analysis for total mortality 156
3.4 DISCUSSION 158
3.5 CONCLUSIONS 164
3.6 KEY POINTS ARISING FROM STUDY TWO 166

CHAPTER 4 167

THE VOLUME-OUTCOME RELATIONSHIP FOR RADICAL CYSTECTOMY IN ENGLAND: THE IMPACT OF EXAMINING OUTCOMES OTHER THAN MORTALITY 167

4.1 INTRODUCTION 168
4.2 METHODS 168
4.2.1 Outcome measures 168
4.3 RESULTS 169
4.3.1 Volume-outcome relationship 170
4.3.2 Interaction effect 176
4.4 DISCUSSION 189
4.4.1 Principal findings 189
4.4.2 Strengths and weaknesses of the study 190
4.4.3 Comparisons with other studies 192
4.5 CONCLUSIONS 193
4.6 KEY POINTS ARISING FROM STUDY THREE 195

CHAPTER 5 196

RISK-ADJUSTED FUNNEL PLOT ANALYSIS OF RADICAL CYSTECTOMY OUTCOMES ACROSS ENGLISH NHS TRUSTS 196

5.1 INTRODUCTION 197
5.2 METHODS 198
5.2.1 Data extraction and model development 198
5.2.2 Creating risk-adjusted funnel plots 198

5.3 RESULTS 201
5.3.1 Bar graph and Caterpillar plot analysis 201
5.3.2 Funnel plot analysis 201
5.4 DISCUSSION 209
5.4.1 Principal findings 209
5.4.2 Strengths and weaknesses of the study 210
5.4.3 Comparison with other studies 211
5.5 CONCLUSIONS 213
5.6 KEY POINTS ARISING FROM STUDY FOUR 214

CHAPTER 6 215

PROVISION OF RADICAL PELVIC UROLOGICAL SURGERY IN ENGLAND AND COMPLIANCE WITH IMPROVING OUTCOMES GUIDANCE 215

6.1 INTRODUCTION 216
6.2 METHODS 217
6.2.1 Data extraction 217
6.2.2 Assignment of radical cystectomy and radical prostatectomy volume bands 217
6.2.3 Assignment of IOG volume bands 218
6.2.4 IOG catchment populations 220
6.2.5 Radical cystectomy referral patterns 220
6.2.6 Data validation 220
6.3 RESULTS 221
6.3.1 Improving Outcome Guidance adherence 221
6.3.2 Pattern of service provision for radical cystectomy and radical prostatectomy 222
6.3.3 Referral patterns for radical cystectomy 233
6.4 DISCUSSION 235
6.4.1 Principal findings 235
6.4.2 Strengths and weaknesses of study 237
6.4.3 Comparison with other studies 238
6.5 CONCLUSION 239
6.6 KEY POINTS ARISING FROM STUDY FIVE 240

CHAPTER 7 241

CONCLUSIONS & FUTURE WORK 241

7.1 CONCLUSIONS 242
7.2 UNANSWERED QUESTIONS AND FUTURE RESEARCH 246

APPENDIX A 251

APPENDIX B 254

REFERENCES 257

LIST OF FIGURES

Figure 1.1 As defined by Donabedian, quality of care can be perceived as an interaction of three key elements 27

Figure 1.2 Multi-dimensional 'quality' measures need to include both clinical pathway and patient-reported measures. 27

Figure 1.3 The 30 day postoperative mortality (A) and morbidity (B) for all major operations performed in the Department of Veterans Affairs hospitals throughout the duration of the National Surgical Quality Improvement Program data collection process. A 27% decrease in the mortality and a 45% decrease in the morbidity were observed in the face of no change in the patients' risk profiles. FY indicates fiscal year. (figure reproduced with permission from reference [28]) 33

Figure 1.4 Examples of potential structural variables which could influence quality of care 39

Figure 1.5 Programme Budgetting – Circulatory System programme budget per capita, expenditure (million pounds) per 100,000 unified weighted population, 2004/05 vs. Mortality from all circulatory diseases, DSR, All ages, 2002-04, Persons. Area A has a low spend per capita and a corresponding high mortality rate. The reverse is true for Area B (reproduced with permission from Department of Health/Crown copyright) 45

Figure 1.6 Conceptual framework of how patient outcome following healthcare intervention is determined by interaction of several dependent components (Redrawn from reference [85]) 59

Figure 1.7 Possible relationships between volume and outcome. Reproduced from reference [88] 64

Figure 1.8 Scoring system for rating the quality of research on volume-outcome relationship. Adapted from reference [85] 67

Figure 1.9 Funnel plot of day case rates for haemorrhoidectomy in London acute NHS Trusts (2005/06). Source data from Hospital Episode Statistics online[127] 88

Figure 1.10 Simulated data used to demonstrate a symmetrical funnel plot 91

Figure 1.11 Caterpillar plot of in-hospital standardised mortality rates following coronary artery bypass grafting in 28 acute NHS Trusts in England, 2004-2005. 95% confidence intervals are plotted. Source data from the Healthcare Commission's 2004/2005 NHS Performance Ratings[130] 93

Figure 1.12 Funnel plot of standardised mortality ratio plotted against number of expected in-hospital deaths for same 28 acute NHS Trusts in England as represented in figure 3. Source data from the Healthcare Commission's 2004/2005 NHS Performance Ratings[130] 94

Figure 1.13 Ratio of % of people waiting greater than 6 weeks for an elective colonoscopy in February 2006 to those in February 2007 in 150 acute Trusts in England. Nine Trusts with missing data were excluded. The control limits are calculated assuming a rate ratio = 1 (no change in rates over the time period). No Trust appears to be failing in its efforts to reducing waiting times for colonoscopy as part of the governmental initiative ensuring that the NHS is progressing as planned towards achieving an 18 week patient pathway. Source data from Hospital Activity Statistics – Diagnostic Statistics[135] 97

Figure 1.14 Mortality rates following paediatric surgery in under-1's in 12 English specialist centres, using HES 1991-1995. The target is the overall average rate of 12 per cent. Reproduced from reference [89] 100

Figure 1.15 Simulated data used to demonstrate an asymmetrical funnel plot 102

Figure 1.16 MRSA cases per 1000 bed days in 139 Specialist and General acute NHS Trusts in England in 2004-5 (single speciality Trusts not plotted). Source data from the Healthcare Commission's 2004/2005 NHS Performance Ratings[130] 106

Figure 1.17 Funnel plots of a) standardised incidence ratios and b) standardised mortality ratios for female breast cancer during the period 2001-2003 for Primary Care Trusts (PCTs) in South West Region, Hampshire and Isle of Wight. Figures reproduced from website of South West Public Health Observatory[159] 108
Figure 2.1 Summary of reported outcomes across extracted studies 123
Figure 2.2 Random effects analysis of studies assessing adjusted mortality following cystectomy for low-volume institutions versus high-volume institutions. 127
Figure 2.3 Random effects analysis of studies assessing adjusted mortality following nephrectomy for low-volume institutions versus high-volume institutions. 128
Figure 3.1 Number of radical cystectomies performed each year in England between 2000/01 and 2006/07 inclusive. 146
Figure 3.2 Crude mortality rates (%) stratified by institutional volume tertiles. 148
Figure 3.3 Crude mortality rates (%) stratified by surgeon volume tertiles 148
Figure 3.4 Summary of odds ratio and confidence intervals for institutional volume-mortality relationships across all four models 153
Figure 3.5 Summary of odds ratio and confidence intervals for surgeon volume-mortality relationships across all four models 154
Figure 3.6 Summary of adjusted probabilities (derived from model 4) for in-hospital and total mortality across institutional and surgeon volume bands 155
Figure 3.7 Scatter plot of Institutional crude mortality rate against institutional annual cystectomy rate (volume) 162
Figure 4.1 Crude outcome rates (%) stratified by institutional volume tertiles 173
Figure 4.2 Crude outcome rates (%) stratified by surgeon volume tertiles 173
Figure 4.3 Summary of odds ratio and confidence intervals for institutional volume-re-intervention relationships across all four models 177
Figure 4.4 Summary of odds ratio and confidence intervals for institutional volume-re-intervention relationships across all four models, stratified by operation-specific and non-operation specific re-interventions 178
Figure 4.5 Summary of odds ratio and confidence intervals for surgeon volume-re-intervention relationships across all four models 179
Figure 4.6 Summary of odds ratio and confidence intervals for surgeon volume-re-intervention relationships across all four models, stratified by operation-specific and non-operation specific re-interventions 180
Figure 4.7 Summary of odds ratio and confidence intervals for institutional volume-complication relationships across all four models 181
Figure 4.8 Summary of odds ratio and confidence intervals for surgeon volume-complication relationships across all four models 182
Figure 4.9 Summary of odds ratio and confidence intervals for institutional volume-readmission and any event relationships across all four models 183
Figure 4.10 Summary of odds ratio and confidence intervals for surgeon volume-readmission and any event relationships across all four models 184
Figure 4.11 Summary of adjusted probabilities (derived from model 4) for each outcome variable across institutional and surgeon volume bands 185
Figure 5.1 Bar graph of crude mortality rate within 30 days of operation across NHS Trusts in England. Horizontal intercept line represents the national average rate of 2.9%. 203
Figure 5.2 Caterpillar plot of crude mortality within 30 days of radical cystectomy for NHS Trusts in England. 95% confidence intervals are plotted and compared with the overall proportion of 2.9% 204
Figure 5.3 Risk-adjusted funnel plots of mortality within 30 days of operation at the Trust level for models one and two 205
Figure 5.4 Risk-adjusted funnel plots of mortality within 30 days of operation at the Trust level for models three and four 206
Figure 5.5 Risk-adjusted funnel plots of re-intervention within 30 days of operation at the Trust level for models one and two 207

Figure 5.6 Risk-adjusted funnel plots of re-intervention within 30 days of operation at the Trust level for models three and four 208
Figure 6.1 Histogram showing the variation in number of Trusts by IOG volume category over the time period studied 223
Figure 6.2 Figure demonstrating percentage of Trusts performing greater than 50 radical cystectomies and radical prostatectomies per year and therefore achieving IOG compliance 224
Figure 6.3 Scatter plot of catchment population versus annual number of combined cystectomies and prostatectomies at the Trust level. The horizontal intercept line indicates the cut-off of 50 cases per year and the vertical intercept line indicates the recommended one million catchment population per Trust 225
Figure 6.4 Percentage of patients, who underwent radical cystectomy, who are over or under the age of 65 years. 227
Figure 6.5 Percentage of patients, who underwent radical prostatectomy, who are over or under the age of 65 years 228
Figure 6.6 Clustered histograms demonstrating the breakdown of the number of Trusts performing a) radical cystectomy and b) radical prostatectomy by volume group across the years 229
Figure 6.7 Correlation between annual cystectomy rate and annual prostatectomy rate at the Trust level. 232

LIST OF TABLES

Table 1.1 Donabedian's seven pillars of quality and eleven buttresses of quality assurance. Adapted from information presented in reference [16] 29

Table 1.2 The IHI Whole System Measures to measure quality of care at the level of a healthcare system. These measures can be organised according to the six aims for healthcare improvement from the IOM. 30

Table 1.3 As suggested by the Patient-Reported Outcomes Measurements Group at the University of Oxford, consideration needs to be given to eight necessary criteria to select the most suitable patient-reported outcome measures instrument(s). Adapted from online information available at http://phi.uhce.ox.ac.uk/inst_selcrit.php. 48

Table 1.4 Internet Resources for Quality of Healthcare 53

Table 1.5 Three fundamental questions that need to be answered when appraising the volume outcome literature. 61

Table 1.6 Summary table of recent literature assessing the volume-outcome relationship for cystectomy, prostatectomy and nephrectomy. 77

Table 1.7 Potential barriers to implementing a selective referral program based on a volume standard. Adapted from reference [90] 82

Table 1.8 Tabulation of the three types of outcome measure distributions used to calculate funnel plot control limits with clinical examples. Online links are provided to download formulae 89

Table 2.1 A summary of outcomes measured and unit of volume analysis at a study level 117

Table 2.2 Display of variation between studies for institutional volume categorisation and factors for which mortality was adjusted 118

Table 2.3 Scoring system used to determine study 'quality'. 120

Table 2.4 Methodological quality scores for studies assessing the volume-outcome relationship for cystectomy 124

Table 2.5 Methodological quality scores for studies assessing the volume-outcome relationship for prostatectomy 125

Table 2.6 Methodological quality scores for studies assessing the volume-outcome relationship for nephrectomy 126

Table 3.1 Breakdown of number of operations for each institutional and surgeon volume bands generated prior to regression modelling 140

Table 3.2 Variation in the structural and process of care measures between the institutional volume groups. Minimum/maximum in parentheses 144

Table 3.3 Summary of mortality outcome measures by year for radical cystectomy (unadjusted data) 147

Table 3.4 Observed numbers of in-hospital and total mortality stratified by institutional/surgeon volume bands 147

Table 3.5 Patient demographics across each of the years analysed 149

Table 3.6 Differences in patient/institution characteristics by provider volume group 150

Table 3.7 Change in ORs for medium and high-volume institutions with cumulative adjustment of each structural variable. Adding the remaining structural/process of care variables led to the final model four odds ratio. 152

Table 3.8 Differences in mortality across both institution and surgeon volume categories, displayed as crude and adjusted Odds Ratios (OR) 157

Table 4.1 Summary of outcome measures by year for radical cystectomy (unadjusted data) 171

Table 4.2 Observed number of events for each outcome variable stratified by institutional/surgeon volume bands 172

Table 4.3 Table summarising the statistically significant ($p<0.05$) associations between structural and process of care variables and outcome measures. Direction of association indicated by up (positive)/down (negative) arrow. (* Approached significance, $p=0.08$) 174

Table 4.4 Differences in outcome measures across both institution and surgeon volume categories, displayed as crude (model one) and case-mix adjusted (model two) odds ratios (OR) 186
Table 4.5 Differences in outcome measures across both institution and surgeon volume categories, displayed as odds ratios (OR) following multi-level modelling (model three) and adjustment for structural variables (model four) 187
Table 4.6 Differences in operation and non-operation specific complication and re-intervention rates across both institution and surgeon volume categories, displayed as odds ratios (OR) following multi-level modelling (model three) and adjustment for structural variables (model four) 188
Table 5.1 Summary table showing trends across Trust volume tertiles for odds ratios and mean predicted probabilities for total mortality and re-intervention within 30 days. 200
Table 6.1 Summary table indicating allocation of Trust/institutional volume bands for radical cystectomy and radical prostatectomy 219
Table 6.2 Table summarising the percentage of the annual total operations (radical cystectomy and radical prostatectomy) performed by each of the Trust volume groups. 230
Table 6.3 Comparison of Trust's volume tertile for radical cystectomy versus radical prostatectomy. Percentages calculated using total radical cystectomy number for each volume group as denominator 231
Table 6.4 Breakdown of the 813 patients referred for radical cystectomy to a Trust differing from that at which their final endoscopic bladder procedure performed. The institutional volume grouping was based-on radical cystectomy provision only 234
Table 7.1 Guidance as developed by the Institute of Medicine for improving the quality of volume-outcome research 243
Table 7.2 Details of the nine principles that the Quality Measurement Task Force used to select eleven indicators of coronary artery bypass grafting quality. Table drawn from information presented in reference [77] 248

LIST OF ACRONYMS

BAUS	**British Association of Urological Surgeons**
HES	**Hospital Episode Statistics**
HRQOL	**Health Related Quality of Life**
ICD-10	**International classification of diseases, 10th revision**
ICU	**Intensive Care Unit**
IHI	**Institute for Healthcare Improvement**
IOG	**Improving Outcomes Guidance**
IOM	**Institute of Medicine**
NHS	**National Health Service**
NICE	**National Institute for Health and Clinical Excellence**
OPCS-4	**Office of Population Censuses and Surveys *Classification of Surgical Operations and Procedures*, fourth revision**
P4P	**Pay for Performance**
PROMS	**Patient Reported Outcome Measures**
PRTO	**Patient-reported treatment outcomes**
PSA	**Prostate Specific Antigen**
PSM	**Positive Surgical Margin**
QMTF	**Quality Measurement Task Force**
STS	**Society of Thoracic Surgeons**
TCC	**Transitional Cell Carcinoma**
UK	**United Kingdom**
UKQIP	**UK Quality Indicator Project**
US	**United States**
VA	**Veterans Affairs**

Chapter 1

LITERATURE REVIEW - INTRODUCTION

The work from this introductory chapter was published as:

Mayer EK, Purkayastha S, Athanasiou T, Darzi AW. Redefining quality of care. J R Soc Med. 2007;100(3):122-4.

Mayer EK, Chow A, Vale JA, Athanasiou T. Appraising the Quality of Care in Surgery. World J Surg. 2009;33(8):1584-93.

Mayer EK, Bottle A, Athanasiou T, Darzi AW, Vale JA. Volume-outcome relationship in surgical urology: Myth or reality? British Journal of Medical and Surgical Urology 2008;1:50-7.

Mayer EK, Bottle A, Rao C, Darzi AW, Athanasiou T. Funnel plots and their emerging application in surgery. Ann Surg. 2009;249(3):376-83.

Mayer E, Chow A, Darzi A, Athanasiou T. How can we assess Quality of Care in Surgery? In: T. Athanasiou (eds.), Key Topics in Surgical Research and Methodology, Springer-Verlag Berlin Heidelberg 2009.

Mayer E, Darzi A, Athanasiou T. The role of volume-outcome relationship in surgery. In: T. Athanasiou (eds.), Key Topics in Surgical Research and Methodology, Springer-Verlag Berlin Heidelberg 2009.

1.1 Appraising the quality of care in surgery

There is a growing global movement towards quality and safety in healthcare and quality improvement initiatives in surgery [1]. Quality of care can be defined in many different ways and in order to measure it, there must be consensus as to what it comprises. Broadly, it can represent an overall impression of delivery of healthcare, but equally it can be defined in more specific terms such as treatment process or outcomes achieved. Patients are more likely to relate to the former, whilst clinicians may be more concerned with the latter. The nature of health and healthcare delivery is also changing with a greater emphasis on prevention, and novel treatments and technologies which may allow treatment of greater numbers of patients, perhaps with more co-morbidity, to a higher standard. Measures of healthcare productivity must be able to reflect this 'health inflation'.

The provision of high quality care is the universal aim of any healthcare system and those that work within it. When the health service is working at its best, it can provide excellent care to our patients, and it is well recognised that high quality care can lead to high quality results. However this is not always achieved. There exist wide variations in the quality of surgical care provision. This variation occurs between countries, regions, hospitals, departments, and surgeons. The delicate interaction of multiple factors at numerous stages of a patient's care pathway means that any single suboptimal episode can result in a cascade effect on the overall quality of care, to the detriment of the person who matters most, the patient.

This introduction forms a narrative review which describes why we need to measure quality of care and what tools we have currently. The requirement for a more structured approach to assess several multi-dimensional aspects of quality of care is also presented along with the components that would be included in such a quality framework. Particular focus is given to describing the volume-outcome relationship, its existing limitations and future potential direction for improvement. Finally the importance of appropriate statistical methodology and data presentation are discussed.

1.1.1 The need to measure quality of care

There is evidence that measuring and reporting quality of care drives improvement. Berwick et al. described two pathways by which this occurs [2]. The "change pathway" describes how the very act of measuring and benchmarking standards itself drives continuing improvement and innovation. Further to this the "selection pathway" describes how publicly released performance data is compared, and better-performing providers are rewarded by 'selection' of that provider. This selection pathway appears to carry more motivational drive than the change pathway. A third pathway, the "reputational pathway" was proposed by Hibbard [3]. This describes a provider's concern for public image or reputation. Public reporting of performance data appears to stimulate quality improvement activity at the hospital level, although it is unclear if this translates into improved effectiveness, safety, and patient-centeredness [4]. Transparency of performance and outcomes reporting, and therefore public accountability, is a way of empowering the patient to inform choice; choice is becoming a legal requirement through the patient constitution [5] and is a common aspect of the patient's perception of quality of care[6].

Beyond public accountability, performance reporting has also been linked with financial incentives in pay-for-performance schemes [7 8]. Five key dimensions have been suggested as the most important influencing determinants of pay for performance programmes:

- Selecting high-impact performance measures
- Making payment reward all high-quality care
- Prioritizing quality improvement for under-served populations
- Appropriate financial rewards
- Choosing when to incentivise individuals or institutions [9]

Financial incentives have been shown to modestly increase improvements in quality among hospitals already engaged in public reporting [10].

1.1.2 Defining quality of care

The Institute of Medicine (IOM) in the US defines quality of care as:

"..the degree to which health services for individuals and populations increase the likelihood of desired health outcomes and are consistent with current professional knowledge" [11]

The American Medical Association defines high quality care as:

"[that] which consistently contributes to the improvement or maintenance of quality and/or duration of life" [12].

Or the definition can incorporate a patient-orientated emphasis such as that of BUPA Hospitals UK [13]:

"…ability to provide the service you want and need - resulting in medical treatment you can rely on and personal care you'll appreciate"

It is clear from these definitions that the term 'quality care' may imply different things to both clinicians and patients. From the clinician point of view, high quality care means up-to-date, evidence based patient care that results in improved clinical outcomes. Although this is also important to the patient, they may be more concerned with aspects of care such as availability, flexibility, reliability and personal touches such as politeness and empathy of medical staff.

The term 'quality of care' can therefore be broadly defined to represent an overall impression of delivery of healthcare, but equally requires some very specific and agreed measures of the treatment process or outcomes achieved. This makes it a complex entity to encompass, requiring an ordered approach.

1.1.3 Assessing quality of care and previous attempts at quality appraisal

Although current definitions of quality of care are applicable, they are deliberately vague and therefore of limited use in defining the assessment of quality of care. Although it can be obvious when high quality care is being provided, providing objective proof of this can be more challenging. The inherent flexibility of healthcare provision must also be considered; with innovation in medical technology and treatments, optimum care standards and therefore the markers of quality of care evolve. It is therefore easy to see why standardising the assessment of quality of care even at a procedural level can be problematic.

Quality of care assessment should include every aspect of a patient's journey through their healthcare system. This would encompass community care, screening where applicable, referral to a specialist, and processes of investigation, diagnosis and treatment. Also needed are details of post-operative management, and follow-up both in the hospital and in the community. In other words, the assessment of quality should be multi-factorial. It is clear that there are countless variables which could be measured. How then to either measure all of them or identify the most pertinent ones?

The principles of using a conceptual framework for organising measures of healthcare quality have been previously recognised. In 1966 Donabedian divided quality of care into three tangible parts; structure, process and outcome [14] (Figure 1.1). Structure is concerned with the actual infrastructure of the healthcare system. This includes aspects such as the availability of equipment, availability and qualifications of staff, and administration. Process looks at the actual details of care including aspects from diagnostic tests, through to interventions such as surgery and continuity of care. Outcome looks at the end result of medical care, traditionally in the form of survival, and restoration of function.

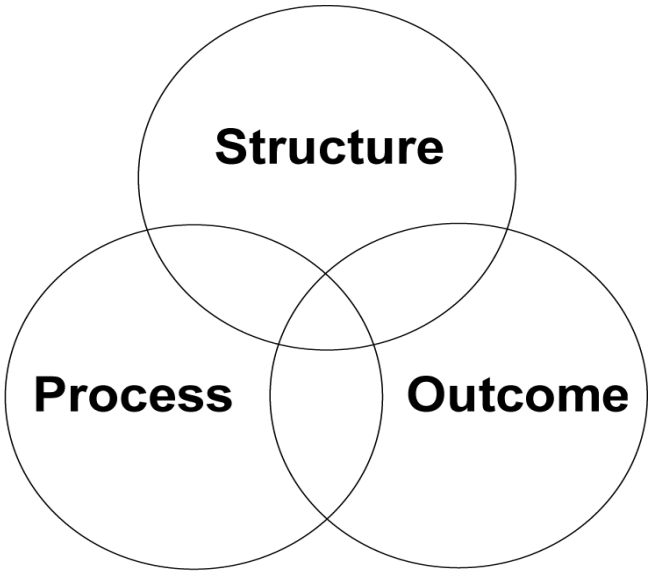


Figure 1.1 As defined by Donabedian, quality of care can be perceived as an interaction of three key elements

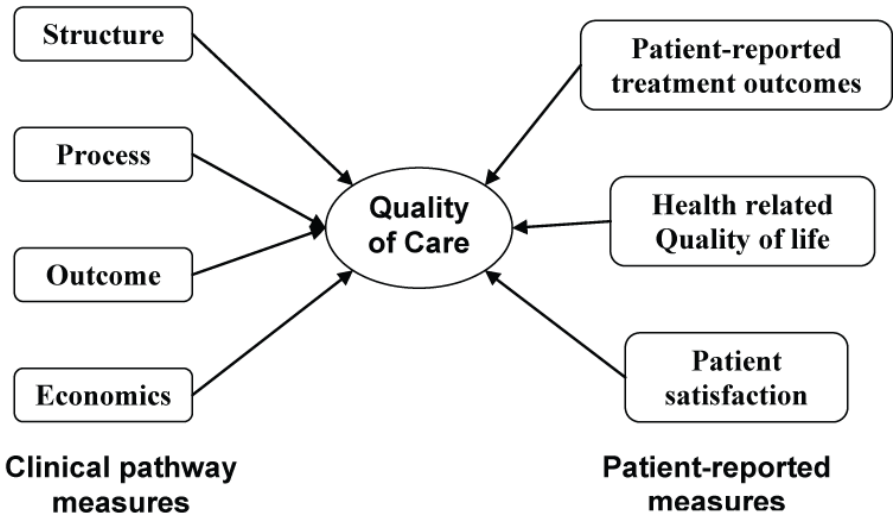


Figure 1.2 Multi-dimensional 'quality' measures need to include both clinical pathway and patient-reported measures.

A fourth element can be proposed for this conceptual model; healthcare economics or its dependent measure productivity. In modern medicine, the availability of financial resources and any resulting financial constraints can impact upon the accessibility and delivery of healthcare services and potentially therefore the provision of high quality care. This is particularly true of publicly funded systems, such as the NHS. The ability of a healthcare provider to deliver a quality of care operating within financial or resource restrictions is an important factor that must be considered. The notion of cost being associated with quality of care is not novel however; Donabedian described several attributes of healthcare which define its quality, the 'seven pillars of quality'. One of these, 'efficiency', related to the "ability to obtain the greatest health improvement at the lowest cost" [15]. Although Donabedian introduces the concept of cost within his quality attributes, the organisation of a contemporary healthcare service is so influenced by business planning that it directly shapes it. For this reason a strong argument can be made for healthcare economics to be included in a conceptual model of quality of care. It could also be argued that healthcare economics forms part of structure and therefore does not require special attention.

Over the years Donabedian built up a framework for defining quality, starting with his structure-process-outcome paradigm, then the seven pillars of quality and finally the 11 buttresses of quality assurance (Table 1.1) which are essential for the design, operation and effectiveness of care [16]. Quality assurance itself can be seen as a combination of system design which results in rough adjustments in performance and quality monitoring, which is responsible for fine tuning of performance.

More recently the IOM [17] and the Institute for Healthcare Improvement (IHI) [18] have also set specific aims for quality measurement (Table 1.2). A review of the NHS has also recognised the need for a greater focus on appraising quality in healthcare [5]. The need for a strong evidence base to support quality improvement initiatives is implicit [19] [20] and conceptual frameworks have been proposed to facilitate the translation of published evidence into policy and managerial decisions for improving quality [21].

Seven Pillars of Quality	The Eleven Buttresses of Quality Assurance
1. **Efficacy** - The ability of care, at its best, to improve health	1. **Interdependency** – Nothing should be viewed in isolation
2. **Effectiveness -** The degree to which attainable health improvements are realised	2. **Organisational Dependency** – Monitoring of care needs to be structured so that it is representative of all organisations throughout the healthcare system; organisations that depend on one another
3. **Efficiency -** The ability to obtain the greatest health improvement at the lowest cost	3. **Consensuality** – Alignment of stakeholders (e.g. healthcare professionals and management) towards a common purpose
4. **Optimality -** The most advantageous balancing of costs and benefits	4. **Congruence** – Quality assurance needs to balance professional accountability with professional autonomy
5. **Acceptability -** Conformity to patient preferences regarding accessibility, patient-practitioner relationships, amenities, effects of care, and cost of care	5. **Credibility** – There needs to be trust in the monitoring process, e.g. completeness and accuracy of data being used
6. **Legitimacy -** Conformity to social preferences concerning all of the above	6. **Relevance** – Quality assurance should be 'tailored' to individual clinician's practice
7. **Equity -** Fairness in the distribution of care and it's effects on health.	7. **Ownership** – Giving ownership of the monitoring process to those being monitored
	8. **Mutuality of Interests** – The self-interests of both the monitored and monitors need to be satisfied by the quality assurance process
	9. **Facilitation** – Using resources or restructuring to overcome the barriers to quality assurance
	10. **Coerciveness** – The legitimacy and necessity of coerciveness to make quality monitoring successful is accepted. The degree of intrusion of the quality monitoring process does however need to be defined
	11. **Personal and Public Virtue** – A commitment to the pursuit of quality as a moral dimension of professional life. This will facilitate any quality monitoring process.

Table 1.1 Donabedian's seven pillars of quality and eleven buttresses of quality assurance. Adapted from information presented in reference [16]

IHI Whole System Measures	IOM Aims for Healthcare Improvement
1. Rate of Adverse Events	Safe
2. Incidence of Non-fatal Occupational Injuries and Illnesses	
3. Hospital Standardised Mortality Ratio (HSMR)	Effective
4. Unadjusted Raw Mortality Percentage	
5. Functional Health Outcomes Score	
6. Hospital Readmission Percentage	
7. Reliability of Core Measures	
8. Patient Satisfaction with Care Score	Patient-Centred
9. Patient Experience Score	
10. Days to Third Next Available Appointment	Timely
11. Hospital Days per Decedent During the last 6 months of life	Efficient
12. Health Care Cost Per Capita	
13. Equity (Stratification of Whole System Measures)	Equitable

Table 1.2 The IHI Whole System Measures to measure quality of care at the level of a healthcare system. These measures can be organised according to the six aims for healthcare improvement from the IOM.

The equality of healthcare provision is a vital part of high quality care. It has been suggested that quality assessment tools have failed to address adequately healthcare inequality across socio-economic groups [22] and few have looked at the impact of healthcare access (a component of healthcare delivery) on health inequality [23].

1.1.4 Benchmarking quality of care

1.1.4.1 Recent initiatives

The need for maintenance of high standards and the improvement of quality of care is well recognised. There are a number of existing programmes dedicated to the improvement of quality of care. The majority of these base their work on performance benchmarking.

Performance benchmarking is a tool that allows organisations to evaluate their practice as compared to accepted best practice. If any deficiencies exist, adjustments can be made with the aim of improving the overall performance. This process must be continuous as healthcare is a continually evolving entity. Currently, healthcare institutions are either benchmarked against national targets or each other as a means of comparison. This approach identifies 'good' and 'bad' outliers and a cohort of 'average' performers. It also serves to identify inequalities that exist, which can then be addressed. This method of benchmarking does help to maintain a nationwide drive to continuously improve services, although there are critics of any system that arbitrarily 'ranks' performance without due consideration for underlying causative factors.

The Healthcare Commission was an independent body that promoted improvements in quality of care in both the NHS and independent health sectors in England and Wales. Its role was to assess and report upon the performance of healthcare organisations to ensure high standards of care. It evaluated performance against targets set by the Department of Health. The Healthcare Commission also looked at clinical and financial efficiency, giving annual performance ratings for each NHS Trust. The areas assessed were generalised, and

included categories such as patient safety, clinical and cost effectiveness, governance, and waiting times. In April 2008 the Care Quality Commission became the independent regulator of health and adult social care services in England and took on the responsibilities of the Healthcare Commission.

The UK QIP (UK Quality Indicator Project) [24] is part of an international programme (IQIP: International Quality Indicator Project) that was started in the USA in 1985. UK QIP is a voluntary exercise based upon the anonymous feedback of comparative data to encourage internal improvement within healthcare organisations. There is no system for publication of results or external judgement of data. By using performance indicators, the aim of the project is not to directly measure quality but to identify areas that require further attention and investigation. Examples of surgical performance indicators include rates of hospital-acquired infections, surgical site infections, in-patient mortality and readmission rates.

A similar project exists in the US alone called the NSQIP: National Surgical Quality Improvement Programme [25]. This nationwide programme was started by the department of Veterans Affairs (VA) to monitor and improve the standards of surgical care across all VA hospitals, and has been slowly introduced into the private sector since 1999 (Figure 1.3).

International benchmarking of healthcare systems has the opportunity to globally improve the quality of healthcare. Sharing of best practice can be beneficial in a supportive network of participating organisations. Such projects include the World Health Organisation's Performance Assessment Tool for Quality Improvement in Hospitals (PATH) project [26] and the Organization for Economic Cooperation and Development (OECD) healthcare quality indicators project [27]. None has been designed specifically for surgery, but some do include surgical performance indicators, such as surgical wound infection and surgical prophylaxis, perioperative mortality and unplanned returns to the operating theatre [24]. An international review of projects on hospital performance assessment concluded that there was "common methodology for the design and selection of

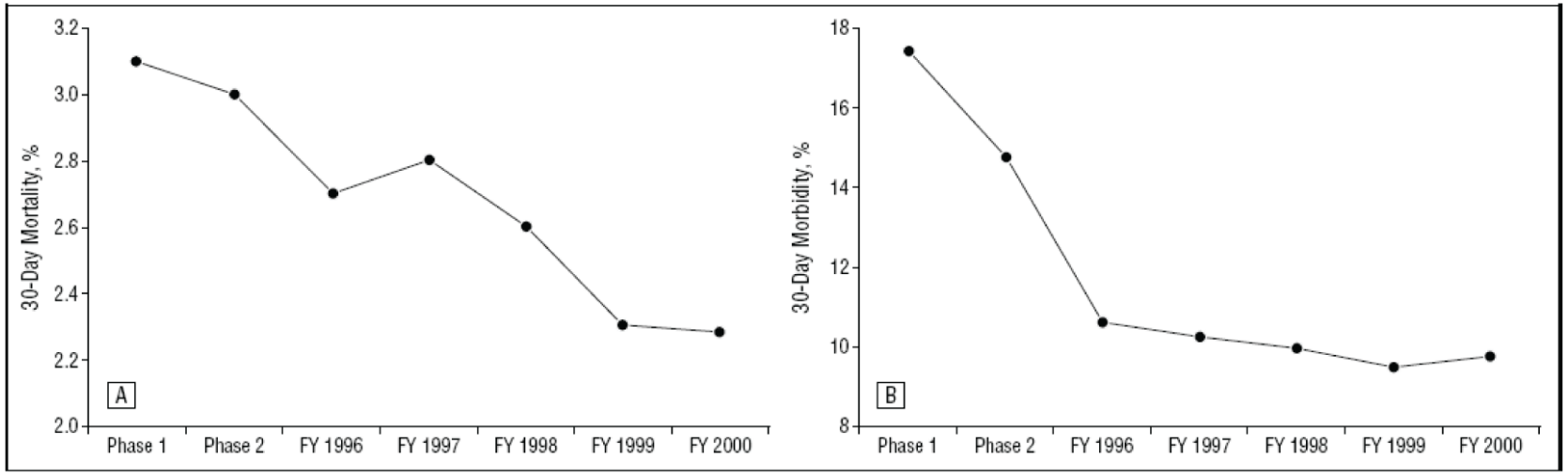


Figure 1.3 The 30 day postoperative mortality (A) and morbidity (B) for all major operations performed in the Department of Veterans Affairs hospitals throughout the duration of the National Surgical Quality Improvement Program data collection process. A 27% decrease in the mortality and a 45% decrease in the morbidity were observed in the face of no change in the patients' risk profiles. FY indicates fiscal year. (figure reproduced with permission from reference [28])

indicators; however, major differences exist with regard to the philosophy, scope and coverage of the projects" [29].

Performance benchmarking is a useful exercise for ensuring that the minimum standard of care that can be expected is attained, but is far too vague and imprecise to inform us if we are delivering high quality care. This is the same problem that any generalisable quality assessment tool will experience, as it also will be unable to appreciate the intricacies of disease-specific high quality healthcare.

1.1.4.2 Pay for Performance Strategies

Pay-for-performance (P4P) programmes use financial reimbursements for clinical providers as a 'reward' for a positive change in performance measures. It is thought that this will help drive further improvements in quality of care. These programmes have gained popularity in recent years with new initiatives in the USA [10 30], UK [7], Australia [31], and Canada [32], being based in both hospital and primary care. P4P programmes typically focus upon process measures as these can detect suboptimal care in a timely manner, whilst being directly under control of the clinician.

There are a multitude of variations of P4P programmes, with incentives being paid either to individual clinicians, clinician groups, clinics, hospitals, or multi-hospital collaborations. Similarly, the amount of incentive required per measure can vary from $2 to $10,000, with incentives received either for reaching absolute thresholds of care, relative thresholds (such as a 30% increase in performance) or even a pay-per-case arrangement.

Although studies have shown that P4P programmes can have positive effects on quality measures, these gains may be only modest [10 33]. Cost-effectiveness is also unclear with some studies showing massive savings [34], and others showing gross overspending [35].

Perhaps the most worrying aspect of P4P programmes is the unintended adverse consequences that can result. Examples of these include 'gaming' strategies where clinicians avoid sick or challenging patients, or reclassify patient conditions, or even claim their incentive when care has not been provided. Similarly patients may receive substantial 'over-treatment' of their medical conditions. In fact, the NHS P4P programme in the UK found that the strongest predictor of improvement in achievement was the exclusion of patients from the programme [7]. On the other hand, clinicians and hospitals serving the more disadvantaged populations may see their income fall as targets and thresholds are difficult to reach.

Although P4P programmes have been shown to improve performance in key clinical areas they can potentially have multiple problems if not subject to careful design and regular evaluation. In order to be successful these programmes must be implemented with the involvement of clinicians from the very start to prevent unintended harm coming to the patient.

1.1.5 The need for a contemporary quality framework in surgery

Despite all of the universally accepted benefits to measuring quality of care in surgery there is currently no universally accepted and/or validated measurement system. An example of how an evidence-based quality framework can be used to improve healthcare has been seen with improvements in stroke services in the NHS. The Department of Health in England recognised the need to improve stroke services by implementing the National Service Framework (NSF) for older people in 2001. This in particular concentrated upon the structure of stroke services as well as the process of care for patients suffering stroke. The Biannual Sentinel Stroke Audit for 2008 has recently been published [36], and demonstrates a continued significant improvement in stroke services. In terms of healthcare structure, 96% of hospitals in the UK now offer specialist stroke services, with an increasing number of specialist stroke unit beds. 98% of hospitals employ a physician with a specialist interest in stroke. There have also been improvements in process of care measures including the uptake of thrombolysis services and

secondary prevention measures. A similar initiative has been beneficial for coronary heart disease [37] and more recently broadly applied to cancer [38].

Quality assessment inevitably includes the measurement of output or outcomes of care. Much of the disagreement with current quality measurement programmes comes from the potential inaccuracy of the benchmarking data or the limited ability of the 'outcome' measure to truly reflect the quality of a healthcare encounter. Traditionally mortality and morbidity have been used as outcome measures and therefore as proxy measures of overall quality of care. Their use has come about in part by the ease with which they can be measured and because they document a definitive end point to interventional treatment. Mortality, however, has a limited ability to discern differences in performance for operations where the mortality rate is low, such as appendicectomy or mastectomy. Even for operations where it is a more appropriate proxy measure for outcome, e.g. coronary artery bypass surgery, changes in the mortality rates over time mean that the 'strength' of mortality as an outcome measure will change accordingly and likewise the weight attributed to other measures of quality will change [39]. A quality framework will therefore need to attribute a 'weight' to a number of different quality measures and this will need to be adaptable over time.

A quality framework would therefore need to move away from the traditional approach of using predominantly outcome measures as a proxy of the overall episode of care and incorporate a number of 'quality' measures that are truly reflective of several dimensions of care within the entire treatment episode. These can be broadly defined as clinical pathway measures, using Donabedian's structure-process-outcome classification [14] and adding the fourth element healthcare economics, and patient-reported measures as outlined in Figure 1.2.

1.1.6 Measuring quality of care

1.1.6.1 Structural variables

The structure of surgical care can be thought of as the "bare bricks" or infrastructure of care. It is involved with details such as equipment, number of

beds, nurse-to-patient ratios, qualifications of medical staff, and administrative structure. It is thought that if surgery occurs in a high quality setting, then surely high quality care should follow. An advantage of measuring structural variables is that the information required is usually fairly reliable and used frequently at a hospital managerial board level. It is however infrequently used in a more clinical and or public domain to help inform the environment in which surgical care is delivered. Logically we need to be certain what correlation exists between these structural variables and quality of care, and this is not well established. Brook et al. have assessed the relationship between patient, physician, and hospital characteristics and the appropriateness of intervention for carotid endarterectomy, coronary angiography, upper gastrointestinal endoscopy.[40] They concluded that the appropriateness of care could not be reliably predicted from standard, easily obtainable data about the patient, the physician, or hospital structural variables. However, coronary angiography and carotid endarterectomy was significantly more likely to be carried out for medically appropriate reasons if performed in a teaching hospital. Hospital teaching status and other associated hospital variables such as size or for-profit status did not however translate into lower postoperative complication or death rates following carotid endarterectomy.[41]

A structural variable that has received more attention than most is the institutional or surgeon volume: the volume-outcome relationship. In this scenario, the volume of patients treated is used as a proxy for the quality of care and then the correlation to important clinical outcome measures determined. On the basis of a large number of studies that show better outcomes for patients treated at high-volume institutions and or by high-volume surgeons, we are seeing a trend of preferential patient referral to high-volume institutions. Promoters of this centralisation of services in the US argue that it is important to help advance the quality of healthcare, e.g. the Leapfrog group [42]. Similarly in the UK, centralisation of oncological services is identified in the Department of Health's Improving Outcomes Guidance framework [43]. Institutional and surgeon volume either independently or in combination are nevertheless rather broad proxy measures for quality of care. Indeed some low-volume providers have excellent outcomes and some high-volume providers' poor outcomes. As a result, research

in this area has started to improve its understanding of the core factors that determine whether or not the institution or surgeon produces better outcomes. Figure 1.4 illustrates potential structural variables for which volume acts as proxy measure and may therefore better inform us of quality of care.

In order to determine which structural variables potentially have the most influence on the quality of care, we first need to determine if correlation exists between them and some dependent end-point; research to date has used clinical outcome measures. Elixhauser et al [44] demonstrated the importance of the ratio of doctors and nurses per bed number, irrespective of the institutional volume, on the mortality rates for paediatric heart surgery. A systematic review published by Pronovost et al [45] also showed that high-intensity intensive care unit (ICU) physician staffing was associated with reduced hospital and ICU length of stay and mortality. Treggiari et al [46] demonstrated in a multicentre study that ICUs that were run by, or associated with, a specialised intensivist had significantly lower mortality rates in patients with acute lung injury (odds ratio = 0.68, 95% CI 0.52-0.89). This association was independent of severity of illness, and consultation by a respiratory physician.

As we better understand the core structural variables that correlate with markers of outcome, it will enable us to assess the degree to which they also influence overall quality of care. It is not unrealistic to imagine that integration of the structural variables, demonstrated to improve quality of care, into institutions irrespective of their caseload volume could further our aim of achieving equality of outcomes for all patients.

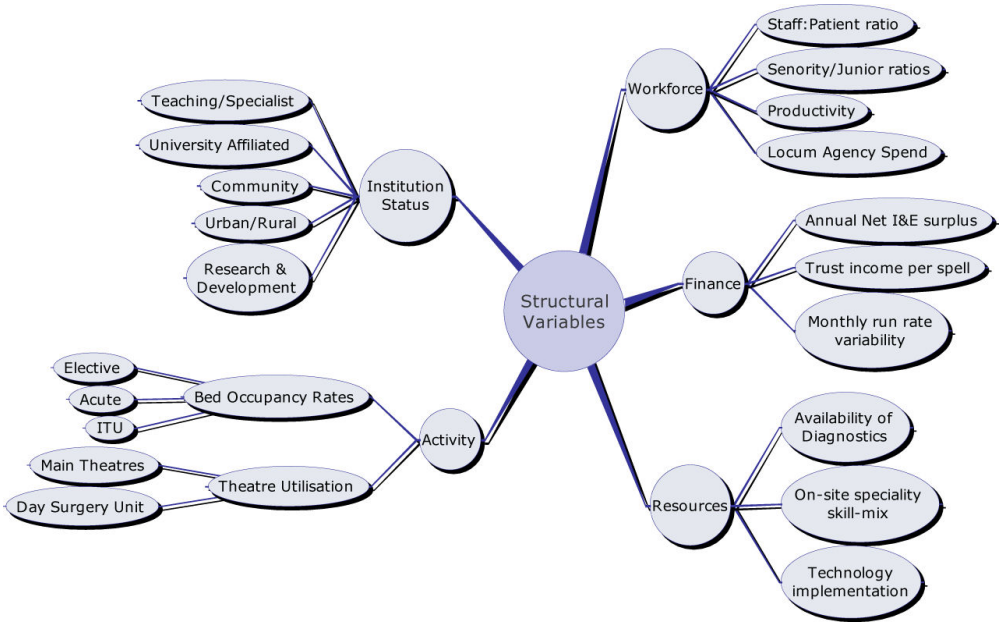


Figure 1.4 Examples of potential structural variables which could influence quality of care

1.1.6.2 Process measures

In surgery, process of care can be thought of as the pre-operative, intra-operative, and post-operative management of a patient. It looks at what is actually done for, and to, the patient. For example, it can look at the availability of screening programmes, the appropriate use of diagnostic tests, waiting times to operation, discharge processes, post-operative follow-up and availability of and willingness to give adjuvant treatments. However the measured processes are only useful if there is evidence to prove that they translate into improved patient care. There is little point, for instance, in ensuring that all patients prior to laparoscopic cholecystectomy have an MRI, when no clinical benefit will be gained. Malin et al. [47] assessed the quality of care for breast and colorectal cancer in the US. They reviewed existing quality indicators, guidelines, review articles, and peer reviewed clinical trials to produce a list of explicit quality measures evaluating the management of newly diagnosed breast and colorectal cancer patients. These were then checked for validity by a panel of experts, and included areas such as diagnostic evaluation, surgery, adjuvant therapy, management of treatment toxicity, and post treatment surveillance. Data was extracted from the patients' notes and via patient questionnaires. Overall adherence to 36 and 25 quality measures specific to the process of care was 86% (95% CI 86-86%) for breast cancer, and 78% (95% CI 77-79%) for colorectal cancer patients respectively. The unique element in this approach was that the group was not trying to correlate process with outcomes, but simply looking at process measures that were agreed, in this instance by evidence base and expert review, to reflect quality of care.

There are a number of potential benefits to the measurement of process as opposed to outcomes in assessing the quality of care. Lilford et al. [48] describe these in detail, but in brief; process measures are less susceptible to, although not exempt from, case-mix bias; their assessment and subsequent improvement will positively reflect on the entire evaluated institutional patient population as opposed to a few outliers with poor outcomes; deficiencies in processes of care are less likely to be seen as a label of poor performance, but instead indicate when and how improvement can be made; and process measures are reflective of the

current state of care, as opposed to the time-delay which is experienced with some outcome measures.

Process measurement is not easy. It is difficult to standardise measurements for process of care in surgery, as this process varies depending upon the surgical pathology. Creating a standard quality measure for all surgery may be impossible. It is more feasible to create measures of process of care for specific pathologies. Examples of where best practice has been defined include diseases with published national guidelines such as those produced by the National Institute for Health and Clinical Excellence (NICE) for cancer of the breast and lung. In the absence of agreed national guidelines, ongoing clinically based research will help us to define evidence-based processes that improve quality of care.

The appropriate use of surgical services is an important process measure in that it not only acts as a very good measure of the quality of care that a patient receives, but also has repercussion on the use of healthcare resources and the economics of healthcare. For a surgical intervention to be appropriate, it must 'have a health benefit that exceeds its health risk by a sufficiently large margin to make the intervention worth doing'. The RAND Corporation has published extensively on the topics of overuse and underuse of healthcare services. One of their largest studies examined the appropriateness of coronary angiographies, carotid endarterectomy, and upper gastrointestinal endoscopy across multiple regions of the US [49]. It found that 17%, 32%, and 17% of these procedures respectively were performed for inappropriate clinical reasons (i.e. overuse). Extrapolation from the literature may indicate that one quarter of hospital days, one quarter of surgical procedures, and two fifths of medications are inappropriately overused [50].

Measuring the process of care can however be incredibly labour and time intensive and will require significant clinical knowledge. There will be a multitude of measurement which can either be obtained prospectively, or gleaned retrospectively, from patient notes. The introduction of electronic coding of patient records may make this an easier task in the future. A recent Cochrane review did find that process measurements in the form of audit are effective in

improving clinical practice [51]. However the costs of measuring process will ultimately have to be weighed against the patient benefit that is gained from any actions taken as a result of those measurements.

In summary there is no doubt that for the majority of surgical conditions, measuring the process of care will provide us with contemporary indicators of quality of care which will be directly influenced by the functionality of a healthcare provider.

1.1.6.3 Outcome measures

Traditionally, quality of care has been judged on outcome measures using endpoints such as mortality and morbidity. They are used because they are easy to measure and recorded with regularity. For an outcome measure to be a valid test of quality it must be linked to and correlate with known processes that when changed will accordingly alter that outcome measure. For example, knowing the number of patients who present with metastases within six months of diagnosis with inoperable liver cancer is an important prognostic outcome. However, it is not a compelling measure of quality, as our ability to influence it is limited.

There are many advantages to using outcomes as a measure of quality of care. Outcomes are well established as an important feature of quality. They can be viewed as the overall effect of care on a patient. Few would doubt the validity of outcomes such as mortality in judging surgical care. Statistics such as mortality rates are understandable at face value, including to the lay person. Consequently there is a natural tendency to rank hospitals according to outcome measures such as mortality rates, with an implied association with quality of care. Examples of organisations that produce rankings according to outcome measures such as mortality rates include the Leapfrog group [30] and the US News "America's Best Hospitals" in the US [52], as well as Dr Foster Intelligence which produces the "Good Hospital Guide" [53], and the Healthcare Commission, now the care Quality Commission, in the UK [54].

The use, however, of outcome measures as the sole indicator of quality of care can be gravely misleading and inappropriate as outcomes are the 'end result' of an entire patient pathway and thus reliant on numerous other variables. The outcome measure itself may not therefore directly correlate with the quality of care. For example, if a patient, who has undergone an operation to remove a colorectal cancer, dies one year after surgery, can we say that he has had poor quality of care? He or she may have received all the best treatments available as guided by the latest evidence based medicine and despite all of this, died. Should their death be assumed to indicate a poor quality of care from the surgical team and allied healthcare professionals? Sometimes the best quality of care in surgery may still result in mortality through circumstances beyond our control. Equally though, the patient may have received the best quality of care throughout their hospital admission, but poor follow-up surveillance and delays in adjuvant treatments may have impacted on the final outcome. Other factors such as the natural history of the disease, the patient's age and co-morbidities often have much larger influences on outcome than surgical care. The method of risk adjustment attempts to compensate for the difference in case-mix between surgical centres. However, the effect of case-mix can never be completely eradicated. Firstly, risk adjustment can not allow for variables that are unmeasured, or not known. Neither can it adjust for the effects of varying definitions (such as the definitions of a surgical site infection) between centres. Risk adjustment can cause increased bias if the risk of that measured factor is not uniform across the compared populations.

This is why, even after adjusting mortality rates for risk, the relationship between quality of care and outcomes such as mortality is inconsistent. Pitches et al. looked at the relationship between risk-adjusted hospital mortality rates and quality or processes of care [55]. They found that a positive correlation between better quality of care and risk-adjusted mortality was found in under half of the papers examined, whilst the others showed either no correlation or a paradoxical correlation. Similarly, Hofer [56] found that the sensitivity for detecting poor quality hospitals based upon their risk-adjusted mortality rates was low at only 35%, whilst the positive predictive value was only 52%. This work has been corroborated with similar models by Zalkind [57] and Thomas [58].

Traditional outcome measures fail to appreciate important patient-specific measures such as quality of life. Recently, NICE has taken the quality adjusted life year into account against a negative financial outlay when deciding to recommend the use of novel oncological medications such as herceptin [59]. There has also been increasing interest in measuring national Patient Reported Outcome Measures (PROMS), as a way of measuring healthcare performance [60]. Successful pilot studies for PROMS data collection were completed in the UK and this has led onto a more formal national programme for PROMS within surgery [61].

Outcomes remain an important method of quality assessment despite their significant limitations around risk adjustment. The use of outcome measures in isolation is clearly inappropriate and in order to improve the use of outcomes as a measure of quality, a multidimensional approach including traditional measures such as mortality and morbidity, but also including more patient reported outcomes such as quality of life, pain scores and so forth, should be encouraged. This will help us to include more patient-centred measures into a currently clinically dominated quality of care assessment.

1.1.6.4 Healthcare Economics

High quality care needs to meet the "productivity challenge" of modern healthcare [5]. A limitation on available resources or financial constraint can be an inhibitor to producing the highest quality of care. With new technologies usually having initial premium costs, and increasing levels of demand from a more educated and aging population, healthcare costs will continue to increase in the future. In the US healthcare expenditure is determined by private insurance companies whilst in the UK, policy has given control to regional strategic health authorities. This 'local' budget control can, and has, led to geographical healthcare inequality. In the UK, this has been termed the 'postcode lottery': the treatments that you are eligible to receive can be dependent upon the area in which you live in accordance with the financial priorities of that area. Indeed the Department of Health has taken this one step further and begun to look at expenditure in a number of different areas of healthcare and correlated this to outcome data [62] (Figure 1.5).

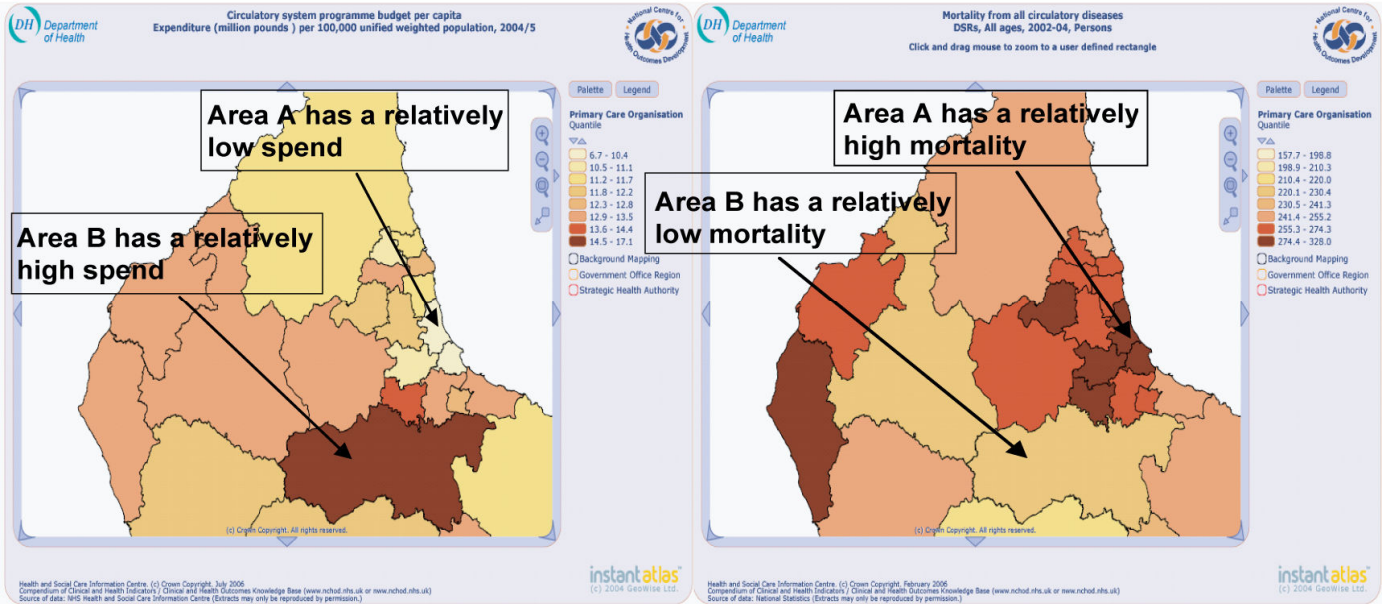


Figure 1.5 Programme Budgetting – Circulatory System programme budget per capita, expenditure (million pounds) per 100,000 unified weighted population, 2004/05 vs. Mortality from all circulatory diseases, DSR, All ages, 2002-04, Persons. Area A has a low spend per capita and a corresponding high mortality rate. The reverse is true for Area B (reproduced with permission from Department of Health/Crown copyright)

In economic terms, productivity is defined as the amount of output created per unit of input. Until recently, NHS productivity was determined using cost (input) and volume measures (output). Volumes of treatment measures such as GP appointments, ambulance journeys, and operations, taken from the National Accounts [63], were taken as clear indicators of how much work the NHS does. This is obviously an oversimplified approach, which ignores important aspects such as quality of care. Due to increasing costs, productivity has therefore been seen to fall in recent years between 0.6 and 1.3% per annum [63]. However, if NHS output is adjusted to account for increased quality of care as well as the increasing value of health, NHS productivity actually demonstrates an increase in productivity from 0.9 to 1.6% per annum [63 64]. Thus we can see how understanding quality of care can have economic benefits as well as increasing public satisfaction.

Although new surgical technologies are usually associated with a higher cost, this cost can at times be counterbalanced by subsequent benefits. A good example of this is the advent of laparoscopic surgery. Given the higher price of laparoscopic equipment compared to standard equipment, along with the surgical learning curve and at times increased duration of procedures, you would be forgiven for thinking that laparoscopic procedures were invariably associated with a higher cost of treatment. However, as shown by Hayes et al. [65] although the initial cost of procedures such as laparoscopic colectomy can be higher, there are overall improvements in cost effectiveness due to savings in reduced recovery days and quality-adjusted life years. Similarly, introducing a dedicated clinical pathway for procedures such as laparoscopic cholecystectomy can provide further cost advantages [66]. The decision-making of the individual surgeon is central to healthcare costs. Using Kissick's decision-making model, Fisher et al have demonstrated that using faecal occult blood testing as a primary screening tool for colorectal cancer can give similar sensitivity to that of colonoscopy, whilst significantly improving access with huge cost savings [67].

These examples show that improving the quality of care that a patient receives, by simply improving the efficiency of healthcare delivery or using evidence-based practice, can result in additional economic benefits. Providing high quality care

does not necessarily have to cost more and may also facilitate the alignment of financial and clinical incentives.

1.1.6.5 Patient-reported treatment outcomes

Patient-reported measures can be classified into patient-reported treatment outcomes, health-related quality of life and patient satisfaction. They are not mutually exclusive, but each provides us with a different dimension to quality assessment.

Patient-reported treatment outcomes (PRTO) greatly enhance quality of care assessment by reflecting the patient's viewpoints on treatment outcomes and typically provide information on symptoms and/or functional status. In an ideal system, PRTO should be integral to clinicians' decision-making processes and the assessment and appropriateness of the care provided. PRTO are especially important within oncology where several different treatment options may exist, and where survival gains can be small with significant treatment side-effects [68]. Numerous questionnaires, across various diseases and treatments, have been designed and clinically validated to record PRTO. The Patient Reported Outcomes & Quality of Life Instruments Database (PROQOLID) provides a comprehensive list of PRTO and quality of life measures [69]. Similar resources will stimulate future benchmarking of PRTO and can be used as an international platform to develop metrics for assessing the quality of patient pathways. Several criteria can aid the selection of the most suitable PRTO measurement tool, Table 1.3.

1.	Appropriateness	Is the instrument content appropriate to the questions which the application seeks to address?
2.	Acceptability	Is the instrument acceptable to patients?
3.	Feasibility	Is the instrument easy to administer and process?
4.	Interpretability	How interpretable are the scores of the instrument?
5.	Precision	How precise are the scores of the instrument?
6.	Reliability	Does the instrument produce results that are reproducible and internally consistent?
7.	Validity	Does the instrument measure what it claims to measure?
8.	Responsiveness	Does the instrument detect changes over time that matter to patients?

Table 1.3 As suggested by the Patient-Reported Outcomes Measurements Group at the University of Oxford, consideration needs to be given to eight necessary criteria to select the most suitable patient-reported outcome measures instrument(s). Adapted from online information available at http://phi.uhce.ox.ac.uk/inst_selcrit.php.

1.1.6.6 Health related Quality of Life

Health-related quality of life (HRQOL) measures must include an evaluative component that assesses impact on general well-being; this can include impact of PRTO. HRQOL measures can either be unidimensional or reflect multiple domains of impact on well-being (multidimensional) and typically include physical, psychological and social components [70]. HRQOL measures can be generic, disease-specific or treatment-specific. HRQOL assessment tools were developed over 20 years ago and initially focused on patients with cancer. The latest version of the Functional Assessment of Cancer Therapy-General (FACT-G) has 27 general questions divided into four primary HRQOL domains: physical well-being, social/family well-being, emotional well-being, and functional well-being [71]. Other cancer-specific assessment tools include questions that relate directly to surgical intervention, such as social embarrassment of stomas (FACT-C). Generic, non-disease specific, HRQOL measurement tools are available and include EQ-5D [72] and SF-36 [73].

The importance attributed to HRQOL measures has been highlighted by NICE by taking the quality adjusted life year into account against a negative financial outlay when deciding to recommend the use of novel oncological medications such as herceptin [59]. As highlighted by Osoba [74] however, the "science of HRQOL assessment has not yet been adequately tested in clinical practice". This is now feasible and, following a successful feasibility pilot in the UK, all English hospitals were required to collect HRQOL data for four surgical procedures, starting in 2009 [61].

1.1.6.7 Patient Satisfaction

Similar to PRTO, the measurement of patient satisfaction is crucial in quality of care assessment as the patient's viewpoints on treatment may be completely different from the viewpoints of a healthcare professional. Patient satisfaction can be divided into two areas; satisfaction determinants - patient variables that can affect overall satisfaction (patient expectations, patient characteristics and psychosocial determinants) - and satisfaction components which refer to a

measure of the patient's perceptions of care actually received (e.g. interpersonal manner, accessibility and convenience, continuity of care and physical environment of care) [75]. Satisfaction determinants naturally influence satisfaction components and for this reason, measurement of satisfaction needs to incorporate both.

The measurement of patient satisfaction is far from simple. When designing satisfaction surveys, four parameters need to be considered that will influence the results; choice of population, timing, type of questionnaire, and the rating of satisfaction [76].

1.1.7 Future direction

The greatest advances in the area of quality assessment will be in the realm of measurement of process and overall performance. Inclusion of structural variables will also need to be considered. This should not be confused with set performance targets, or blindly following clinical guidelines. These newly developed assessment scores should allow us to implement changes that will not only improve that score, but more importantly improve the quality of care delivered to our patients. For example a performance measure that tells us a certain proportion of the population underwent a particular desired process is not enough. It does not give us any indication on how to improve quality. It is important to know why certain members of the population failed to achieve this goal. Was it through contra-indications to that process, lack of communication, lack of compliance or something else? This information can give further understanding of what changes are needed to improve the service that is provided. In short, knowing that improvement is needed is important, but more helpful is the knowledge of how to improve.

Engagement in this process by institutions and clinicians is crucial. They have been reluctant to date as the public reporting of performance data has had a 'name-and-shame' style by inappropriately ranking them against each other without duly considering institutional variations that cannot be adjusted for, but which result in explicable varying performance. Better methods of visually

presenting performance data that avoid arbitrarily ranking healthcare providers, that are interpretable at face value to the lay person, and which still continue to identify Trusts which need special attention, will help to engage all stakeholders in future performance benchmarking.

1.1.8 Public health implications

The assessment of quality of care is a public health issue which is becoming a dominant theme in structuring modern healthcare. The rigorous and accurate measurement of quality is an essential component for the improvement of public health services, and answering public accountability. The methods by which quality is assessed has the potential to dictate healthcare policy well into the future, and as practicing surgeons, we must all be well educated on this topic. Some examples of the current assessment of quality of care can be gathered from the internet sources listed in Table 1.4.

The field of cardiothoracic surgery has long been aware of the push towards quality improvement and public accountability. In the 1980's, the Society of Thoracic Surgeons (STS) initiated one of the largest data collection operations in medicine, resulting in the STS National Adult Cardiac Surgery Database. It is now the largest and most comprehensive single specialty database in the world. It allows not only surgeons and Trusts to compare their results, but is also freely publicly available and patients can identify their own surgeon's outcomes. With increasing emphasis placed upon quality and performance measurement, the STS set up the Quality Measurement Task Force (QMTF) to create a comprehensive quality measurement program for cardiothoracic surgery. The results of this program have been published [77] [78] and represent the most up to date and rigorous methods by which quality assessment can be performed.

Undoubtedly, with further investigation and reporting of the factors driving quality of care, inequality of healthcare provision will be uncovered. No one doubts that all patients should have equity of quality of care and it can result in more lives saved. The Leapfrog group in the US now recommends that there is a certified critical care specialist available for their ICUs, and estimated that this

restructuring could save more than 54,000 lives in the US per year [79]. But can the current healthcare infrastructure manage geographical fluxes in demand that may result from patients mobilising their freedom of choice and seeking out 'better care'? Often institutions that are currently able to provide higher quality healthcare can do so only under the constraints of their current patient population demand. Any reasonable increase in this demand can have a negative impact and subsequently lead to a worsening of the quality of their healthcare provision.

1.1.9 How to design healthcare quality reforms

Objectives to improve quality of care are well recognised, but the implementation of systems in order that these objectives are met, is far from straight forward. Translating research evidence into the everyday structure and processes of a healthcare system is feasible, but made difficult by the variation that exists across healthcare systems and between healthcare providers. Leatherman et al. [80] describe a conceptual framework to facilitate the designing of health system reforms that consists of three aspects:

- A taxonomy to organize the available evidence of potential quality enhancing interventions (known as the QEI project)

- A multi-tiered approach to select and implement interventions in a health-care system at four levels: national, regional, institutional and the patient–clinician encounter

- A model to guide the adoption of a balanced portfolio approach to quality improvement—recognizing the prudence of simultaneously employing professional, governmental and market levers for change.

Organisation	URL:
The Leapfrog Group	**http://www.leapfroggroup.org/**
The National Surgical Quality Improvement Program (NSQIP)	**https://acsnsqip.org/login/default.aspx**
The International Quality Indicator Project (IQIP)	**http://www.internationalqip.com/**
The Healthcare Commission	**http://2007ratings.healthcarecommission.org.uk/homepage.cfm**
The Institute for Healthcare Improvement (IHI)	**http://www.ihi.org/ihi**
Agency for Healthcare Research and Quality (AHRQ)	**http://www.ahrq.gov/qual/measurix.htm**
The National Committee for Quality Assurance (NCQA)	**http://web.ncqa.org/**
The Institute of Medicine (IOM) Health Care Quality Initiative	**http://www.iom.edu/CMS/8089.aspx**
The National Association for Healthcare Quality (NAHQ)	**http://www.nahq.org/**
Quest for Quality and Improved Performance (QQUIP)	**www.health.org.uk/qquip**

Table 1.4 Internet Resources for Quality of Healthcare

The QEI project encompasses several aspects of quality improvement, such as effectiveness, equity, patient responsiveness and safety. It itself forms part of a wider initiative called the Quest for Quality and Improved Performance (QQUIP); a 5-year international collaborative research project between the University of North Carolina, School of Public Health, London School of Economics, University of York and University of Cambridge. The limitations of using evidence-base to bring about health reform are recognised, such as publication bias and difficulties in translating evidence from one healthcare system into another, but early results of the QEI project are generating some good examples of focused quality interventions.

The integration of evidence-based interventions needs to occur across all levels of a healthcare system in order that predictable systemic improvement in quality arises. This "multi-tiered approach to building predictable systemic capacity for improvement" describes three key factors; 'horizontal coherence', the interaction of several different types of quality interventions; 'vertical coherence', the interaction of a quality improvement intervention across the multiple levels of the healthcare system; 'coherence in accountability', the balance between professionalism and professional accountability, centralised governmental control and market forces. Coherence in accountability forms the components for a "balanced portfolio approach to quality improvement" and recognises that individually professionalism, government or market factors cannot generate sustainable quality change.

1.1.10 How to achieve healthcare quality improvement

As highlighted by Glickman et al.[81], there has been proportionally more attention directed towards the 'process' and 'outcome' components of Donabedian's structure, process, outcome framework for quality. In today's modern healthcare system, 'structure' consists of important organisational and managerial components that are the enablers for driving forward multidimensional quality improvement agendas. Glickman et al. describe these organisational characteristics from a management perspective; executive management, including

senior leadership and board responsibilities, culture, organisational design, incentive structures and information management and technology.

The distinctive aspect to this work is the combination of business and medical viewpoints to provide a contemporary operational definition of structure that updates Donabedian's 'physical characteristics' description. This framework engages managerial capabilities crucial to achieving healthcare quality improvement.

1.1.11 Conclusions

Assessing quality of care in surgery is an important and essential part of maintaining and improving patient care. The very act of measurement serves to determine current standards and provides a baseline against which improvement can be made locally and/or nationwide. Benchmarking between providers will assist in identifying inequalities at a provider or regional level. Further investigation of the causative factors will discover pockets of best practice and local innovation that can then be disseminated more widely. Although traditional assessments of quality have been heavily influenced by a number of clinical outcome measures, such as mortality and morbidity, we have shown that these are clearly inadequate in isolation and do not provide a reliable assessment of the quality of surgical service.

As described by Donabedian some 40 years ago, quality of care can be explained by three key elements; structure, process and outcome. Treatment outcome measures will still form an important part of quality assessment as they are easily understandable to the clinician and patient alike, and outcomes such as postoperative mortality remain an important endpoint. We will see expansion of the use of patient reported outcomes such as quality of life, and current health status, in order to achieve a well rounded viewpoint on quality care.

Measurement of structural and process of care variables must be used in combination with outcome measurements and have the significant advantage that they are less influenced by factors such as case-mix and patient co-morbidities.

This will help to overcome the methodological difficulties of producing suitable adjusted data. These structural variables and process measures are not, however, currently widely or routinely collected and it will require a labour-intensive undertaking, and undoubtedly require additional resources, in order for this to change.

Healthcare economics is a further key element to assessing quality of care and has significant impact upon modern surgical care which is heavily influenced by continually evolving technological and biosurgical innovation. A lack of available finances will always act as an inhibitor to delivering the highest quality of care. The combination of surgical innovation, making more people eligible for treatment, with an aging and increasingly demanding public, means that financial constraints will remain a considerable factor for the foreseeable future.

A quality of care assessment tool should be multi-factorial, taking into account the entire patient treatment episode. It should include up-to-date process measurements gleaned from evidence based medicine and national guidelines. It should consider patient-centred as well as disease-specific clinical outcome measures and incorporate structural variables indicating effective and efficient healthcare delivery. In this way we can be confident that we can obtain the most accurate and valid assessment of quality of care in surgery.

1.1.12 Key points for appraising the quality of care in surgery

- The assessment of quality of care in surgery is an important and essential part of improving patient care.

- Practising clinicians must be aware of the advances in quality of care assessment, as they have the potential to drive healthcare policy well into the future.

- Three key determinants of quality of care have been defined; structure, process and outcome.

- Traditional measures of quality have concentrated mostly on outcome measures such as morbidity and mortality figures. However outcome measures alone are inadequate to accurately assess quality of care.

- Process of care measures are fast becoming a major method of assessment of quality care, as they can provide up-to-date assessments that are founded in evidence based surgery.

- There are already established programmes using process of care measures to benchmark standards and reduce inequality in surgical care. These programmes are generic and give us only an indication of minimum standards.

- In the future, quality of care assessments should be pathology-specific, taking into account the structures, processes and outcomes of the whole patient episode of care, and combining them with patient-centred measures such as quality of life and patient satisfaction.

- The translation of available research evidence into everyday healthcare provision can be the limiting factor to 'frontline' quality improvements. It can be facilitated by conceptual frameworks for designing health system reforms and engaging contemporary managerial capabilities.

1.2 The role of volume-outcome relationship in surgery

It has long been postulated that improved outcomes in healthcare can result from treating greater numbers of patients and is explained by "practice makes perfect" [82]. Over the last 10 years, we have seen an acceleration of more formal research in this area and particularly within the surgical specialities as a result of their interventional nature. Numerous studies have reported on a higher volume, better outcome association with the implication that the quality of care that patients receive can be greatly influenced by this relationship. Despite the considerable evidence, many uncertainties remain about the true relationship. Some low-volume surgeons and or institutions have excellent outcomes and some high volume surgeons and or institutions have poor outcomes. An alternative explanation for the volume-outcome relationship may therefore be that healthcare providers, either at an institutional level or surgeon level, that display better outcomes receive more referrals and as a result treat greater volumes of patients. This is known as the selective-referral hypothesis [82]. Volume, therefore, does not automatically result in better outcomes for patients and as such is an inexact indicator of quality of care. On its own, volume is acting as a surrogate marker for the other numerous and complexly interacting factors within a patient's treatment episode that combine to determine their outcome, favourable or not (Figure 1.6).

We are seeing an increasing trend towards more detailed performance monitoring of surgeons and surgical healthcare providers. More so, this collated data is not solely for 'internal' consumption, but is being made publicly available to allow the better informed patient to exercise a degree of choice as to where they receive their treatment. A number of organisations, including the Healthcare Commission previously and now the Care Quality Commission [83] and Center for Medical Consumers [84], produce data at a surgeon and institutional level which report on number of operations performed and outcomes of those operations, although a direct link between the two is not advertised.

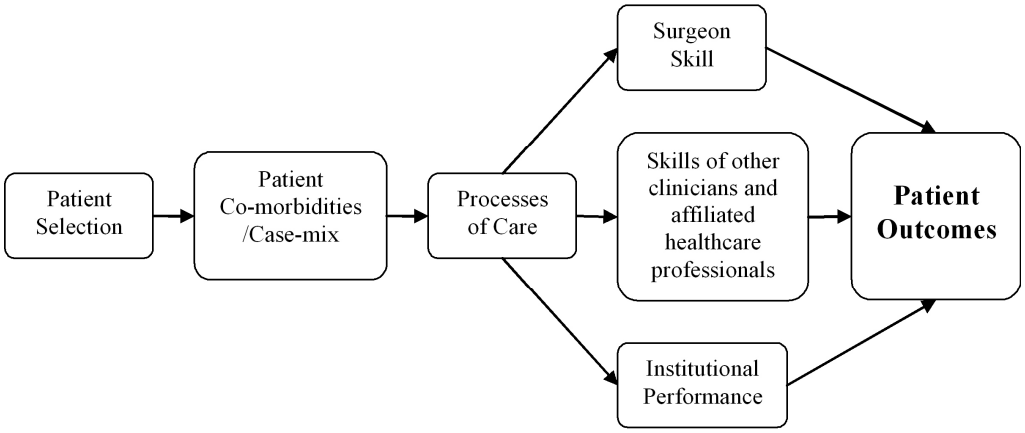


Figure 1.6 Conceptual framework of how patient outcome following healthcare intervention is determined by interaction of several dependent components (Redrawn from reference [85])

There are vast implications to the patient and healthcare providers alike of acting upon the published evidence in this area. When clinicians are appraising the volume-outcome relationship literature, in order to practise evidence-based surgery, three fundamental questions need to be answered [86], Table 1.5.

This section of the introduction will discuss the methodological basis for assessing the volume-outcome relationship in surgery and some of the limitations that surround it. It will explore the commonly used outcome measurements with their possible alternatives and consider the interaction between the surgeon's volume and institutional volume and their effect on patient outcome. Finally, the health policy implications of incorporating volume-outcome relationship research into health service provision will be explored.

1) Are the results of the study valid?

Is the database accurate?

How was the volume determined?

Was the volume analysed as a continuous or a categorical variable?

If categorical, were volume groups determined a priori?

Is the primary outcome appropriate?

Are patients the same among volume groups? How were the data analysed?

Was multivariable analysis used to adjust for important prognostic differences?

Was a sensitivity analysis performed to test for statistical robustness?

2) What are the results?

What was the magnitude of the results?

How precise was the estimate of the treatment effect?

3) Will the results help me care for my patient?

Table 1.5 Three fundamental questions that need to be answered when appraising the volume outcome literature.

1.2.2 Methodological framework for assessing volume-outcome relationship

1.2.2.1 Data sources & Data quality

The majority of research exploring the volume-outcome relationship in surgery has relied on large administrative databases for its source data, as extracting information direct from patient charts would be too time consuming and expensive. The inferences that we make from this research are therefore solely reliant on the accuracy of these databases and it comes back to the old adage; 'what you put in, is what you get out'. The nature of healthcare funding in the US by provider reimbursement through private health insurance companies means that the size and comprehensiveness of the associated databases makes them the most exploitable for volume-outcome research and as such the majority of publications in this field originate from the US. The administrative databases such as The Medicare Provider Analysis and Review Files and the Healthcare Cost and Utilization Project Nationwide Inpatient Sample can be linked to epidemiological databases such as Surveillance, Epidemiology, and End Results (SEER) to improve the sophistication of the data extraction, although this use of administrative data as opposed to clinical data inhibits the degree of risk adjustment that can be performed. Equally some administrative databases derived from managed care plan enrolees, although encompassing thousands of patients, may not be representative of the general population by virtue of only including patients older than 65.

In England and Wales, Hospital Episodes Statistics (HES) is a database routinely populated within the health service for administrative purposes and not specifically for clinical audit. It has contained admitted patient care data since its inception from 1986 onwards. Its use as a data source in health service research has been limited by worries over the completeness and accuracy of the data input at a patient coding level. However, over time the quality of the HES database has improved and particularly so since the introduction of the 'payment by results' initiative in 2002 as a means to provide a transparent, rules-based system for paying healthcare providers, which is linked to activity and adjusted for case-mix. As a result, it is being used more frequently as a more reliable source database for

exploring the volume-outcome relationship in surgery in England. Indeed there is now evidence to suggest that HES has similar discrimination to clinical databases when used in risk prediction models for death [87].

The availability of centrally collated administrative databases should not distract from the usefulness of data originating from a single centre or close network of centres, which tend towards relatively smaller caseloads. The ability in these settings to more closely quality control the data acquisition and subsequent risk adjustment and analysis overcomes some of the issues of administrative databases.

1.2.2.2 Data presentation

Although studies have demonstrated a correlation between volume and outcome, it is unclear whether the relationship is continuous, step-wise, or has a single clear cut-off (Figure 1.7). Until this has been determined, the optimal means of displaying data cannot be established. Handling volume as a continuous variable will improve the chance of detecting an outcome difference along the volume gradient and equally may reveal a cut-off volume after which there is no further change in outcomes achieved. Presenting volume as a continuous variable is typically done by scatter graph. Although methodologically superior, visual difficulties can occur when apparent correlations are seen, but are not statistically proven because of a large number of low volume providers with a zero outcome, that are not visually obvious [88]. Volume is often therefore analysed and displayed as a categorical variable. Under these circumstances, it should be handled as a minimum of three volume groups, of approximately equal size, which should be determined prior to analysis. Display is then in the form of histogram and importantly with confidence intervals displayed.

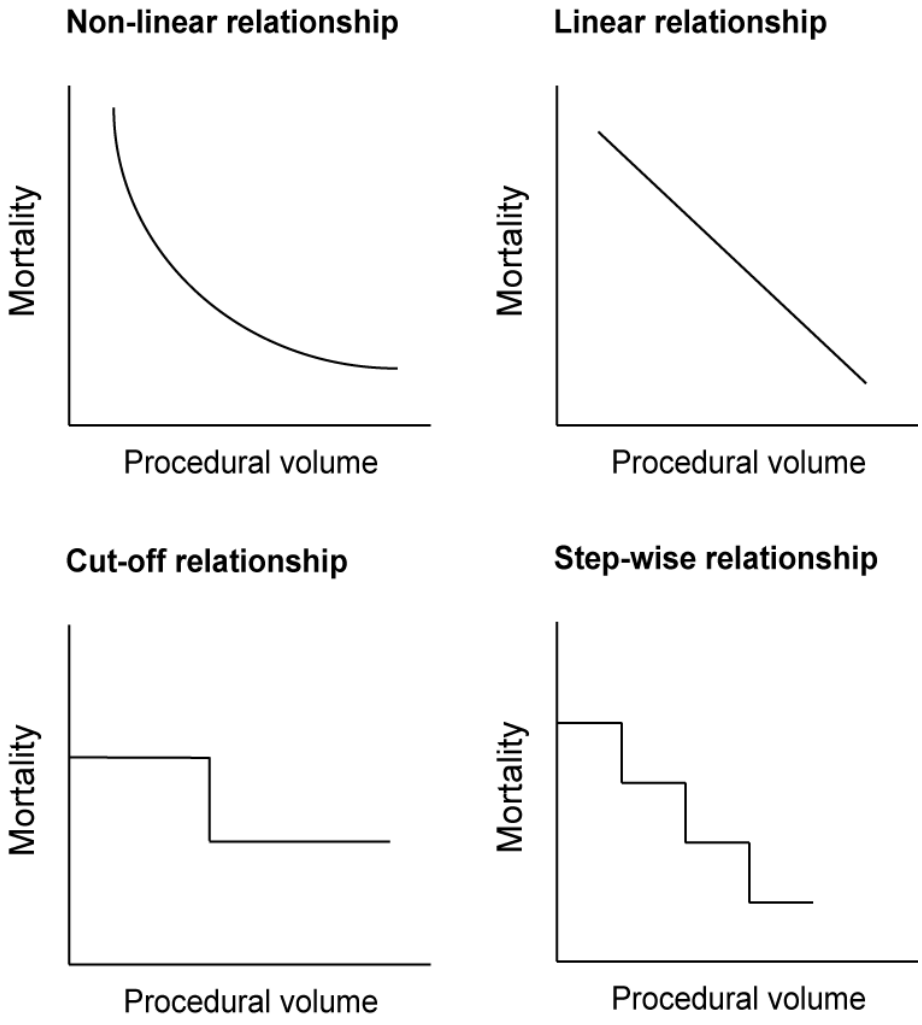


Figure 1.7 Possible relationships between volume and outcome. Reproduced from reference [88]

Funnel plots have been used extensively to assess the quality of aggregated data in meta-analysis. This requires plotting a measure of study precision against treatment effects. They are therefore ideal for making an initial assessment of volume of cases (measure of precision) against outcome. The most public example of this in England surrounded the independent Bristol Royal Infirmary Inquiry into paediatric cardiac deaths. By plotting the mortality rates in under 1s for paediatric cardiac surgery against caseload undertaken for each of 12 English Hospitals, it becomes evident that there is an appreciable decrease in mortality rate as the total number of operations performed increases, (Figure 1.14). A formal statistical test of the volume-outcome relationship can then be performed using regression analysis with an appropriate error structure. This is described in more detail by Spiegelhalter [89]. The basic structure of a funnel plot is no more than a scatter graph. Unique is the construction of control limits as a function of the measure of precision which do not depend on the data being plotted. They are formed by calculated confidence intervals and typically represent 2 and 3 standard deviations. As the volume of cases (denominator) increases, the confidence interval narrows and a 'funnel' shape forms. Any plotted data point that lies within the control limits is said to be acting under common cause variation and as such its performance is within acceptable normal variation. Any data point lying outside of the control limits (see Figure 1.9) is acting under special cause variation and as such is a 'true' outlier. The implications of this are that such outliers need to be carefully considered before inclusion in volume-outcome correlations as they will inevitably have a significant impact, which may only be the result of extenuating circumstances (e.g. methodological errors). The application of funnel plots in surgery is further discussed later in the introduction.

1.2.2.3 Methodological limitations

The quality of the data and methods used to analyse it will necessarily affect the results and substance of any conclusions. The quality of commonly used data sets has already been discussed and is generally well recognised in the literature. More recently there has been a growing awareness of the methodological limitations of data handling and presentation. The majority of studies exploring the volume-

outcome relationship in surgery support an inverse relationship, such that the greater number of cases performed by a healthcare provider improves the patient related outcome, typically reported as mortality. The magnitude and degree of statistical significance of the reported correlation between studies does however vary considerably and much of this variation can be attributed to heterogeneous study design. The IOM held a workshop in 2000 to review the current understanding of the relationship between volume of health services and health-related outcomes [90]. As part of this, Halm et al created a quantitative method of assessing the research design of volume-outcome studies, such that higher scores would reflect increasing likelihood of the study's ability to discern generalisable conclusions about the nature and magnitude of the relationship between volume and outcome. The designed scoring system assessed 10 integral methodological criteria including representativeness of the dataset, risk adjustment, clinical processes of care assessment and methods of volume analysis [85], (Figure 1.8).

The authors then scored 88 studies extracted by systematic review of the literature. Possible total quality scores ranged from 0 to 18. The mean total quality score was 7.8 ± 1.9, with a median score of 8 (interquartile range 6 to 9). Only 18% of studies achieved a score greater than 10. A similar process was subsequently repeated for studies assessing volume-outcome relationships for oncological operations between 1984 and 2004 [91]. Again no study scored greater than 11 out of a possible 18 with the vast majority scoring between 7 and 10. The implication of this finding, that at best existing study design is only modest, limits the transferability of the reported correlations to clinical practice. It also raises the question whether much of the volume-outcome relationship will be voided by more methodologically robust research?

Question	Characteristic	Score
1.	Representativeness of sample:	
	Not Representative	0
	Representative	1
2.	Number of hospitals or Surgeons	
	Hospitals < 20 and Surgeons < 50	0
	Hospitals ≥ 20 or Surgeons ≥ 50	1
	Hospitals ≥ 20 and Surgeons ≥ 50	2
3.	Total sample size (cases)	
	< 1000	0
	≥ 1000	1
4.	No. of Adverse Events	
	<20	0
	21-100	1
	>100	2
5.	Unit of analysis	
	Hospital or Surgeon	0
	Both Separately	1
	Both Together	2
	Both Together + further component	3
6.	Appropriateness of patient selection	
	Not measured	0
	Measured separately	1
	Measured and analysed	2
7.	Volume	
	Two categories	0
	Multiple categories	1
8.	Risk Adjustment	
	None	0
	Administrative data only	1
	Clinical data only	2
	Clinical data + C >.75 and H/L test positive	3
9.	Clinical processes of care	
	Not measured	0
	One	1
	≥ 2	2
10.	Outcomes	
	Death only	0
	Death + other	1

Figure 1.8 Scoring system for rating the quality of research on volume-outcome relationship. Adapted from reference [85]

Volume, being easily measured, is used as a proxy for the expertise of a surgeon or institution. It is assumed that any volume categorised differences in outcome are the result of better surgical practices resulting from undertaking more cases and so improving expertise. Many studies use standard statistical methods that assume patients to be independent observations, whereas the outcomes of patients treated by the same provider (irrespective of volume treated) are likely to be similar. Indeed the potential for this phenomenon, known as clustering, to occur is greater the smaller the provider caseload. Similarly, standard statistical methods do not consider that outcomes between providers with similar volumes can vary considerably. These factors will serve to falsely accentuate any derived volume-outcome relationship. An example of this was published by Panageas et al. [92] who performed a reanalysis of three previously published volume-outcome studies using statistical methods for analysing clustered data; a random-effects model and generalized estimating equations. They demonstrated that for colectomies, prostatectomies and rectal cancer surgery attenuation of the volume-outcome relationship occurred, depending on the outcome measured, when adjustment was made for clustering in addition to case-mix and volume of operations performed. The demonstration of clustering on its own is an important entity and the authors provide an example of a simple but plausible explanation "Some colon cancer surgeons are more likely to perform colostomies (thereby potentially avoiding anastomotic leaks and postoperative infections), while others are substantially more likely to attempt primary re-anastomosis. This variation in surgical practice leads directly to observed variation in these outcomes when analyzed on a surgeon-by-surgeon basis". Consideration of the presence of clustering is therefore an important aspect to volume-outcome relationship assessment and will go some way to identifying unexplained variation in outcomes. It likely reflects important differences in processes of care that will provide opportunities for improving quality of care.

The variability in the observed outcome will naturally be greater the fewer times it occurs, i.e. for low volume providers. For high volume providers the observed outcome, having occurred a sufficiently large number of times, (as determined by the underlying true rate of occurrence), is likely to be more representative of the

underlying true rate and thereby display less variability. So when analysing and presenting crude observed outcomes, there will be bias against low volume providers as the observed outcome will vary substantially from the true underlying outcome rate, often as the result of a few outliers. Some of this bias can be removed by considering the observed outcome as a rate of the expected outcome, a calculated estimate of the true outcome rate. As applied to mortality, this ratio of observed over expected mortality is known as the standardised mortality ratio. The calculated expected mortality rate can be adjusted for a number of confounding variables such as age, gender, race, co-morbidities, deprivation scores and volume and as such improve the robustness of any remaining volume-outcome relationship.

1.2.2.4 Outcome measures

Morbidity/Mortality/Length of Stay/Re-admission rates

One of the most important aspects to researching the volume-outcome relationship is identifying the outcome to be assessed. Availability of measures recorded consistently and reliably in administrative databases has meant that mortality, either inpatient or 30-day, has predominated. Mortality is however not always the best outcome measure and indeed may result in the absence of any volume-outcome relationship, when one exists for more procedure-specific outcomes [93]. Mortality is unlikely to be useful as an outcome measure if the mortality rate for the operation of interest is too low to discern differences between high and low volume providers. This phenomenon likely explains why the volume-outcome (mortality) relationship appears strongest for complex operations such as oesophagectomy and pancreatectomy and weakest for operations such as colectomy and carotid endarterectomy, in the existing literature. Depending on the operation of interest and its associated mortality rate, procedure-specific morbidity could be a more appropriate outcome measure to detect a volume-outcome relationship. Indeed there is evidence to suggest that among high-volume surgeons, those who performed well for one morbidity endpoint, performed well in others [94].

Mortality measured as long-term survival may overcome the problems described above for operations with low mortality rates. Needless to say it is reflecting very different quality components as compared to 30-day mortality. Long-term survival will be affected by surgical technique, e.g. positive margin rates for oncological procedures, follow-up diagnostics, availability and thresholds for giving adjuvant treatment, to name a few. As a result, operations which demonstrate no volume-outcome relationship because of their low 30-day mortality can display volume-outcome relationship for long-term survival [95].

Hospital lengths of stay and 30-day re-admission rates are other important outcome measures that were infrequently assessed [96]. In the last three to four years we are seeing them more frequently included [97]. They act as important outcome measures reflecting the efficiency of post-operative care pathways and intensity of discharge planning. Besides the disadvantages to the patient, prolonged length of stay and high re-admission rates carry a significant financial burden and are therefore politically motivated. Emergency re-admission rates can be a clinically useful indicator, especially when the risk of death following surgery is negligible. This is particularly true in the day surgery setting, or the elective setting where the severity of a suitable condition of interest displays little variance. The studied populations, as a result, tend towards homogeneity and assessment of the re-admission rate as an outcome measure allows comparison of like-with-like units of interest [98].

1.2.3 Limitations of case-mix adjustment

It is clearly advantageous to adjust outcome data for as many variables as possible thereby eliminating confounding explanations for demonstrated correlations. Limitations in data availability will naturally limit the degree of adjustment. Administrative databases as the main data source usually record data that allows adjustment for age, sex, ethnicity, socio-economic status, method of admission and hospital size (teaching status). Adjustment for patient co-morbidities is less readily available, albeit arguably the most important. It has been reasoned in terms of methodological quality that risk adjustment only from administrative databases is of relatively inferior quality and should originate from medical records or

prospectively designed clinical registries [85]. The resources that would be required in order to fulfil this when dealing with thousands of administrative patient records in a retrospective fashion generally make this impractical. In order to circumnavigate this issue, much work has been done on manipulation of administratively recoded co-morbidity variables. Using predictive modelling incorporating the Charlson co-morbidity score, it has been shown that the UK based routinely collected HES administrative database can be used to predict risk of mortality following surgical intervention with similar discrimination to national speciality-specific clinical datasets. In this regard, performance suitably adjusted for patient case-mix can be determined [87].

In addition to adjusting for patient orientated variables, the processes of care that a patient experiences as they progress through their treatment episode will clearly impact on the end outcomes. The variability that exists has been clearly demonstrated, using patient pathways for treatment of colorectal and breast cancer as example. The study analyses adherence to quality measures incorporating clinical domains representative of the entire patient episode: diagnostic evaluation, surgery, adjuvant therapy, management of treatment toxicity and post-treatment surveillance. Further examination was made of eight components of care integral to these clinical domains: testing, pathology, documentation, referral, timing, receipt of treatment, technical quality and respect for patient preferences. Although overall adherence to these quality measures was high, significant adherence variability was identified (13 to 97% for breast cancer and 50 to 93% for colorectal cancer) [47].

As the quality of administrative databases improve both in terms of data acquisition and handling, so will our ability to adjust for a number of confounding variables. Statements concerning the existence of volume-outcome relationships will as a result be more robust in nature.

1.2.4 Quality of life measures

Outcome measures such as mortality, morbidity and length of stay are used frequently because of the ease with which they can be measured and the

abundance of data held in administrative databases. Studies have, where possible, assessed endpoints that directly affect a patient's quality of life, such as long-term incontinence following radical prostatectomy [93]. There is, however, a paucity of studies which use validated health surveys incorporating patient values using qualitative indicators (SF 36, EQ 5D). More recently, there is awareness that outcome measures which better assess the quality of life of patients following surgical intervention can add an important aspect to any outcome assessment tool. The Department of Health initiative for all English hospitals to collect Patient Reported Outcome Measures across four surgical procedures [61], signifies an important shift away from centrally collating and reporting 'pure' clinical outcome measures, to ones that encompass patient-orientated experiences of their treatment.

1.2.5 The influence of the Surgeon and/or Institution

Surgeons will care for their patients in ways that they feel appropriate. Although this will be affected by the institution's structure in which they work, they may have had the opportunity to partially influence the structure themselves. The surgeon and institution could therefore be seen as a single influencing factor due to their integrated nature and yet they reflect very different influences on the patient's final outcome. A surgeon's volume may be indicative of his or her technical skill and case selection, whereas an institution's volume will encompass the perioperative processes and multidisciplinary personnel that will all impact on a patient's quality of care, some of which the surgeon will not be able to directly influence. This issue is at the very heart of volume-outcome research as it directly questions which is used as the independent variable, the surgeon or the institution.

1.2.5.1 The role of the Surgeon

A number of studies have shown a statistically significant relationship between surgeon volume and mortality across a variety of operations [96]. The lack of adjustment for hospital volume in many cases, however, has resulted in much questioning of the true role of the surgeon's caseload in surgical outcomes. Yet studies have shown that even when you adjust for the influence of hospital

volume on patients' outcomes, the surgeon volume-outcome relationship persists. Indeed in some instances, the surgeon volume accounted for the entire apparent volume effect [99], and adjusting for surgeon volume can remove the hospital volume-outcome relationship [100]. This clearly demonstrates the potential importance of the surgeon's volume. However, even among high-volume surgeons, inter-surgeon differences in performance do result in significant variability in patient outcome [94]. Independent of prior case volume, surgeons' achieving better patient outcomes could attract greater numbers of referrals and so be rewarded with a higher volume of cases (the selective-referral hypothesis). Similarly, high volume hospitals might attract surgeons who are already achieving better outcomes, which will in part explain the hospital volume-outcome relationship.

As the methodological quality of volume-outcome studies improves we may see that the surgeon and their caseload have a greater influence on outcomes for more technically complex operations. This does not mean to say that the hospital volume will cease to play a role, as the quality of intensive care services, ward nursing and physiotherapy remain integral to achieving the desired patient outcomes.

1.2.5.2 The role of the Institution

The institution in which a surgeon works will impact on the patient outcomes achieved. High volume hospitals are likely to be larger facilities and will provide a broader range of specialist and sub-specialist services. Associated academic and training programmes may result in higher staff to patient ratios, availability of innovative treatments & technologies and on-site multidisciplinary referral networks. The important question is 'does this translate to improved outcomes for patients'?

The importance of the structure and processes within an institution, aside from the surgeon's technical expertise, is suggested from the studies assessing the outcomes of cancer patients managed non-surgically. Although confounded by differences in treatment regimens, lymphoma and testicular cancer patient

mortality appears to be lower for patients receiving their treatment at a comprehensive cancer centre [101]. For colorectal resections, Harmon et al. [102] demonstrated that in terms of inpatient mortality, medium volume surgeons could achieve results similar to high volume surgeons when operating in medium or high volume hospitals but not low volume hospitals. Similarly the outcomes achieved by low volume surgeons showed some improvement with increasing hospital volume, although they never equalled those of high volume surgeons. The relationship between hospital volume and mortality following colorectal surgery appears not only to affect inpatient mortality, but also 30-day, 2-year and overall mortality [103]. This study corroborated the findings of Hillner et al. in showing that hospital volume was the greater predictor of each measure of mortality as compared to surgeon-specific mortality. Even when the authors used stoma formation rates as the outcome, which might be presumed to be almost solely determined by surgeon technique and preference, after adjusting for case-mix, hospital volume was as an important independent predictor as surgeon-specific volume. The exact reasons for this association have not been proven, but availability and quality of postoperative care available to a patient may well influence the surgeon in their choice of operation performed [103].

It is now generally agreed that ongoing volume-outcome relationship research should focus at the institutional level and try and identify factors such as processes of care for which volume is acting as a proxy measure. The impact of surgeon volume should not however be disregarded as the surgeon and the institution in which he operates remain integrally linked.

1.2.5.3 Immeasurable factors

The direction of causality in the volume-outcome debate has not been proven, however the 'practice makes perfect' hypothesis alludes to factors which have been thought immeasurable, such as the surgeon's skill. These factors have been implicated both at a surgeon level but also at an institutional level. While the majority of volume-outcome studies have used large administrative databases to investigate the relationship, in order that they identify small differences in

outcomes, access to clinical data extracted from medical records is likely to reveal differences in processes of care that were once considered immeasurable factors. This approach was used successfully by Thiemann et al. to investigate the association between hospital volume and survival after acute myocardial infarction in elderly patients [104]. A third of the survival advantage associated with treatment at higher volume hospitals could be attributed to the use of aspirin, thrombolytic agents, beta-blockers, angiotensin-converting–enzyme inhibitors, and revascularization. While availability to technology for angioplasty and bypass surgery was greater in high volume centres; this was not independently associated with overall mortality. No predominant mechanism was identified to explain outcome differences between high and low volume institutions, but as alluded to by the authors, this is not surprising as the mechanisms and processes involved within a patient's care pathway are multifactorial and interlinked.

Variations in processes of care may go someway to explaining why even for a cohort of high volume institutions, there are good and bad performers and why volume per se does not always adequately reflect quality of care in terms of its process. If such relatively subtle differences in processes of care can be responsible for altering outcomes, then could low volume hospitals, which enforce clinical best practice throughout the patient treatment pathway, achieve outcomes approaching those of high volume hospitals? There is evidence to suggest that when low volume institutions have as many residents/interns and registered nurses per 100 beds as high volume institutions, overall mortality rates for paediatric heart surgery and heart transplants appear equivalent [44].

Further volume-outcome research should focus on incorporating clinical data with administrative databases. This will identify important differences in the structure and processes between good and badly performing institutions that may or may not relate to caseload volume. Before we attribute outcome differences to immeasurable factors, we must first ensure that all measurable factors have been identified and considered.

1.2.6 Existing literature in urology

The majority of volume-outcome research in urology, most of which has been published in the last 5-10 years, has focused on surgical urology and covers mainly cancer operations. Studies have included radical cystectomy [105-108], surgical repair of bladder exstrophy [109], radical prostatectomy [110-113], nephrectomy [97 114-116] and TURP [117]. An important message comes from non-surgical urology, where comprehensive cancer centres appear to achieve lower mortality rates for testicular cancer patients. Although confounded by differences in treatment regimens, the importance of non-surgeon factors on patient outcomes is reinforced [101]. The majority of studies to date originate from the US and use administrative databases as their source data.

Most studies have assessed the more complex procedures of cystectomy and prostatectomy, and the most commonly measured outcome is either 30-day mortality or in-hospital mortality. Studies suggest a higher volume-lower mortality relationship for both cystectomy and prostatectomy, when comparing low with high volume providers (Table 1.6), although the relationship is not always statistically significant. Categorisation of volume means that the volume bands used to define low, medium and high providers etc. differ across studies, making direct comparison difficult. Other outcomes commonly measured include length of stay [97] and postoperative complications [118].

Studies have assessed longer-term morbidity such as late urinary complications after radical prostatectomy (e.g. bladder neck obstruction, urethral stricture) and results indicate that the rate is reduced if performed by high-volume surgeons or in high-volume institutions [93]. It is unclear if the same association is true for long-term continence. No association between number of prostatectomies performed and overall (non-cancer specific) mortality rates was found.

Study	Year	Operation	Country of origin	Unit of analysis	Volume handling	Outcome measures	Volume-outcome relationship	Methodological adjustment
Hollenbeck et al. [120]	2007	C	US	I	Terciles	1,3	1 ↓, 3 ↓	a,c
Barbieri et al. [179]	2007	C	US	I	Quintiles	1,2,3	1 ↓*, 2 ↔, 3 ↓*	-
Gilbert et al. [186]	2008	C	US	I	Quintiles	1	1 ↓	a,c
McCabe et al. [106]	2005	C	England	I	Continuous	1,3	1 ↓*, 3 ↔	-
Elting et al. [108]	2005	C	US	I	Terciles	1,2	1 ↓, 2 ↓*	a,c,d
Klein et al. [181]	2008	P	US	S	Continuous	6	6 ↓*	b,d
Gilbert et al. [186]	2008	P	US	I	Quintiles	1	1 ↓	a,c
Judge et al. [121]	2007	P	England	I	Quintiles	1,2,3	1 ↔; 2 ↓*; 3 ↓*	a
Hollenbeck et al. [164]	2007	P	US	I	Deciles	1,3	1 ↓*, 3 ↓*	a,d
Chun et al. [119]	2006	P	Germany	S	Continuous	6	6 ↓*†	b
Gilbert et al. [186]	2008	N	US	I	Quintiles	1	1 ↑	a,c
Davenport et al. [114]	2006	N	UK	I	Dichotomous	1,2,7,8	1 ↔, 2 ↔, 7 ↓, 8 ↓	-
Taub et al. [115]	2004	N	US	I	Terciles	1,3	1 ↓, 3 ↔	a
Konety et al. [116]	2006	N	US	I	Terciles	1	1 ↔	a,c,d
Goodney et al. [97]	2003	N	US	I	Continuous	3,4	3 ↔, 4 ↔	a

↓ = Inverse correlation between volume and outcome (as volume increases, outcome improves)
↑ = Positive correlation between volume and outcome (as volume increases, outcome worsens)
↔ = No correlation
* Signifies statistically significant result

† = Statistically significant independent predictor effect lost when the surgeon with the highest SV, who contributed about half of total number of cases analysed, was removed from the analyses.

Operation
C = Cystectomy
P = Prostatectomy
N = Nephrectomy

Unit of analysis
I = Institution
S = Surgeon

Outcome measures
1 = Mortality
2 = Complications
3 = Length of stay
4 = Readmission rates
5 = Morbidity
6 = Oncological outcome
7 = Conversion rate
8 = Transfusion rate

Methodological adjustments
a = Patient characteristics
b = Disease characteristics
c = Structural &/or process of care variables
d = Clustering
e = Multi-level modelling

Table 1.6 Summary table of recent literature assessing the volume-outcome relationship for cystectomy, prostatectomy and nephrectomy.

Immediate oncological efficacy as indicated by positive surgical margin (PSM) rates has been assessed following radical prostatectomy. Two studies have assessed the relationship between the rate of PSM following radical prostatectomy and surgeon volume [111] [119]. Both studies were single centre and highlight the benefit of prospectively collected data in this regard as results were also adjusted for pathological variables. There is a suggestion that surgical volume does decrease PSM rates, although in both studies a single surgeon with very high volume (accounting for a third to half of the total dataset) biased the results [119]. Similarly, patients treated at lower volume institutions are at increased risk of initiation of subsequent adjuvant therapy with radiation therapy, medical hormone ablation or orchiectomy following radical prostatectomy, indicating higher rates of failure of cancer control [110].

Although there has been preference towards measuring institutional volume and outcomes, only one study has specifically looked at processes of care in the patient's treatment pathway and their influence on institutional outcome variability [107]. This study identified that there were many processes of care before, during, and after radical cystectomy that differ between high and low volume institutions. Adjusting for these process differences accounted for 23% of the volume-mortality effect. Similarly, adjusting for institutional structural variables has been shown to attenuate the institutional volume-mortality relationship by up to 59% following cystectomy [120]. A patient treated in a low volume institution with a high registered nurse-to-patient ratio may have the same risk of mortality following cystectomy as a patient treated in a higher volume institution with a lower nurse-to-patient ratio [108].

Across all outcomes, most studies assess the relationship of the institutional volume with outcome as opposed to the surgeon-outcome relationship. The exact reason for this is unclear, but may reflect the feeling that a change in institutional processes could bring about improvements in outcomes more readily than changes at the surgeon level. Only a minority of studies combined both surgeon and institutional volume, and only one incorporated a multilevel modelling approach. This study confirmed the importance of using this statistical technique by

demonstrating that, for mortality following cystectomy, the proportion of the effect of surgeon volume attributable to hospital volume and vice versa was 46% and 39% respectively [99].

Only three published studies to date have looked at the UK's volume-outcome relationship: one for cystectomy [106], one for prostatectomy [121] and one for nephrectomy [114]. Using HES data, McCabe et al reported a significant inverse relationship between inpatient mortality rate and the mean number of cystectomies/year calculated over a five-year period at the institutional level. The study design is unfortunately limited by many methodological shortcomings and further conclusions cannot be made [106]. Also using HES data, Judge et al. demonstrated that length of hospital stay and rates of some short- and long-term postoperative complications following radical prostatectomy were lower in high-volume hospitals, although magnitudes of the effects were small [121]. Davenport et al. reported on the data obtained from the BAUS UK national laparoscopic nephrectomy audit over a three year period. Although data submitted to the database increased over the audited period, the study was not sufficiently powered to be able to primarily assess the volume-outcome relationship. However, early results suggest lower mean conversion rates and mean transfusion rates for institutions performing more than one case per month. These higher-volume centres showed a lower rate of minor morbidity, but conversely a higher rate of major morbidity; this may reflect case-selection and case-mix factors [114].

1.2.7 Public health implications

1.2.7.1 Policy change and Healthcare restructuring

Volume outcome research has enormous potential to impact on current and future policy initiatives. It is for this very reason that its validity must be assured. The Leapfrog Group is a coalition of America's largest corporations and public agencies that buys healthcare on behalf of their employees, dependants, and retirees and was founded in November 2000. One aspect of their mission statement is to advance the safety, quality and affordability of healthcare. With this premise, they acted as early proponents for selective referral to high-volume

institutions for five surgical procedures. Although better understanding of the volume-outcome research has necessarily caused a redefinition of the 'basket' of operations suitable for regionalisation, the message has not wavered and indeed has gained momentum with a number of other organisations joining in. Surely whatever the underlying causative factors, centralisation of healthcare can only be beneficial for the patient? Opponents, however, describe a number of valid disadvantages to centralisation; long travel times for patients and family of patients particularly in rural areas, alterations of local referral patterns that could destabilise other non-centralised healthcare services, the potential for a two-tier healthcare system generating yet further health inequality, inhibiting continuity of care because of segmentation of preoperative, perioperative and postoperative care and the possible overburdening of high-volume centres which could negatively affect quality of care.

Policy change is the 'end result' of health service research and acts as the initiator for healthcare restructuring. Restructuring and subsequent organisation of healthcare based on the centralisation model would implement a selective referral strategy. While the evidence might indicate probable or potential improvements for patients in terms of treatment outcomes, it is unable to incorporate the variability that exists in terms of disease incidence, patient demographics, existing healthcare resource and patient choice. In this regard, flexibility and adaptation needs to be exercised during any implementation process to reflect local requirements.

Change can be accompanied by a period of instability and uncertainty. Potential barriers to implementing a selective referral process are summarised in Table 1.7.

Restructuring in order to facilitate the selective referral process is not the final step in ensuring improved outcomes for patients. As discussed, high volume per se is no guarantee of improved outcomes. The continued evaluation of processes and systems of care within existing or newly created high volume institutions will act as a quality control mechanism, but also continue to enhance our

understanding of the determinants of improved outcome thereby facilitating any future modifications.

1.2.8 Research & Ethical implications

The last five years has seen a great increase in health services research exploring the volume-outcome relationship within surgery, yet there remain many unanswered questions. At the turn of the century, the American Institute of Medicine outlined the feedback of participants of a workshop 'Interpreting the Volume–Outcome Relationship in the Context of Health Care Quality' and their proposals for future warranted research actions [90]. Proposals broadly fell into four categories; Implementation of related research, new areas of research, methodological development and health services research data infrastructure. The improved quality of research since publication of this consensus has reflected in part some of the recommendations. Improvements have generally been seen in the field of methodological development, where authors have examined outcomes other than mortality, such as functional status, quality of life and longer term outcomes and better adjusted for confounding variables such as processes and systems of care. Outcomes and risk adjustment tools have also been tailored more towards being procedure specific as opposed to generic within surgery.

As with all translational research, the implementation of health services research into directing future developments in service provision remains the end objective. With the potential for such important policy implications as described previously, we must be certain not only of our methodological rigour, but also that we have asked the correct question. Many groups have pursued defining a minimum annual caseload above which better outcomes can be assured. This relies solely on the basis that volume acts as a surrogate marker for factors resulting in better outcomes for patients. As our understanding of the volume-outcome relationship has developed, we know this is not always true. Should we therefore not focus more on exploring factors that can be shown to correlate with improved outcomes and then verify if they are independently reliant on volume of delivery and scales of economy? This approach would better identify local variations in service provision and better influence policy change.

A potential for decrements in quality of care at higher volume providers
Patients' preferences for care close to home
Patients' lack of resources to travel to hospitals that are far away
Patients who need immediate treatment or are too unstable to transfer
Loss of access in areas where low-volume services have been closed (e.g., cardiac surgery)
Resistance from surgeons and hospitals to cooperate in quality monitoring efforts;
Effects on marketplace structure and competition:
Increased market power of high-volume hospitals (e.g., prices could rise)
Barriers to entry of new competitors (e.g., it is difficult to start at high volume)
Potential for medically inappropriate admissions to boost volumes to meet cut-offs.

Table 1.7 Potential barriers to implementing a selective referral program based on a volume standard. Adapted from reference [90]

Not surprisingly, volume-outcome research is not spared of ethical considerations. Varying service provision demographics means that widespread and all-encompassing service re-configuration based on volume-outcome research will have to overcome a number of challenges. It has been suggested that this may be particularly true in rural areas and as a result it has been suggested that "initial implementation efforts be focused in urban areas with dense concentrations of hospitals, where success is more likely" [90]. During this implementation phase there is clearly the potential for a 'two-tier' healthcare system separating rural and urban populations.

Policy change based on this research can only occur once we are certain of its validity and that validity must be confirmed by a number of stakeholders and in particular those who will not have commercial gain. This is not to say that we should discourage healthcare providers from carrying out volume-outcome research, but we need to be happy that the output is robust and demonstrates equity to all patient groups.

1.2.9 Conclusion

Research of the volume-outcome relationship in surgery will continue to thrive in the current healthcare environment and will help shape its future structure. As we have dissected apart the relationship, we have unveiled a complex interaction of numerous outcome, process and structural factors. Although the 'practice makes perfect' or 'selective-referral' hypotheses are not the sole determinants of the causal relationship, they will continue to form a part. Continuing research in this area should be conducted under the guidance of a methodological framework that incorporates many of the methodological limitations and ideals discussed within this chapter and equally emphasise, and account for, the importance of the hierarchical nature of patients' outcomes within modern healthcare.

Great strides have been made and our understanding today far surpasses that of 20 years ago. This is however not the end and continued research will form an important component of society's wish and the medical communities' desire to improve the quality of care that patients receive.

1.2.10 Key points for volume-outcome relationship in surgery

- We have seen an appreciable increase in research evaluating the volume-outcome relationship in recent years

- Existing explanations for the relationship have revolved around the 'practice makes perfect' and/or 'selective-referral' hypotheses.

- Most studies are restricted to using administrative databases as their source and are thereby limited by insufficient case-mix adjustment

- A specifically designed rating system has demonstrated that the methodological quality of existing volume-outcome research is modest at best

- Although the majority of studies have supported an inverse correlation between volume and outcome, appropriate refinement of the statistical methodology attenuates this relationship

- Mortality has been most frequently used as an outcome measure; expansion to more procedure-specific and discriminatory outcomes would be appropriate and should include process, structural and quality of life measures

- Development of multilevel modelling based on a methodological framework is needed to help direct future research and the clinical translation of evidence in this area

- This health service research is crucial for ongoing advances in healthcare delivery and the subsequent quality of care that patients receive.

1.3 Funnel plots and their emerging application in surgery

The concept of exploring, modelling and displaying the variation in performance is not new. Well over 80 years ago, the attempt to identify unacceptable variation in performance was first developed by Walter Shewhart, a physicist and engineer. He described two coexistent causes of variation, using a managerial setting: "common cause variation", which is intrinsic to any stable process, and "special cause variation", which signals the effects of external factors upon a process. The first step in reducing the effect of the factors causing the special cause variation is to acknowledge their existence.

As highlighted recently, surgical outcomes are being more commonly measured as a facet of quality improvement initiatives [122]. Although this occurs more intensively in private healthcare delivery, similar changes are evident in publicly funded systems such as the NHS. Increasing accountability within healthcare as part of governmental policy has meant that surgeons and surgical outcomes have lent themselves to inspection as a result of the interventional nature of their discipline. This has led to the reporting of data reflecting performance and quality of care for specific surgical specialities [123]. Maximising productivity of healthcare output involves continuous quality assessment which is becoming increasingly important as health demands increase, although resources remain often finite. Healthcare institutions need to become more involved in the process of quality assessment, including the use of appropriate and practical quality control tools to measure and benchmark performance. In the setting of a value-based competitive healthcare delivery system, performance may have an impact on future development and fund allocation.

Several quality control tools have been introduced and tested in healthcare but their use has been limited by the fact that their importance has been recognised within the managerial and statistical circles, but relatively ignored by clinicians. Graphical tools such as funnel plots that have been previously implemented within the industry setting were considered more practical, although their uptake within the healthcare setting has been relatively slow to follow. Subsequent to

their validation as a tool for assessing performance outcomes [124], we are now seeing increasing application in several aspects of healthcare assessment. As with other graphical tools such as control charts or CUSUM charts, funnel plots offer the opportunity to statistically define "control limits" around measurable outcomes. This allows for adjustment of unmeasured factors and makes them extremely useful for the monitoring of performance in surgery, either at a surgeon, institutional, regional or national level or indeed by means of qualifying aggregated data in reported literature.

1.3.1 Principles of funnel plots

1.3.1.1 Funnel plots versus ranking

The ability to perceive variation in data is greater with graphical than tabular display. The method of presentation may however influence our interpretation of the results. Marshall et al. used a randomised controlled trial design to assess the effects of using two forms of data presentation on outlier identification by health service managers [125]. Three different case studies of service provider mortality outcome data were displayed using either a ranked histogram (or league table – ranking of institutions' performance from 'worst' to 'best', e.g. from highest to lowest mortality rate) or a control chart. They first simulated death rates consistent with common-cause variation so that labels like 'good' and 'bad' performance in the league table were spurious (and hence any investigation unnecessary, which the control chart avoided); in the second case study there were two units with genuinely high rates and in the third there were no such units. Although this study was limited by its low response rate (47%) and blinding of the authors to the characteristics of the participants, the league table group wanted to investigate a greater number of units than were actually out of control, whereas the control chart group wanted to investigate on average about the right number. The presentation of outcome data using control charts may therefore reduce the inappropriate labelling of service providers as outliers, which can occur with 'league table' formats.

1.3.1.2 The control limits of a funnel plot

The control chart as developed by Shewhart in the 1920s has been adapted to generate the funnel plot; both adhere to the basic theory of statistical process control [126]. Constructing funnel plots involves plotting data on a scatter plot and then superimposing 'control limits' around the data points (Figure 1.9). The control limits typically represent two and three standard deviations from the mean. The method of calculation of the standard deviation is affected by the data distribution which is typically binomial or Poisson (Table 1.8). The smaller the sample size (denominator) the wider the control limits become, reflecting the greater the variability due to chance and the degree of uncertainty. Therefore, as the sample size (denominator) increases, the control limits converge and a 'funnel shape' forms. Data points lying within the control limits are said to be consistent with common cause variation, or natural variation inherent in any stable process. This can be a result of chance (random) variability or equally reflect marginal differences in external factors. For example, a degree of variation between hospitals for lengths of stay following complex elective surgery can result from factors such as postoperative care pathways, rehabilitation facilities, discharge facilitators, local socioeconomic deprivation, and availability of carers. It is accepted that common cause variation will compensate to a degree for differences in the data points as a result of unmeasured factors, but it should not be used to overcome basic inadequacies in data quality.

Data points outside of the control limits are said to be consistent with special cause variation. The causes of this variation can be internal or external to a process, intentional or unintentional and transient or permanent. Mortality differences following emergency repair of ruptured aortic aneurysm between Trusts could result from decisive factors such as surgical error, availability of open or endoluminal repair techniques and availability of intensive care support.

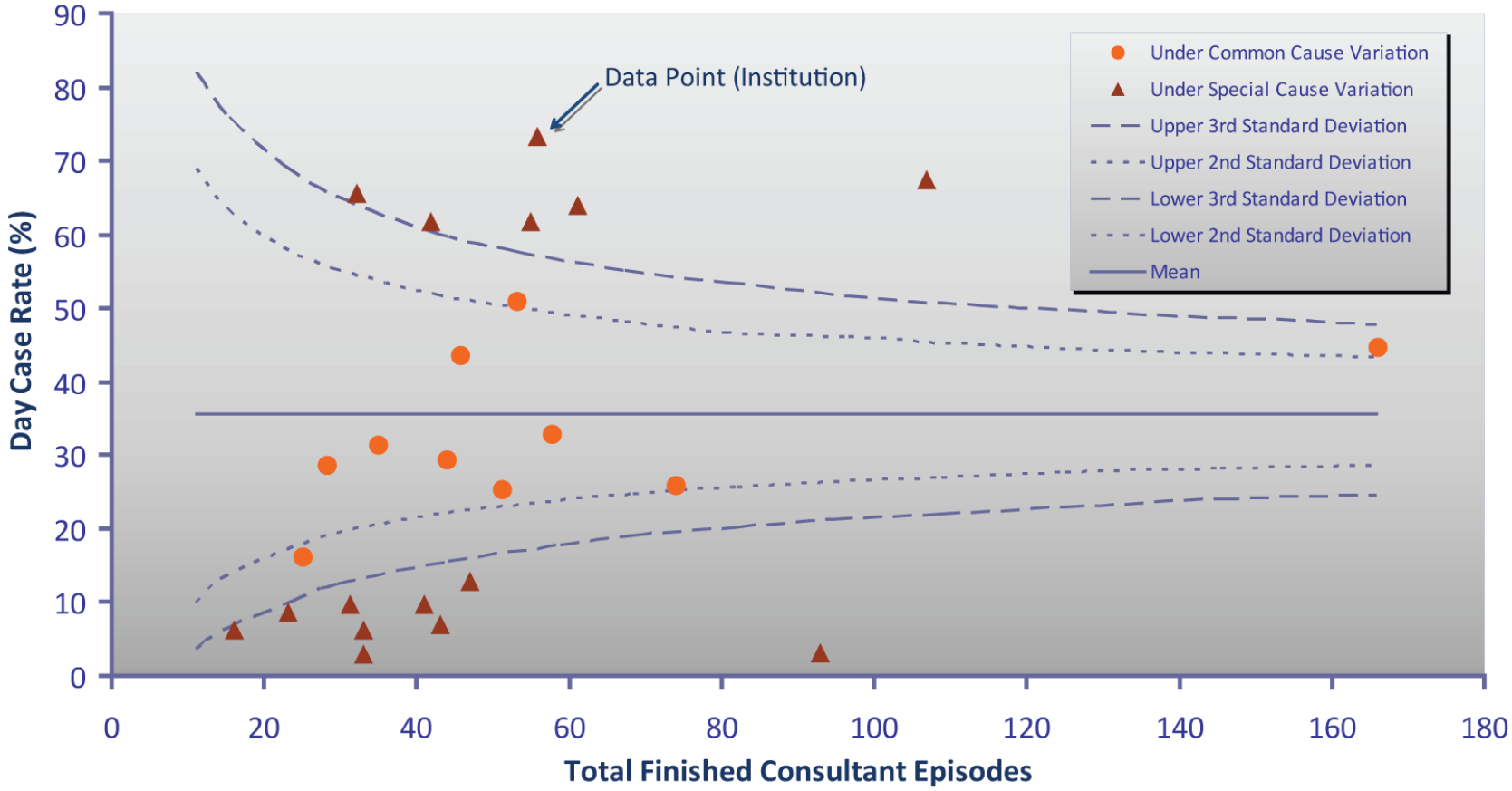


Figure 1.9 Funnel plot of day case rates for haemorrhoidectomy in London acute NHS Trusts (2005/06). Source data from Hospital Episode Statistics online[127]

Distribution for outcome measure	Clinical examples	Example web site for formulae/Excel macros to calculate control limits †
Normal	**Duration of operation, Intensive Care Unit length of stay (transformed as required, e.g. using logarithm)**	**http://www.qualitydigest.com/pdfs/empiric.xls**
Poisson	**Infection rate, standardised mortality ratios (for low-mortality rates e.g. <5%), number of post-op falls**	**http://statpages.org/confint.xls**
Binomial	**Crude death rate, day case rate**	**http://statpages.org/confint.xls**

Table 1.8 Tabulation of the three types of outcome measure distributions used to calculate funnel plot control limits with clinical examples. Online links are provided to download formulae

†Exact and approximate formulae for calculating confidence intervals as appropriate for the distribution of the outcome measured [proportions, rates (whether indirectly or directly standardised), and continuous outcomes (both for cross-sectional data and changes)] are described in detail in Appendix A of Spiegelhalter DJ. Funnel plots for comparing institutional performance. Stat Med 2005; 24:1185-202 (reference [89]).

1.3.1.3 The choice of axis in funnel plots

Interpretability of a graphical tool is very much influenced by the choice of axis. When a funnel plot is used to detect publication bias, it is typical to plot a measure of the study size on the vertical axis and a measure of the treatment effect on the horizontal axis. As a result the "funnel" is in an 'up-down' direction (Figure 1.10). The application of funnel plots in assessing outcomes in healthcare has generally seen a switching of the axis with the 'health outcome' plotted vertically against a measure of sample size horizontally as in Figure 1.9. The funnel is therefore in a 'left-right' direction. There is no apparent statistical reason for this change, although as a result data points will fall above or below an overall 'mean line', potentially visually representing under or over achievement. This effect is contradictory to the ethos of the funnel plot and the result of this variation on interpretation of the data has not been formally assessed.

1.3.2 Funnel plots and institutional or surgeon's performance

1.3.2.1 Performance indicators

Traditionally, there has been hesitation at publicly reporting comparative surgical outcome data, probably reflecting concerns over the impact of its misinterpretation. In contrast to league tables already discussed, funnel plots draw attention only to those units that lie outside the control limits rather than the (often spurious) ranking of each of them. On the basis of their performance data, units of interest can be split into three groups: those whose performance is as expected, those who are performing unexpectedly well and those whose performance is unexpectedly poor.

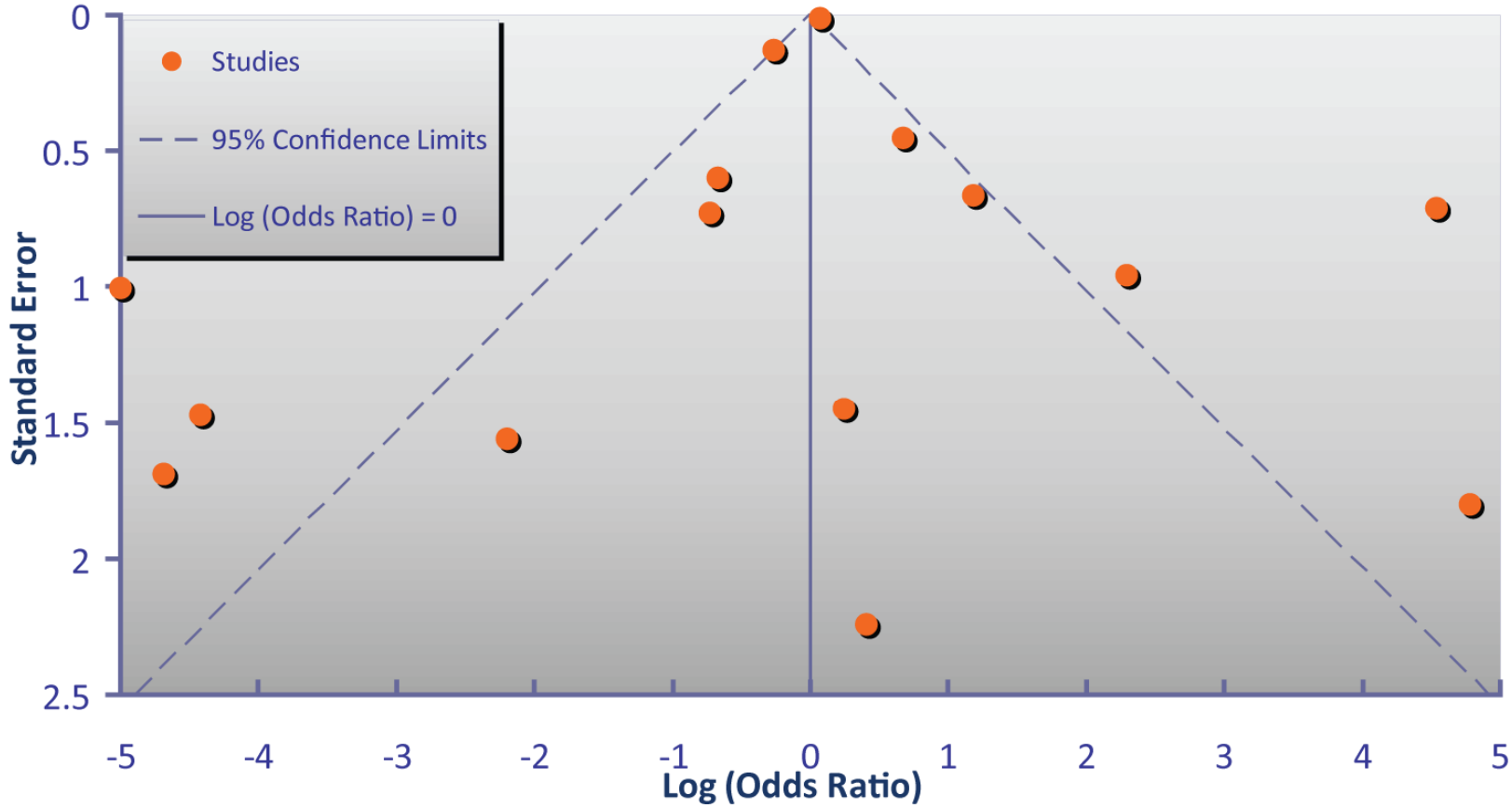


Figure 1.10 Simulated data used to demonstrate a symmetrical funnel plot

The most recent and most publicly transparent surgical outcome data in the UK started as a result of the collaboration between the Healthcare Commission and the Society for Cardiothoracic Surgery in Great Britain and Ireland [123]. For the first time, cardiac surgery survival data for both institutions and individual surgeons is available via the World Wide Web. Although there is no direct adjustment of the data, an expected survival rate range (confidence intervals of the expected survival rate) is indicated alongside each of the observed survival rates. Using a caterpillar plot, assessment of performance of cardiac surgical units nationally shows that 12 institutions with a standardised mortality ratio (SMR) greater than 100 could be labelled as 'poor performers' if the confidence intervals were not displayed or their meaning not understood (Figure 1.11). Representation of the data as a funnel plot plausibly changes the interpretation (Figure 1.12). The majority of Trust rates were consistent with common cause variation. Trust X was ranked as the worst performer on the caterpillar plot, when in fact its rate was consistent with common cause variation as indicated by the funnel plot and therefore had acceptable performance. Only Trust Y fell outside of the 99.8% control limits and appeared to be truly divergent. Although anonymous and not publicly advertised, the Vascular Society of Great Britain and Ireland has also used funnel plots to perform preliminary institutional comparison of mortality data following aortic aneurysm repair [128].

These advantages of funnel plots over ranking systems has possibly facilitated individual units in having the confidence to academically publish their performance results, although the effect of external influences, such as public demand for increased transparency, cannot be ignored. Both funnel plots and cumulative funnel plots have been used to present in-hospital outcome data for percutaneous coronary intervention of five consultant cardiologists. Funnel plots display summaries of overall performance for each operator and cumulative funnel plots display individual operator's performance on a case series basis [129].

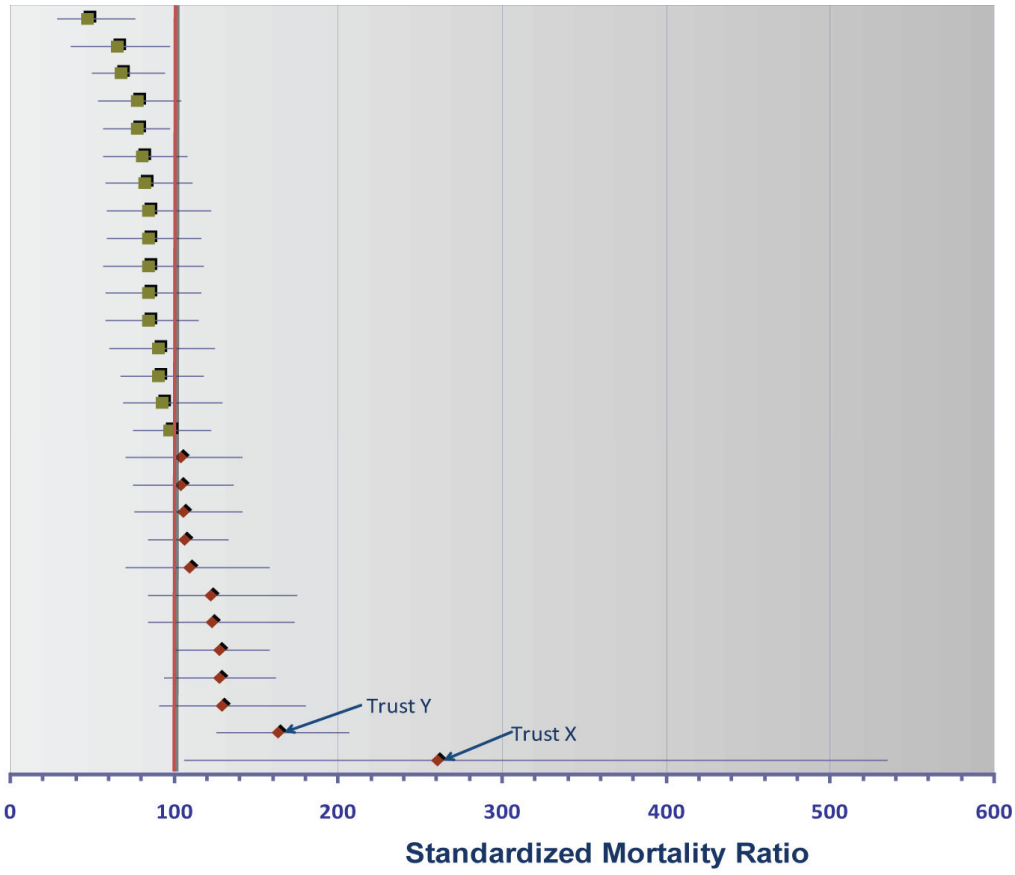


Figure 1.11 Caterpillar plot of in-hospital standardised mortality rates following coronary artery bypass grafting in 28 acute NHS Trusts in England, 2004-2005. 95% confidence intervals are plotted. Source data from the Healthcare Commission's 2004/2005 NHS Performance Ratings[130]

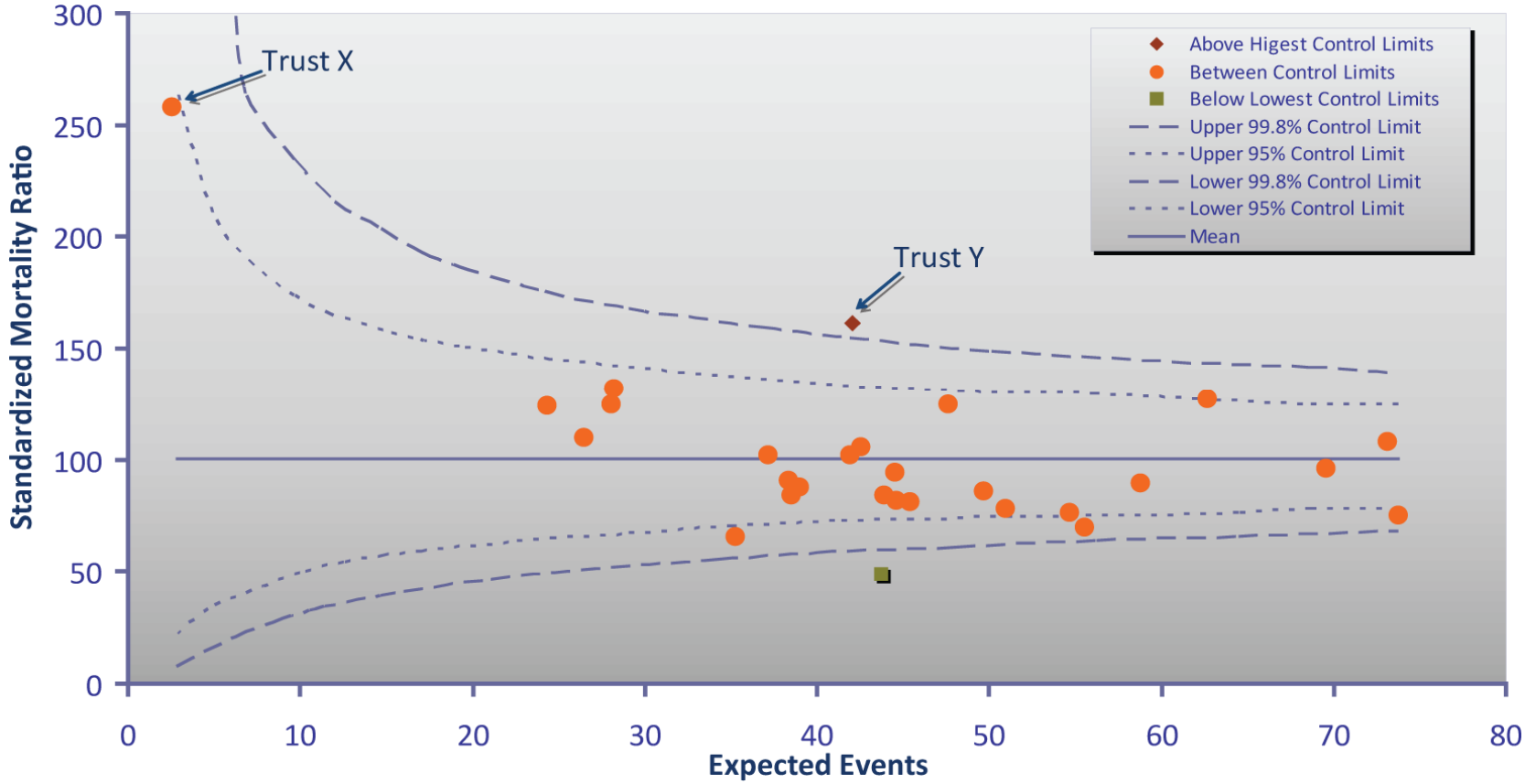


Figure 1.12 Funnel plot of standardised mortality ratio plotted against number of expected in-hospital deaths for same 28 acute NHS Trusts in England as represented in figure 3. Source data from the Healthcare Commission's 2004/2005 NHS Performance Ratings[130]

1.3.2.2 Need for Adjustment

Funnel plot methodology will 'tolerate' factors that result in 'random' variation between units of interest, but will not compensate for frankly heterogeneous data. Practically useful performance indicators should reflect healthcare processes or outputs that are common to all units of interest and ideally be adjusted for aspects such as the treated population's case-mix, or use a relatively homogenous patient population. This issue was highlighted by Mason et al. who used an administrative dataset to compare outcomes in urology [98]. They concluded that two measures, 30-day emergency re-admission rates after day-case care and after elective admissions for a therapeutic endoscopic procedure on the outlet of the male bladder or prostate, were of most use because of the reasonable homogeneity of the populations represented (any patients with a cancer diagnosis were removed from analysis). Standard funnel plots were constructed using 95% and 99% control limits. Of the measures indicated above, only eight and seven hospitals respectively were outside of the 99% control limits. By chance alone, this figure would be expected to be two hospitals (200 hospitals analysed). This suggests that there is very little variation in outcome across the NHS as assessed by the measures outlined above, but it does rely on reasonable matching of case-mix. The same group identified that in the absence of homogeneity there existed no measure of mortality that could be practically used to compare the performance of gynaecological units [131]. However, again the emergency readmission rates after day-case and elective abdominal hysterectomy could be used as indirect indices. Under these circumstances the funnel plot, although more meaningful than a ranking histogram, must be interpreted with caution.

An existing limitation in the UK as compared with the US of comparing surgical outcomes has been the relatively poor quality of speciality-specific data collection. Recent improvements in adjustment of nationally collated administrative databases for patient co-morbidity now potentially allows for great expansion in risk-adjusted analysis [87].

1.3.2.3 Sequential monitoring with funnel plots

As for histograms, funnel plots are used typically to provide a 'snapshot' analysis and are therefore subject to variation over time through chance alone. Sequential monitoring of performance can be achieved by the construction of cumulative sum (CUSUM) charts, which plot a cumulative number of events against time. The events, such as 30-day mortality following an operation, can be directly observed and then compared with a predicted value (observed-expected statistic), or adjusted for pre-operative risk factors (risk-adjusted CUSUM) [132] [133]. Further methodological modifications allow for specific charts to be applied depending on the context of performance assessment. All however recognise the importance of adjusting for individual patient risk [134]. Funnel plot methodology can also be adapted to identify performance changes over time; a rate ratio (change in performance indicators observed at two points in time) assesses the relative change in performance and is particularly useful when looking at the impact of a change in healthcare delivery or performance against 'time-fixed targets' (Figure 1.13).

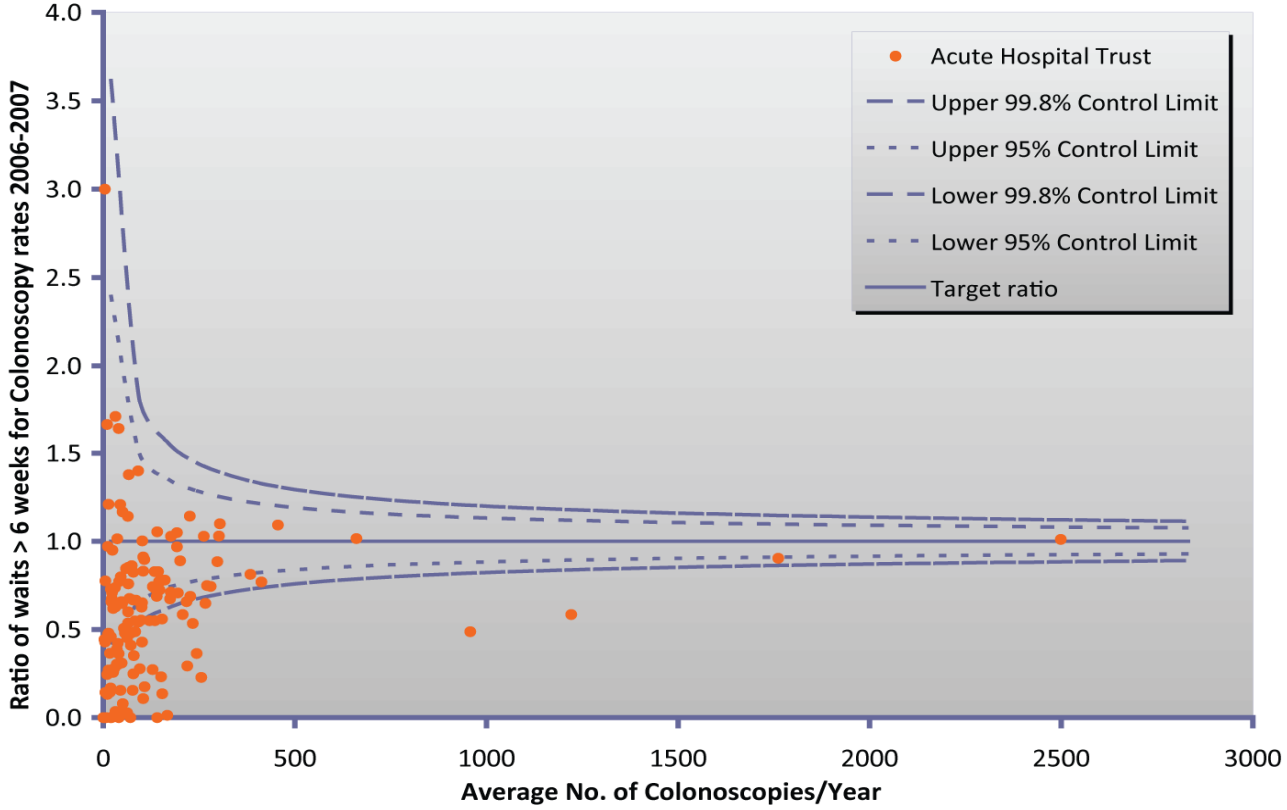


Figure 1.13 Ratio of % of people waiting greater than 6 weeks for an elective colonoscopy in February 2006 to those in February 2007 in 150 acute Trusts in England. Nine Trusts with missing data were excluded. The control limits are calculated assuming a rate ratio = 1 (no change in rates over the time period). No Trust appears to be failing in its efforts to reducing waiting times for colonoscopy as part of the governmental initiative ensuring that the NHS is progressing as planned towards achieving an 18 week patient pathway. Source data from Hospital Activity Statistics – Diagnostic Statistics[135]

1.3.3 Over-dispersion and how to deal with it

When assessing potential indicators of performance we may see the majority of data points sitting outside of the control limits (over-dispersion). This is in itself counter-intuitive and could either reflect insufficient adjustment of the data or an uncontrolled underlying process. If the goal of the analysis is simply to describe the variation between units, then no modification is required. However, if the goal is to detect units with divergent performance, one can adjust the data using robust statistics and bring the process back under 'control'. These methods for dealing with over-dispersion are well covered in the review by Spielgelhalter [136], but in summary include clustering, where data points (institutions) are stratified into homogenous groups to improve the comparison of like with like; comparing data points with a 'normal range' instead of a precise target; and inflation of the control limits, either by the incorporation of an over-dispersion factor or by using a random effects model. A view is held by some that the presence of considerable over-dispersion is indicative of the poor discriminatory value of a performance indicator and its use should be withheld.

1.3.4 Volume-outcome relationship

Funnel plots are ideally suited to exploring the volume-outcome relationship in surgery by plotting an observed event rate against a volume of cases. It has long been believed that institutions, or indeed surgeons, that undertake greater numbers of particular operations achieve improved surgical outcomes over 'low-volume' institutions [137]. This has precipitated service reconfiguration in the UK, with centralisation of specialist services and similarly in the US, selective referral to high-volume institutions driven by proponent groups such as Leapfrog [42]. The Leapfrog group has placed emphasis on a point cut-off for volume of procedures, such as oesophagectomy, coronary artery bypass grafting and elective abdominal aortic repair that should be undertaken by an institution with the aim of reducing the annual death rate for patients undergoing these procedures [138]. Birkmeyer et al. looked at the potential benefits of adherence to more recent Leapfrog group standards and suggest that the annual death could be reduced by as much as a third [139]. This study plotted the average adjusted mortality rates for institutions

categorised into high or low volume groups based on a cut-off volume threshold, for five different procedures. Although the mortality data was adjusted for age, gender, race, admission acuity and co-existing diseases, there were limited further adjustments to allow for institutional, and indeed surgeon factors, that will influence the patient outcome (mortality) and the greater variability of outcomes that will exist for the lower volume institutions due to sample size constraints. For each of the five procedures, the low-volume institutions are shown to have statistically significant higher mortality rates, although the lack of adjustment for important influential factors and statistical methods chosen will have affected the results. By plotting standardised mortality ratio against volume of cases, one can assess for a volume-outcome relationship whilst still allowing for common cause variation and simultaneously identify truly divergent institutions, be they high or low volume. This approach was used in the national enquiry into paediatric cardiac surgery mortality at UK hospitals [140]. It was clear that there existed an inverse relationship between mortality following paediatric cardiac surgery in under 1's and annual caseload. It was also evident that one of the 12 hospitals had a truly divergent performance sitting outside of the upper 99.8% control limit. All other variation between hospitals was consistent with common cause variation (Figure 1.14) [89].

While there is a large amount of research corroborating the positive correlation between volume and outcome, there are numerous studies to suggest no correlation. Of interest we see that when simple scatter graphs are used to plot adjusted mortality, e.g. observed versus expected mortality, as an indication of true mortality, against volume, correlations are not so apparent [141]. Moreover, recent papers have explored the methodological limitations of existing techniques of representing volume-outcome relationships. They highlight that the greater variability of outcomes from centres performing low volumes of cases must be considered, but there is no suggestion for the use of funnel plots as an alternative to displaying the data [88]. Funnel plots would be ideal in this setting with their easily understood axis. They provide an initial indication of potential volume-outcome relationships, which can then be formally tested using a regression model [89]. It is important to emphasize that observed outcome rates are more variable in

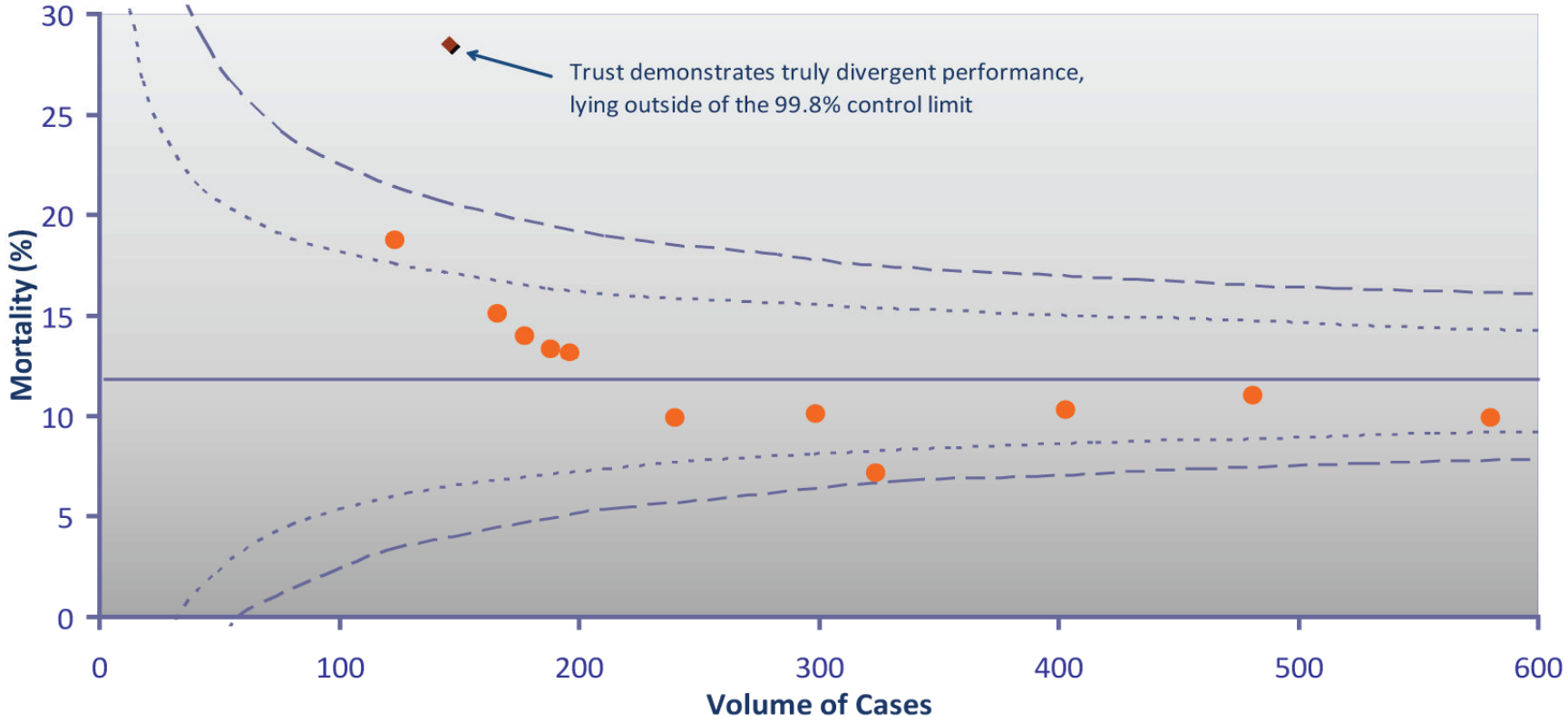


Figure 1.14 Mortality rates following paediatric surgery in under-1's in 12 English specialist centres, using HES 1991-1995. The target is the overall average rate of 12 per cent. Reproduced from reference [89]

low-volume centres because of their small sample size. For example, with five cases and a real (underlying) death rate of 20%, observed death rates of 0% to 100% will be much more common than in centres with 50 or 500 cases, and control charts allow for this size-dependent sampling variation.

The inherent adjustment for common cause variation allows for the variation in institutional processes of care that have yet not been fully identified and/or measured. The presence of truly divergent institutions as identified by the funnel plot, either as good or bad performers, will provide us with the best opportunity to understand the underlying factors driving quality of care. This can have important implications for policy change as identification of absolute cut-off values for volume threshold may not be optimal.

1.3.5 Assessment of publication bias in the literature and funnel plots

The assimilation of published data and subsequent interpretation forms an important component of practising evidence-based surgery. One of the most frequent uses of the funnel plot to date is in assessing the quality of aggregated data in meta-analysis [142]. By plotting a measure of study precision (vertical axis) against treatment effects, for each of the studies within a meta-analysis, the resultant symmetry or indeed asymmetry can give an indication of the presence of publication bias. Publication bias describes the phenomenon whereby systematic reviews and meta-analysis do not necessarily include all available studies of interest. This occurs because the likelihood of finding studies during systematic searches is related to the results of those studies [143], such that smaller studies without statistically significant results are less likely to be published [144]. In the absence of publication bias, a symmetrical inverted funnel (hence the name funnel plot) is generated because the treatment effect estimates from smaller studies scatter more widely at the base of the graph with those of larger studies narrowing towards the apex (Figure 1.10). If a meta-analysis is overrepresented by larger studies with more significant results, and therefore smaller studies are not included, asymmetry of the 'funnel' such as that seen in Figure 1.15 occurs. The problem of publication bias in the literature is not insignificant; previous research suggests that as many as half of meta-analyses may be subject to some level of

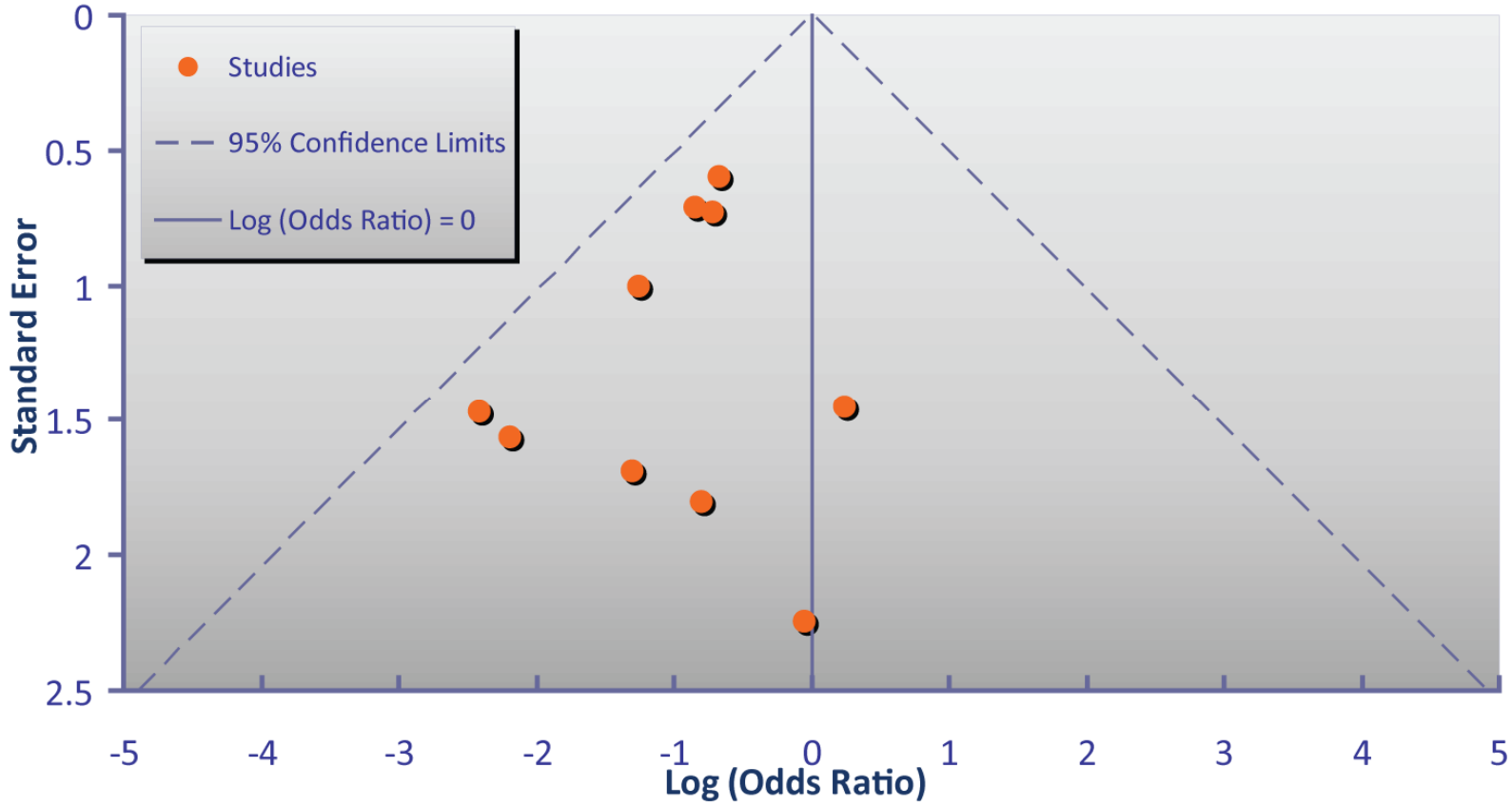


Figure 1.15 Simulated data used to demonstrate an asymmetrical funnel plot

publication bias [145]. Furthermore, the absence of these 'missing' studies as a result of publication bias could result in about 5-10% of meta-analyses being interpreted incorrectly [145]. Publication bias can be grouped with location bias under a heading of selection bias when categorising causes of funnel plot asymmetry. Location bias describes factors such as the checking of reference lists of other studies and reviews increasing the likelihood of locating 'supportive' studies (citation bias), only incorporating studies reported in English language (language bias), studies published in journals not indexed in one of the major databases may not be included (database bias), and multiple publications originating from single studies results [142]. Other categories, and therefore causes of asymmetry, include poor methodological quality of smaller studies, true heterogeneity because of study size differences, chance, and artefact as a result of the selected measure of study precision and/or treatment effect [146]. Asymmetry of funnel plots cannot therefore be automatically attributed to 'bias' and other explanations need to be considered.

Funnel plot construction requires a measure of study precision plotted against treatment effect. If inappropriate measures are selected, funnel plots can lose their ability to detect publication bias. Tang et al [147] studied a group of published meta-analyses with asymmetrical funnel plots suggestive of selection bias. In over 85% of the meta-analyses studied, a change in the definitions of study precision and /or treatment effect used to construct the funnel plot resulted in the funnel plot becoming symmetrical. This raises an issue about the certainty with which conclusions can be made concerning the influences of bias for asymmetrical funnel plots. In order to try and improve the standardisation of axis choice, Sterne et al. [148] presented guidelines that could be applied to funnel plots of meta-analyses with binary outcomes. They concluded that in most cases standard error is the most preferable measure of study precision, although precision, variance and the inverse of variance are also valid. A log ratio measure (log odds ratio or log risk ratio) is best used as a measure of treatment effect; a measure's consistency across trials and its ease of interpretation is an important consideration [148].

Methods used to detect publication bias, of which funnel plots are one example, should themselves be subject to an evidence base. This will help to reduce costs to patients and society through incorrect treatment decisions being made based on misleading methodological processes [149].

1.3.6 Public health issues related to surgery and application of funnel plots

Presentation of healthcare data is an important aspect of patient education both in terms of health behaviour modification and public accountability. These data must be easily interpretable, requiring minimum background understanding and account for as many confounding factors as possible, so as not to be grossly misinterpreted. It is therefore common to see 'high level' data presented, often with graphical summation. Such comparative healthcare data commonly appear in the public domain via the internet, newspapers or national news broadcasts and the lay person draws conclusions at face value. Funnel plots have been used to display indicators of public health across the English Regions as commissioned by the Chief Medical Officer in 2003 [150].

Ranking of data can allow for variation in outcomes by chance alone, through the addition of confidence intervals or Bayesian ranking. These methods are not always readily understood, both by healthcare professionals and the public, and there remains a natural tendency to focus on the rank of the organisation. This has the danger of generating a two-tier healthcare system as one group of patients travel to obtain perceived optimum healthcare, whilst others, unable or unwilling to travel will persist with local healthcare. We may see division of healthcare provision across social class, which contradicts the ethos of equality of healthcare. Certain institutions may not be able to manage the increased workload generated by a favourable rating. League tables can themselves result in inequalities through generation of unequal demand. This, in combination with initiatives such as pay for performance, could lead to financial disadvantages for institutions that in truth are not underperforming. A sudden fall in income may then result in failure of healthcare services for all.

One area of interest in the study of inequalities in healthcare has focused on geographical/institutional differences in access to healthcare services. Several government initiated service delivery targets such as operative waiting lists [151], MRSA infection rates [152], referral to treatment times for cancers [153] have been implemented to reduce existent inequalities. Not surprisingly, differences in local infrastructure and population and disease demographics have inhibited a complete abolition of inequality. The continued identification of inequalities is therefore important to enable targeting of areas where further and greatest impact can be made. Funnel plot methodology, as for performance assessment, has been used to identify where inequalities exist, although the potential for more widespread use exists.

MRSA infection of hospitalised patients causes significant morbidity and mortality and financial burden to hospitals. A government target announced in 2004 of halving the MRSA bacteraemia rates over a three year period to 2008 has made this a key issue for NHS Trusts and the National Patient Safety Agency [154] , and it remained a performance indicator assessed by the Healthcare Commission [155] and now by the Care Quality Commission [156]. It is inevitable that a degree of variation between Trusts in rates of reported MRSA bacteraemia will ultimately be caused by the treated patient population, variability in community MRSA rates of the Trust's catchment population, MRSA prevalence of inter-hospital transfers and chance alone. Whilst there has been some adjustment to the data publicly released, the figures have not included the above significantly weighted factors. As a result, hospitals have been spuriously ranked and the public misinformed. Some of the limitations can be overcome using the funnel plot methodology. For clinicians, epidemiologists and public health doctors, this non-specific data adjustment may severely limit the usefulness of the funnel plot; however, for the non-statistically minded public it will reduce the misdirected criticism of institutions that are in fact performing entirely within acceptable limits. Figure 1.16 is a funnel plot of MRSA cases per 1000 bed days against the number of bed days. Each acute, non-specialist Trust is represented as a data point. It is not uncommon for the scatter of results to be considerable when assessing performance.

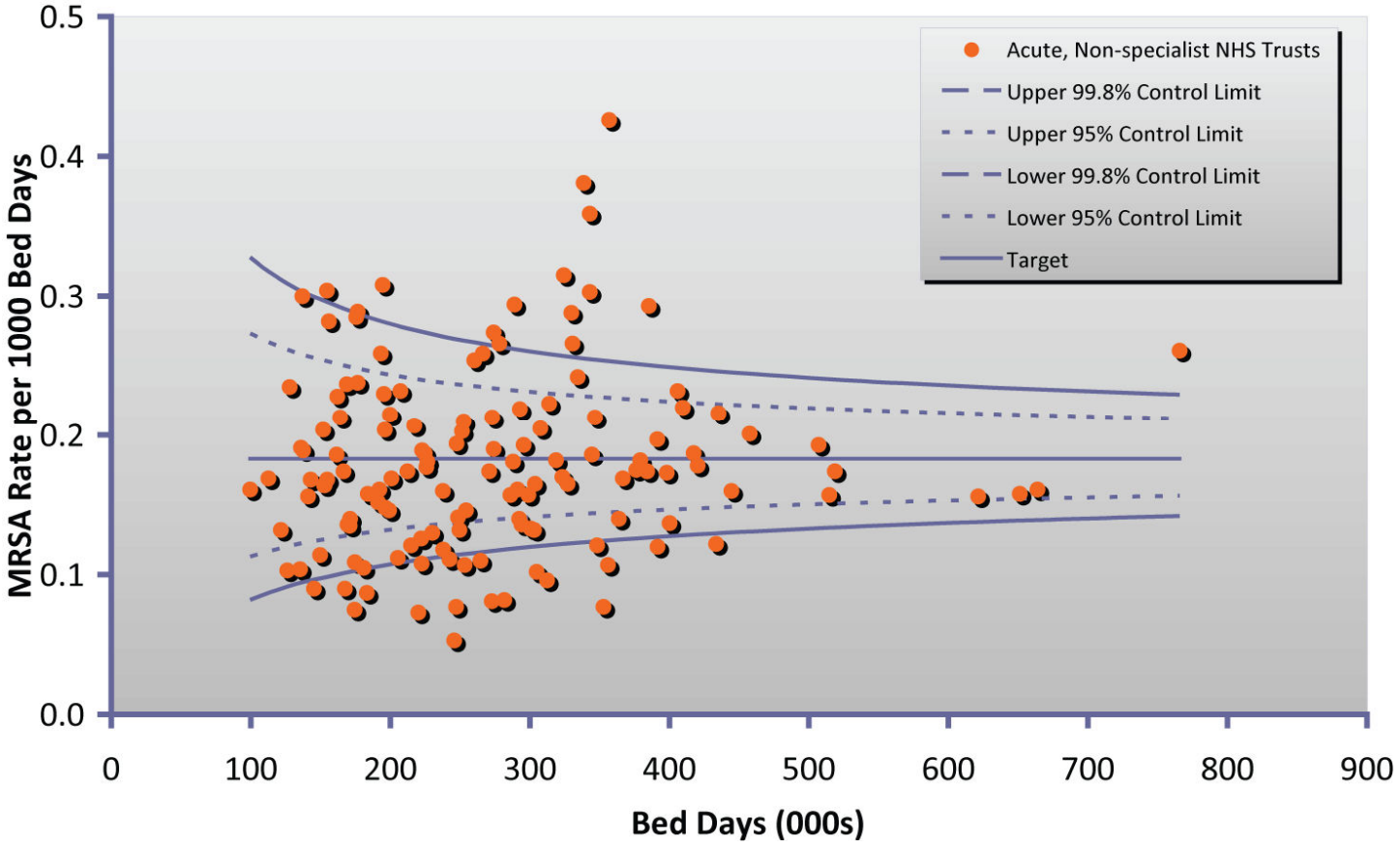


Figure 1.16 MRSA cases per 1000 bed days in 139 Specialist and General acute NHS Trusts in England in 2004-5 (single speciality Trusts not plotted). Source data from the Healthcare Commission's 2004/2005 NHS Performance Ratings[130]

The funnel plot, however, demonstrates that 13 Trusts are clearly outside of three standard deviations above the mean. Equally however, there are 17 Trusts that fall outside of three standard deviations below the national mean MRSA rate. The divergence of these Trusts potentially indicates that this special cause variation is beneficial to reducing MRSA infection rates. The processes within these better performing Trusts can be assessed to determine transferability nationally to reduce MRSA rates.

There exist a number of problems when measuring changes in rates of an outcome over a time period, such as chance variability, regression to the mean and a low power to detect true underlying changes [157]. These are worsened when handling low event numbers, such as MRSA bacteraemia cases at individual hospitals. Furthermore, infectious disease cases demonstrate a 'clustering' phenomenon, resulting in over-dispersion [157]. All of these factors will complicate distinguishing between changes in underlying risk and an observed (estimated) change in rate. Funnel plots have been used to demonstrate these described difficulties with monitoring rate changes and also to demonstrate the uselessness of ranking under these circumstances [157]. However, monitoring rate changes for outcomes like MRSA bacteraemia is likely to continue and recommendations for future improvements, such as analysing changes in MRSA rates over at least three-year periods, will help to improve the robustness of the interpretation [157].

The Association of Public Health Observatories facilitates joint working across a network of 12 Public Health Observatories to produce information, data and intelligence on peoples' health and healthcare for practitioners, policy makers and the wider community [158]. To truly understand the health needs of the population, one needs not only to measure incidence and prevalence of disease, but also identify regional differences in these and the outcomes of patients treated with those diseases. Inequalities should trigger review into the underlying causative factors and lead to corrective measures aiming to improve the overall health of the studied population.

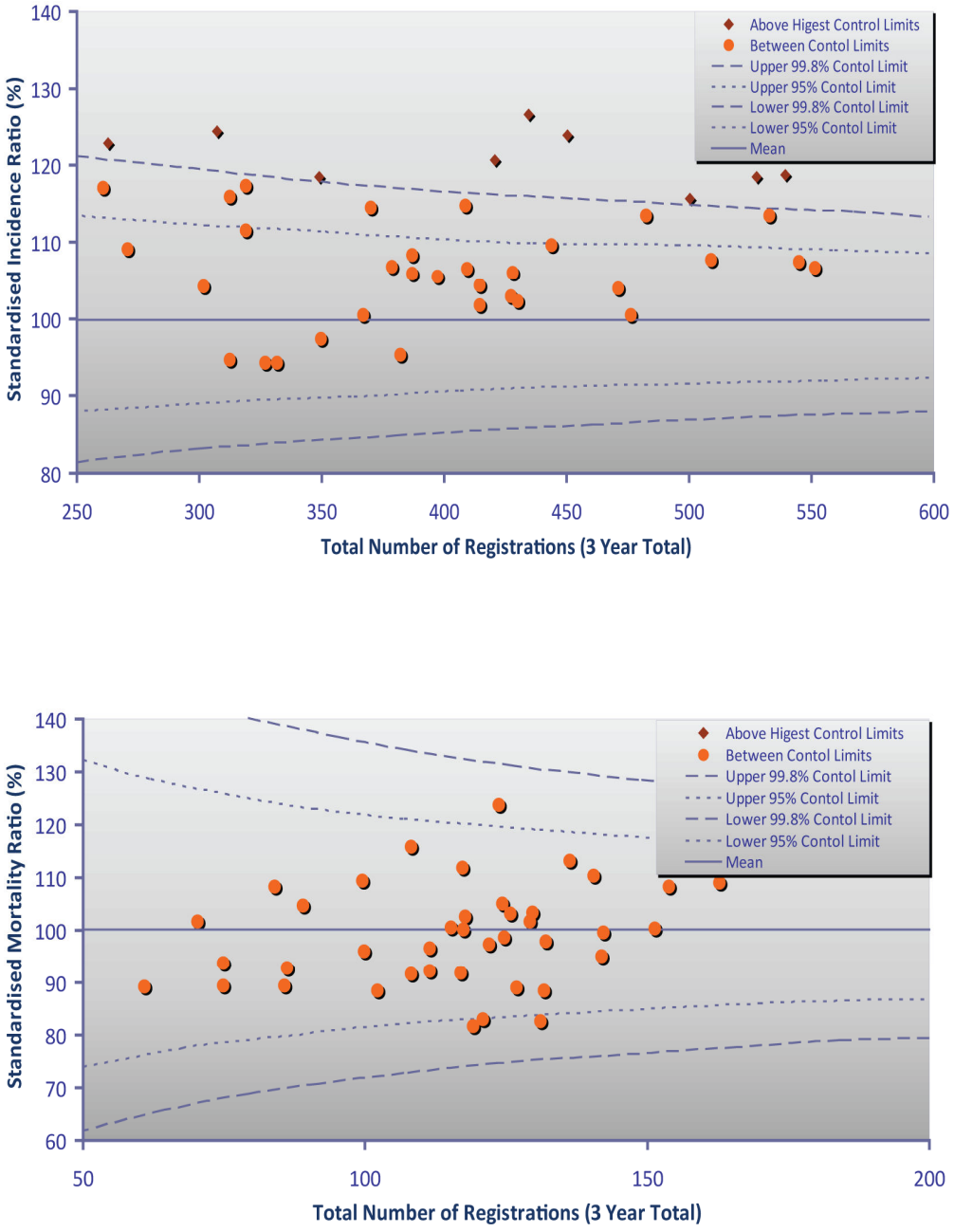


Figure 1.17 Funnel plots of a) standardised incidence ratios and b) standardised mortality ratios for female breast cancer during the period 2001-2003 for Primary Care Trusts (PCTs) in South West Region, Hampshire and Isle of Wight. Figures reproduced from website of South West Public Health Observatory[159]

Figure 1.17, whilst we see a number of PCTs with an SIR above the upper control limit, all PCTs show common cause variation with respect to their SMR. There may be important differences in the disease case-mix between PCTs. Equally it might be that PCTs with higher SIRs, despite the demand upon the screening, referral and treatment pathways, are efficient and effective enough to maintain a relatively low mortality.

Local examination of the processes of care will confirm or refute this claim, but there is the potential to disseminate the experiences within this PCT to improve the health of the entire region. By assessing outcomes, which may include mortality rates, across several disease groups, one can identify within a region if inequality exists and if too many resources are being directed at a select few disease groups at the expense of others.

1.3.7 Conclusions

The application of funnel plots within healthcare and more specifically surgery is increasing but is still to be established as the norm. They are relatively simple to construct, easy to interpret and overcome the limitations of ranking. When assessing performance, they make adjustment for a degree of variation between the units of interest (surgeons, institutions etc.) that result from factors not directly influential or measurable. Funnel plots should not be used as a way to overcome fundamental inadequacies in data quality. Funnel plots can be widely applied acting as a screening tool which, for indicators of performance, outcome or inequality, seeks to identify those areas worthy of further investigation either as areas of potential good practice or areas of concern.

Funnel plots are also important for evaluating evidence emerging from aggregated data such as in meta-analysis, which could influence the practice of evidence-based surgery. Crucial though is the selection of measures used to construct the funnel plot, as this can greatly influence any conclusions made.

Funnel plots should play an increasingly important role as a quality control tool in future surgical performance assessment.

1.3.8 Key points for funnel plots and their emerging application in surgery

- Funnel plots are good for identifying which processes are not yet in control.

- They are practical tools, easy to use, and offer the opportunity to statistically define control limits around measurable outcomes.

- Funnel plots are suited to exploring the volume-outcome relationship in surgery by plotting an observed event rate against a volume of cases.

- Funnel plot methodology avoids the spurious ranking of healthcare providers based on performance and outcomes.

- This alternative performance display may help future quality improvement initiatives to engage healthcare providers in improving transparency of outcomes data.

Chapter 2

ASSESSING THE QUALITY OF THE VOLUME-OUTCOME RELATIONSHIP IN URO-ONCOLOGY

The following chapter was published as:

Mayer EK, Purkayastha S, Athanasiou T, Darzi A, Vale JA. Assessing the quality of the volume-outcome relationship in uro-oncology. BJU Int. 2009;103(3):341-9.

2.1 Introduction

The introduction has highlighted how over the last decade there has been increasing focus on the assessment of the volume-outcome relationship for complex surgery. Proponents for selective referral to high-volume institutions such as the Leapfrog group [42] have based their views on a number of studies suggesting that there exists a positive correlation between volume and outcome. Centralisation of specified services, such as oncological surgery, could therefore in principle improve patient outcomes and so reduce postoperative morbidity and mortality. More recently however, the methodological limitations of existing measures of the volume-outcome relationship have been highlighted by several investigators [88 96]. This has led to careful reconsideration to determine for which operations a volume-outcome relationship truly exists. To facilitate this, the IOM proposed a scoring system against which the methodological quality of published studies could be determined [160].

The development and application of this relevant scoring system to assess the strength of the current best available evidence (methodological quality of volume-outcome relationship studies) may have significant implications in clinical practice and policy setting. Also it is important to note that the volume-outcome literature within urology has never been systematically assessed in this way.

Within England and Wales, there is evident organisational change for the delivery of oncological surgery as a result of the IOG. IOG is a component of the current UK government's health policy, and identifies mechanisms for continually improving the treatment of cancer patients [161]. These changes as applicable to urology have been based on the principle that for urological cancers (including those of the prostate, bladder, & kidney), patients who receive their surgery and post-operative care in hospitals treating larger numbers of patients, are likely to experience better outcomes. The latest guidance for radical urological pelvic cancer surgery dictates that a specialist urology team should have a total catchment population of at least one million and perform at least 50 combined total procedures [161]. It is estimated that this population will produce well over the

minimum cut-off of 50 combined total numbers of prostatectomies and/or cystectomies. There is also guidance as to which patients with renal carcinoma should be managed by a specialist urological cancer team at a cancer centre [161].

The aim of this study, by using a previously validated scoring system, is to systematically quantify the methodological quality of published studies reporting on the volume-outcome relationship for the three major and common uro-oncological surgical procedures; cystectomy, prostatectomy and nephrectomy. This would also allow for correlation of study quality scores and reported outcomes, further understanding of the significance of the reported volume-outcome relationships and facilitation of evidence based clinical and policy decisions. Quality scoring the urological literature to date identifies how future research in this field can be improved and establishes this readily useable scoring system within urology, allowing for future benchmarking.

2.2 Materials and Methods

2.2.1 Study Selection

Multiple searches were performed of electronic databases including: PubMed, EMBASE & Cochrane collaboration. The last search was performed on the 1st May 2007. Articles published in English language literature were searched for using the keywords "outcome", "volume", "urology", "mortality", "morbidity", "cystectomy", "nephrectomy" and "prostatectomy". The bibliographies of included articles were reviewed for further relevant citations and searches were performed for specific authors known to have multiple publications in this area.

2.2.2 Primary inclusion criteria

Studies were included if they met the following criteria:

- Patient cohort treated from 1980 onwards
- Relevant data referring solely to operations for malignant disease

- Health outcome(s) assessed as dependent variables

- When at least hospital and/or surgeon volume was an independent variable

These inclusion criteria are similar to those used by Halm et al. [85]; it was decided to include studies from 1980 onwards such that changes in methodological quality and indeed the magnitude of the volume-outcome relationship would be observed if present. As highlighted by Halm et al. [96] clinical and surgical care prior to 1980 were considerably different because of available treatments and surgical techniques, limiting the relevance to modern day outcomes.

When articles examined more than one procedure, the data for each procedure were considered and analysed as a separate study.

2.2.3 Other considerations

When multiple studies evaluated data derived from the same dataset with any overlap in the studied time period for the same procedure, the following rules were applied:

a. Where studies examined more than one outcome measure, the data for each outcome measure were considered and analysed separately

b. For articles assessing the same procedure, with the same outcome measures and using extracted data from the same database, the study with the lower quality score was excluded. If studies had equal quality scores, the one incorporating the greater time period was included.

2.2.4 Data Extraction

Two authors (EM & TA) independently assessed articles for inclusion/exclusion and quality scored the studies. Any disagreement was resolved by consensus opinion.

2.2.4.1 Outcome measures

All reported postoperative outcomes were considered in the extracted studies. A summary of the assessed outcomes is provided in Table 2.1. Variation in the definition of the outcome measures was present. This heterogeneity only became important for the purpose of the random effects analysis, which was used for mortality. Mortality was defined by studies as either in-hospital mortality only or alternatively as, in-hospital or within 30 days of the index procedure. In order to perform the random effects analysis, mortality odds ratios for different volume providers were used. Frequency of reporting meant that this was only possible at the level of the institution. The highest volume group was maintained as the reference group for each study, such that the mortality odds ratio reflected the ratio of the odds of mortality in a low volume hospital compared to the odds in a high volume hospital. On each occasion, adjusted odds ratios were used, although there was variation between studies in the factors used for adjustment. A summary of the variables by which the extracted odds ratios were adjusted is displayed in Table 2.2. Similarly, variation was found in the categorisation of institutional volumes, Table 2.2.

2.2.4.2 Quality assessment of studies

Extracted studies were assessed quantitatively for quality using the pre-defined scoring system tailored to measure the degree to which the study design is more likely to reveal more accurate information about the magnitude and nature of the relationship between volume and outcome [85]. Question 10 of this scoring system (Table 2.3) was slightly modified in order to incorporate studies which assessed outcomes other than mortality.

	Unit of analysis		Outcome Measures								
Study	Institution Volume	Surgeon Volume	Mortality	Complications	Length of Stay	Re-admission rates	Processes of care	Positive Surgical Margin	DxT/ hormone requirement	Longterm morbidity	Overall Survival
Cystectomy											
Birkmeyer et al. (2003) [99]	✓	✓	✓								
Goodney et al. (2003) [97]	✓				✓	✓					
Elting et al. (2005) [108]	✓		✓	✓							
Hollenbeck et al. (2007) [120]	✓		✓		✓						
Hollenbeck et al. (2007) [107]	✓		✓				✓				
Konety et al. (2006) [118]	✓			✓							
Konety et al. (2005) [105]	✓	✓	✓		✓						
McCabe et al. (2005) [106]	✓		✓		✓						
Prostatectomy											
Begg et al. (2002) [93]	✓	✓	✓	✓						✓	
Ellison et al. (2000) [162]	✓				✓						
Hu et al. (2003) [113]	✓	✓	✓	✓	✓					✓	
Karakiewicz et al. (1998) [163]		✓	✓								✓

Yao et al. (1999) [112]	✓				✓	✓					
Chun et al. (2006) [119]		✓						✓			
Eastham et al. (2003) [111]		✓						✓			
Ellison et al. (2005) [110]	✓								✓		
Hollenbeck et al. (2007) [164]	✓		✓		✓						
Nephrectomy											
Birkmeyer et al. (2002) [165]	✓		✓								
Goodney et al. (2003) [97]	✓				✓	✓					
Davenport et al. (2006) [114]	✓		✓	✓							
Konety et al. (2006) [116]	✓		✓								
Taub et al. (2004) [115]	✓		✓		✓						

Table 2.1 A summary of outcomes measured and unit of volume analysis at a study level

	Volume Categorisation (annual operation number)		Adjustment Variables								
Study	Lowest Volume Institution	Highest Volume Institution	Age	Sex	Race	Year of procedure	Co-existing conditions	Admission type (elective/ non-elective	Measures of social deprivation/ socio-economic status	Cancer grade /or stage	Hospital Factors (e.g. ownership, location, teaching status, bed numbers etc.)
Cystectomy											
Birkmeyer et al. (2003) [99]	<4	>10	✓	✓	✓	✓	✓	✓	✓		✓
Elting et al. (2005) [108]	≤3	>10	✓	✓	✓		✓		✓		✓
Konety et al. (2005) [105]	<1.5	≥2.75	✓	✓	✓		✓		✓		✓
Hollenbeck et al. (2007) [120]	1 - 5	≥20	✓	✓	✓		✓	✓	✓		
Hollenbeck et al. (2007) [107]	0.26 - 2.11	≥6.1	✓	✓	✓	✓	✓	✓	✓	✓	
Nephrectomy											
Birkmeyer et al. (2002) [165]	<7	>31	✓	✓	✓	✓	✓	✓	✓		
Konety et al. (2006) [116]	≤3	>8	✓	✓		✓	✓		✓		✓
Taub et al. (2004) [115]	1 - 14	≥34	✓	✓	✓		✓	✓			

Table 2.2 Display of variation between studies for institutional volume categorisation and factors for which mortality was adjusted

2.2.5 Endpoints

The primary endpoint of this study was to determine the methodological quality scores for studies exploring the volume-outcome relationship in cystectomy, prostatectomy and nephrectomy. The secondary endpoint was to establish correlation between study quality scores and the magnitude of the reported volume-outcome relationship within each of the three operative categories.

2.2.6 Statistical Analysis

For ratio measures of treatment effect, the data were transformed to logarithms (for example as a log odds ratio and the standard error of the log odds ratio). Identical outcomes across studies were combined in logarithmic scales using the generic inverse variance method to determine a pooled effects estimate. The inverse variance method is so named because the weight given to each study is chosen to be the inverse of the variance of the effect estimate (i.e. one over the square of its standard error). A variation on the inverse variance method is to incorporate an assumption that the different studies are estimating different, yet related, treatment effects. This produces a random effects meta-analysis, and the simplest version is known as the DerSimonian and Laird method. To undertake a random effects meta-analysis, the standard errors of the study-specific estimates are adjusted to incorporate a measure of the extent of variation, or heterogeneity, among the treatment effects observed in different studies. The size of this adjustment can be estimated from the treatment effects and standard errors of the studies included in the meta-analysis.

Treatment effects were extracted as ratio measures in the form of an odds ratio (OR) and were therefore converted to natural logarithms. Logarithm standard errors were calculated using the Cochrane Collaborative formula [166].

Because patients operated on in different institutions have varying risk profiles and selection criteria for each surgical technique vary, random effect models were used in this study.

Question	Characteristic	*Score*
1.	Representativeness of sample:	
	Not Representative	0
	Representative	1
2.	Number of hospitals or Surgeons	
	Hospitals < 20 and Surgeons < 50	0
	Hospitals ≥ 20 or Surgeons ≥ 50	1
	Hospitals ≥ 20 and Surgeons ≥ 50	2
3.	Total sample size (cases)	
	< 1000	0
	≥ 1000	1
4.	No. of Adverse Events	
	<20	0
	21-100	1
	>100	2
5.	Unit of analysis	
	Hospital or Surgeon	0
	Both Separately	1
	Both Together	2
	Both Together + further component	3
6.	Appropriateness of patient selection	
	Not measured	0
	Measured separately	1
	Measured and analysed	2
7.	Volume	
	Two categories	0
	Multiple categories	1
8.	Risk adjustment	
	None	0
	Administrative data only	1
	Clinical data only	2
	Clinical data + C >.75 and H/L test positive	3
9.	Clinical processes of care	
	Not measured	0
	One	1
	≥ 2	2
10.	Outcomes†	
	Single outcome measured	0
	≥ 2 outcomes measured	1

Table 2.3 Scoring system used to determine study 'quality'.

†Minimally adapted from original scoring system [85] so as to include studies that assessed outcomes other than mortality.

In the tabulation of our results, squares indicate point estimates of treatment effect (logarithm of odds ratio), with 95% confidence intervals indicated by horizontal bars. The diamond represents the summary logarithm odds ratio from the pooled studies with 95% confidence intervals. Analysis was conducted using the statistical software Review Manager Version 4.2 (The Cochrane Collaboration, Software Update, Oxford).

2.3 Results

A total of 25 articles fulfilled the initial inclusion criteria. Five articles reported on more than one of the three operations of interest. Because the data for each procedure was considered and analysed as a separate study, the 25 articles provided data for 31 studies. Application of the secondary exclusion criteria, to stratify studies using the same datasets to evaluate the same outcomes for the same operations, excluded nine studies, leaving a total of 22 studies for final analysis. The studies were of the following origin; 18 American, two UK, one Canadian and one German. Year of publication ranged from 1998 to 2007, with 19 of the 22 studies being published in the last five years.

The final 22 studies consisted of eight cystectomy, nine prostatectomy and five nephrectomy series. All studies used administrative databases as their data source except for two studies (one single institution prospectively collected dataset and one speciality society prospectively collated database). All but one study had a total sample size greater than 1000 cases. Although nearly all of the studies used administrative databases as their source of data, in only about half of these were the databases representative of the general population. The reason for this originates from many US studies using administrative databases only holding data on patients greater than 65 years of age.

The frequent use of administrative databases is also the explanation for studies persistently scoring poorly for certain data adjustments. In particular, only 18% and 14% of studies respectively addressed, to any degree, clinical processes of care or clinical data risk adjustment. No study explored the appropriateness of

patient selection. 15/22 and 3/22 studies solely used the institution or surgeon as the volume measure respectively. Only four studies appropriately explored the impact of both the institution and surgeon volume on the outcome measures. 15/22 studies assessed at least two outcome measures. In the main, studies assessed mortality (inpatient and/or 30-day) or length of stay as their outcome measures. Frequencies of outcomes measured at a study level are summarised in Figure 2.1.

Across all of the studies the total quality scores ranged from 4 to 11 out of a maximum possible score of 18. The median total quality scores within each of the operation types were 8.5, 9 and 8 for cystectomy, prostatectomy and nephrectomy respectively. Table 2.4, Table 2.5, Table 2.6.

2.3.1 Results of the random effects model

Six studies measured mortality as an outcome following cystectomy. The odds ratio (OR) of mortality between high and low volume providers, including a measure of uncertainty, was extractable from five of these six studies. Overall, mortality was almost twice as high in institutions categorised as low-volume providers (OR=1.88, 95%CI 1.54-2.29) (Figure 2.2). Similarly for nephrectomy mortality was higher in institutions categorised as low-volume providers with an aggregated OR of 1.28, 95% CI 1.10-1.49 (Figure 2.3). Mortality outcome data was extractable from only one of the four prostatectomy studies that assessed it, preventing an analysis of aggregated results. There was insufficient consistency of reporting to allow for random effects analysis to be performed for the remaining outcome measures.

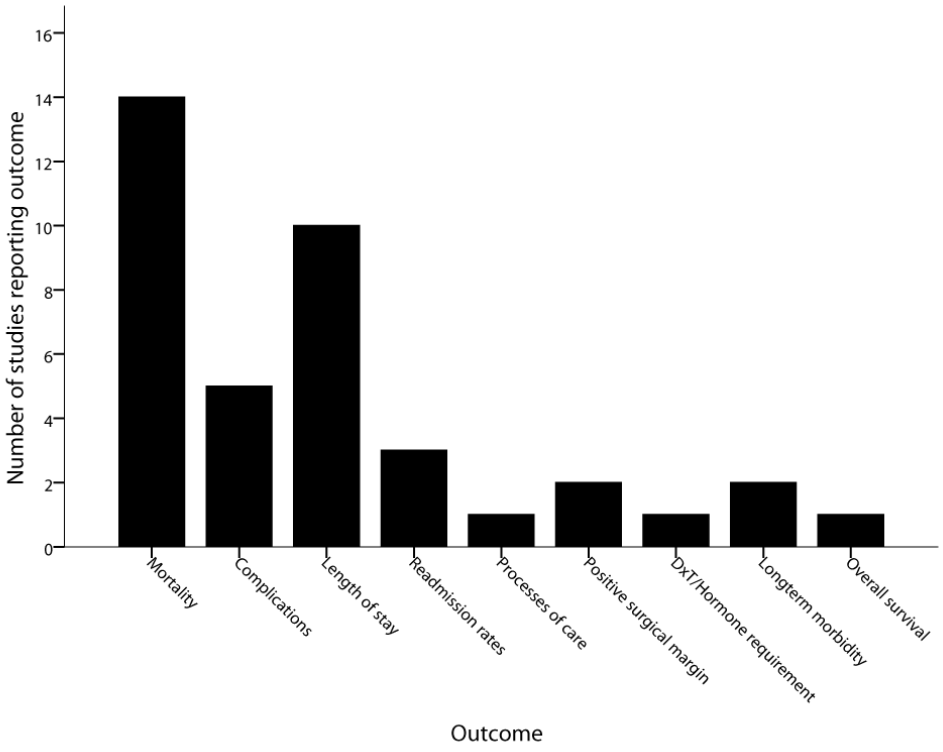


Figure 2.1 Summary of reported outcomes across extracted studies

Study	Scoring Question	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total Quality Score
	Maximum Score	**1**	**2**	**1**	**2**	**3**	**2**	**1**	**3**	**2**	**1**	**18**
Birkmeyer et al. (2003) [99]		0	2	1	2	2	0	1	1	0	0	9
Goodney et al. (2003) [97]		0	2	1	2	0	0	1	1	0	1	8
Elting et al. (2005) [108]		1	2	1	1	0	0	1	1	0	1	8
Hollenbeck et al. (2007) [120]		1	2	1	1	0	0	1	1	2	1	10
Hollenbeck et al. (2007) [107]		0	2	1	2	0	0	1	1	2	0	9
Konety et al. (2006) [118]		1	2	1	2	0	0	1	1	0	0	8
Konety et al. (2005) [105]		1	2	1	2	2	0	1	1	0	1	11
McCabe et al. (2005) [106]		1	2	1	2	0	0	1	0	0	1	8
Median score		**1**	**2**	**1**	**2**	**0**	**0**	**1**	**1**	**0**	**1**	**8.5**

Table 2.4 Methodological quality scores for studies assessing the volume-outcome relationship for cystectomy

Study	Scoring Question	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total Quality Score
	Maximum Score	**1**	**2**	**1**	**2**	**3**	**2**	**1**	**3**	**2**	**1**	**18**
Begg et al. (2002) [93]		0	2	1	2	2	0	1	1	0	1	10
Ellison et al. (2000) [162]		1	2	1	2	0	0	1	1	0	1	9
Hu et al. (2003) [113]		0	2	1	2	3	0	0	1	0	1	10
Karakiewicz et al. (1998) [163]		1	1	1	2	0	0	1	1	0	1	8
Yao et al. (1999) [112]		0	2	1	2	0	0	1	1	0	1	8
Chun et al. (2006) [119]		0	0	1	2	0	0	1	2	0	0	6
Eastham et al. (2003) [111]		0	0	1	2	0	0	1	2	1	0	7
Ellison et al. (2005) [110]		0	2	1	2	0	0	1	2	0	1	9
Hollenbeck et al. (2007) [164]		1	2	1	2	0	0	1	1	0	1	9
Median Score		**0**	**2**	**1**	**2**	**0**	**0**	**1**	**1**	**0**	**1**	**9**

Table 2.5 Methodological quality scores for studies assessing the volume-outcome relationship for prostatectomy

Study	Scoring Question	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Total Quality Score
	Maximum Score	**1**	**2**	**1**	**2**	**3**	**2**	**1**	**3**	**2**	**1**	**18**
Birkmeyer et al. (2002) [165]		0	2	1	2	0	0	1	1	0	0	7
Goodney et al. (2003) [97]		0	2	1	2	0	0	1	1	0	1	8
Davenport et al. (2006) [114]		0	2	0	1	0	0	0	0	0	1	4
Konety et al. (2006) [116]		1	2	1	2	0	0	1	1	1	0	9
Taub et al. (2004) [115]		1	2	1	2	0	0	1	1	0	1	9
Median Score		**0**	**2**	**1**	**2**	**0**	**0**	**1**	**1**	**0**	**1**	**8**

Table 2.6 Methodological quality scores for studies assessing the volume-outcome relationship for nephrectomy

Question 1 - Representativeness of sample, Question 2 - Number of hospitals or doctors, Question 3 - Total sample size (cases), Question 4 - No. of adverse events, Question 5 - Unit of analysis, Question 6 - Appropriateness of patient selection, Question 7 – Volume, Question 8 - Risk adjustment, Question 9 - Clinical processes of care, Question 10 - Outcomes

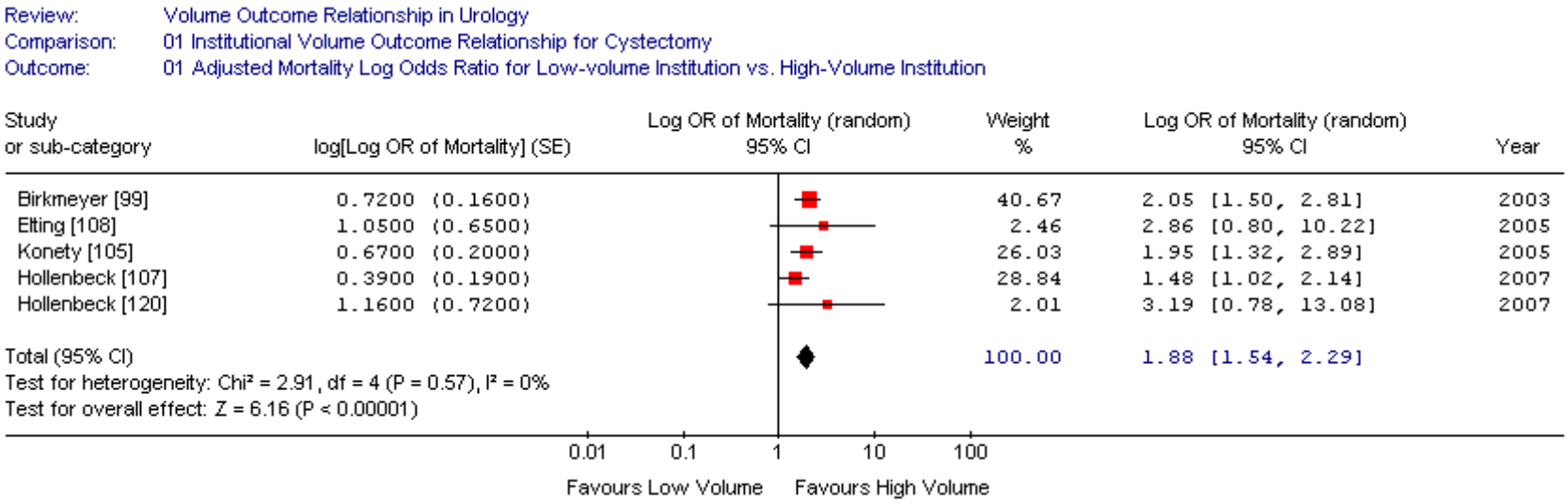


Figure 2.2 Random effects analysis of studies assessing adjusted mortality following cystectomy for low-volume institutions versus high-volume institutions.

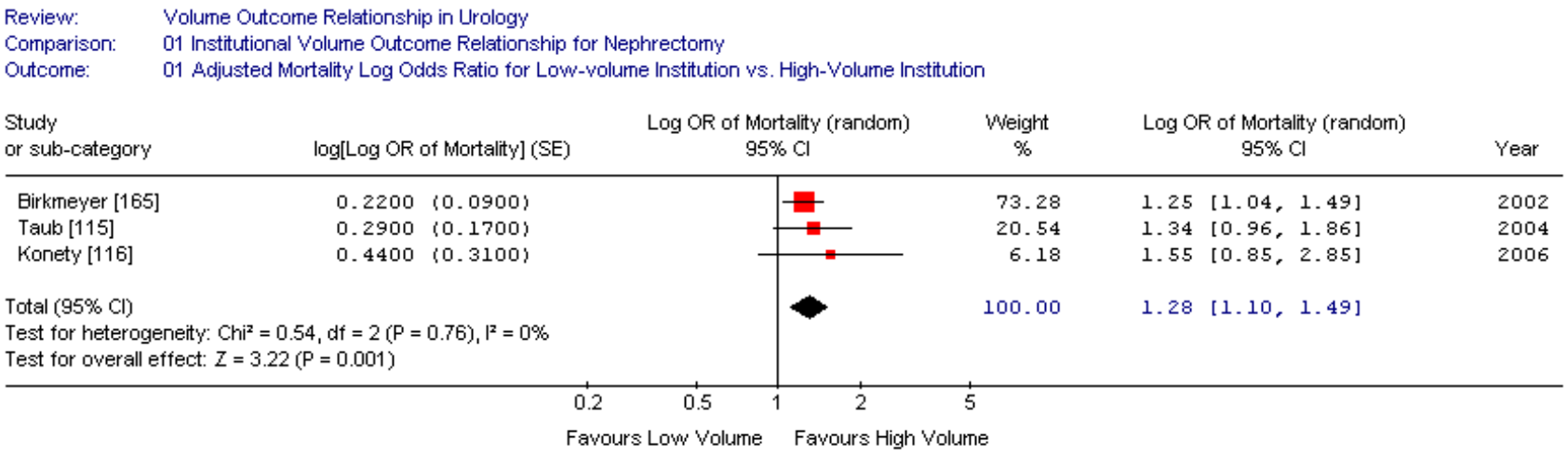


Figure 2.3 Random effects analysis of studies assessing adjusted mortality following nephrectomy for low-volume institutions versus high-volume institutions.

2.4 Discussion

There has been considerable volume-outcome research within the surgical specialities and the relationship remains controversial. Within urology, there is even evidence to suggest that a high-volume surgeon can perform poorly in a high-volume hospital, and a low-volume surgeon can perform well in a low-volume hospital [167]. Previous authors have identified that the methodological quality of reported studies is not as good as might be expected [91 96]. This study has specifically examined and quantified, using a previously validated scoring system, the methodological quality of volume-outcome research in the field of cystectomy, prostatectomy and nephrectomy. Similar to research in other surgical specialities we have shown that for these index procedures in uro-oncology, the methodological quality of the best available evidence was only modest at best. Studies most frequently did not adjust for appropriateness of patient selection and varying clinical processes of care as they rely on administrative databases which do not routinely record such variables.

For cystectomy and nephrectomy it was possible to perform a random effects analysis which showed that for the extracted and quality scored studies, postoperative mortality was higher in low-volume institutions than high-volume institutions. The magnitude of this volume-outcome relationship is greater for the more technically difficult cystectomy than nephrectomy, which is in keeping with previous evidence that has also supported the theory that the magnitude of the relationship is greater with more complex surgery [165]. Similar analysis could not be repeated for mortality following prostatectomy as it was not possible to extract the required data.

Although we could have performed a pooled analysis for length of stay, it was not considered a valid option due to the fact that an insufficient number of studies provided 'adjusted' length of stay as an outcome, making analysis meaningless. In addition, length of stay can vary for reasons other than quality of care such as, for example, funding mechanisms and organisation of care. Random effects analysis was also not possible for the other remaining outcome measures, due to their

infrequent reporting. This point is of most concern as there is a strong argument that for operations such as prostatectomy and nephrectomy, where mortality is infrequent, other outcome measures are more appropriate to help discern differences between providers. For example, positive surgical margin (PSM) rates and postoperative PSA failure rates would be more relevant as measures of cancer control following radical prostatectomy, and morbidity measures such as urinary continence and erectile dysfunction are important from a quality of life perspective. Of the two papers that examined PSM rates and surgeon volume following radical prostatectomy, Eastham et al. demonstrated a significant trend towards higher surgical volume and lower PSM rates, although absolute differences were small. More interestingly, even after adjusting for disease case-mix, surgical date and surgical volume, the individual surgeon remained an independent risk factor for positive margins [111]. Chun et al. demonstrated that surgical volume was a statistically significant predictor of PSM, although when they removed the bias of the one surgeon who contributed over half of the total studied patient cohort, the relationship was not evident [119]. Similarly, relevant outcomes following radical nephrectomy are very different to those following partial 'nephron sparing' nephrectomy and nephroureterectomy. It is important to mention that the pooled effects analysis of mortality following nephrectomy was performed without distinguishing between the different surgical approaches. This was necessitated by the fact that the extracted studies either did not define exactly what they had classified as nephrectomy [165], or analysed their data after having aggregated the different surgical approaches [115 116]. Similarly it was not possible to subanalyse extracted studies reporting on radical cystectomy depending on whether the patient had undergone continent or incontinent urinary diversion.

The secondary endpoint of correlating study quality scores and the magnitude of the reported volume-outcome relationship across the three operative categories was not achievable due to the low number of studies from which relevant data could be extracted and outcome heterogeneity.

The vast majority of studies included had chosen only to explore the relationship of either surgeon or institution volume with outcome. A surgeon's volume may be

indicative of his or her technical skill and case selection, whereas an institution's volume will encompass the perioperative processes and multidisciplinary personnel that will all impact on a patient's quality of care, some of which the surgeon will not be able to directly influence. It is important that the relationship of the institution and surgeon with outcome is assessed as a bare minimum, and ideally multilevel modelling performed to understand the predominant underlying factors (Figure 1.6). The degree of influence of the institution and surgeon will change dependent on the operation; those that require more technical skill such as partial nephrectomy may be more dependent on the surgeon volume than operations which rely more heavily on pre and postoperative input from associated specialities and allied healthcare professionals, such as in cystectomy.

Multilevel modelling takes into account that patients are 'nested' within surgeons and surgeons are 'nested' within institutions, forming a three-level hierarchical structure. The power of multilevel modelling comes from the distributional assumptions that it makes, which allows the simultaneous estimation of between institution variance and institution level predictors, a process which can be repeated for surgeons. These issues cannot be addressed by simplistic statistical models that ignore the structure of the data. Multilevel modelling allows for the influence of each 'level' to be determined independent of others. Only four studies of the 22 combined in some form the influence of the institution and surgeon volume, but only one study explicitly used multilevel modelling. This study demonstrated that the proportion of effect of surgeon volume attributable to hospital volume and conversely hospital volume attributable to surgeon volume, on the odds of operative death between low and high volume providers, were 46% and 39% respectively, reinforcing the importance of a multilevel modelling approach to outcome analysis.

This study incorporates previously designed and validated methodology for quality scoring volume-outcome research, but is unique in its specific focus on uro-oncology. With over 80% of our final cohort of studies having been published in the last five years and 50% since the most recent systematic reviews in the same field [168] [169], it highlights the requirement for this contemporary examination

of the volume-outcome relationship. Most studies however, continue to focus on mortality and length of stay as the outcome of interest.

The use of an almost identical scoring system to assess the methodological quality of the studies as used previously by Halm et al. [85] and Killeen et al. [91] allows us to draw some comparisons. Killeen et al. performed a systematic review of published volume–outcome data relating specifically to surgery for cancer up to 2004 [91]. They determined that the median quality score for cystectomy, prostatectomy and nephrectomy respectively was 10, 9 and 8. This was however limited by a small number of included studies. Halm et al. [85] reviewed the volume-outcome literature across all of healthcare and established that the median quality score across all articles up to 1999 was only 8, less than 50% of the available 18 points. Our study demonstrated an overall median quality score of 8.5. This encompasses our strategy of excluding the study with the lower quality score if two studies assessed the same procedure, with the same outcome measures and used extracted data from the same database. In addition we minimally modified the scoring system to allow for inclusion of studies looking at outcomes other than mortality. This would, if anything, serve to slightly raise the quality score assigned to a study. Despite these considerations, more recent volume-outcome research has therefore not improved, in terms of methodological quality, as compared to previous years. This continues to limit the transferability of the reported correlations to clinical practice. More importantly, it also raises the question whether much of the volume-outcome relationship will be voided by more methodologically robust research? Volume-outcome research has enormous potential to impact on current and future policy initiatives. It should therefore conform to an agreed and accepted methodological and statistical validity before it is 'translated' to healthcare reform.

The application of this quality scoring system within uro-oncology allows for future research to be benchmarked and can be applied to the development of a quality framework within uro-oncology. It is readily useable and can be interpreted by people who may not have an in-depth knowledge of the area, by incorporating many methodological shortcomings common across studies. It

therefore goes further than a previous study which scored uro-oncology volume-outcome research only for case-mix adjustment [168]. In this study, the authors performed a systematic review of the literature to investigate if patients undergoing radical prostatectomy, cystectomy or nephrectomy have improved health outcomes if treated at high volume hospitals or by high volume surgeons. They concluded that this was true for cystectomy and prostatectomy, but simultaneously described limitations in data adjustment that could potentially invalidate their conclusion [168].

Although our study only used 22 studies for the final analysis, the decision to use strict inclusion/exclusion criteria means that the included studies have used data from either a different dataset or a different time period to assess the outcome of interest. This ensures no duplication of the patient populations studied. Previous reviews have excluded studies reporting only from single centres on the basis of potential bias. Single or few centre studies however can provide greater detail with regard to patient and pathological variables that should be adjusted for. Equally, they could allow for measurement of appropriateness of patient selection and specific processes of care, all valuable insight for shaping future research. Two of the final 22 included studies were single or few centre studies; both studies exploring the relationship between surgeon volume and PSM rates following radical prostatectomy adjusting for pathological data [111 119].

The majority of the studies originated from the US which may limit the transferability of the conclusions to a European, and in particular the English, healthcare system, because of the inherent 'structural' differences in the healthcare systems. This is particularly relevant with regard to the organisational change resulting from IOG in England and Wales with only two UK studies meeting inclusion criteria. Furthermore, these studies each achieved a low methodological quality score of eight and four respectively [106 114] (see Table 2.4 & Table 2.6).

The benchmarking of future research utilising a validated scoring system such as that used in this study will help to confirm adherence to a quality framework and

validate the strength of it's conclusions and translation to changing healthcare policy and clinical practice.

2.5 Conclusions

Centralisation of radical urological pelvic cancer surgery forms an important component of IOG in urology. Service reconfiguration relies on the translation of robust health service research to support and enable it. This study has demonstrated, by using a previously validated quality scoring system, that the methodological quality of volume-outcome research as applied to cystectomy prostatectomy and nephrectomy is only modest at best. It can also be said, by comparison to previous similar studies, that despite the continued expansion of research in this area, study 'quality' has remained relatively static. There is a paucity of data applicable to the English healthcare system. The following two studies of this thesis will explore the impact of improved statistical techniques on volume-outcome research as related to radical cystectomy in England. Radical cystectomy was chosen because of the greater magnitude of volume-outcome relationship demonstrated by the random effects analysis. Radical cystectomy fulfils the criteria as an index procedure for assessing the volume-outcome relationship in that it is performed with sufficient frequency and has mortality and morbidity rates sufficient, in principle, to be able discern differences between providers of varying volume.

2.6 Key points arising from study one

- The methodological quality of existing volume-outcome research is only modest at best

- Recommendations made by the Institute of Medicine in 2000 for improving the quality of volume-outcome research appear to have had limited impact

- The majority of volume-outcome research has been conducted in the US, which limits transferability to the English healthcare system

- Random effects analysis of pooled data shows that for both cystectomy and nephrectomy, postoperative mortality was higher in low-volume institutions than high-volume institutions

- The magnitude of the volume-outcome relationship appears to be greater for the more morbid cystectomy operation as compared to radical nephrectomy

- The incorporation of more robust methodology when analysing existing administrative datasets could improve the validity and usefulness of volume-outcome research.

Chapter 3

THE VOLUME-MORTALITY RELATIONSHIP FOR RADICAL CYSTECTOMY IN ENGLAND

The following chapter was published as:

Mayer EK, Bottle A, Darzi AW, Athanasiou T, Vale JA. The volume-mortality relation for radical cystectomy in England: retrospective analysis of hospital episode statistics. BMJ 2010 Mar 19;340:c1128. doi: 10.1136/bmj.c1128.

3.1 Introduction

The previous study has quantified the methodological limitations of existing volume-outcome research. To incorporate all of the IOM recommendations for improving the quality of volume-outcome research, future research would need to incorporate data that has been prospectively collected specifically for this purpose. Some of the methodological shortcomings of the existing studies arise from the fact that they have used administrative datasets and these cannot be improved upon. It is also evident however, that some of the methodological shortcomings, such as multilevel modelling, are not limited by the use of administrative datasets. It is also feasible that by the novel manipulation of, and by combining contemporary administrative datasets, data can be appropriately adjusted for structural and process of care confounders.

The IOM methodological scoring system [85] highlighted several important statistical and methodological considerations: the hierarchical structure of the institutional and surgeon influences (multi-level modelling), appropriate handling of the provider volume variable, need for adjustment of structural/process of care measures, and measuring outcomes other than mortality. The aim of this study is to create an improved statistical/methodological framework, taking these issues into account and use it to investigate the institutional and surgeon volume-mortality relationship for radical cystectomy in England. The subsequent chapter then further examines the volume-outcome relationships by expanding the model to include outcomes other than mortality

3.2 Methods

3.2.1 Data extraction

Data for inpatient elective cystectomies (ADMIMETH=11, 12 or 13 with the OPCS4 code M34 occurring in any procedure field in any episode) were taken from Hospital Episode Statistics (HES) for the six financial years 2000/1 to

2005/6 and from the Secondary Users Service for 2006/7. HES is the national statistical administrative database for England of the care provided by NHS hospitals. Each record in HES constitutes a finished consultant episode, which covers the continuous period during which a patient, admitted to a hospital, is under the care of the same consultant, whose General Medical Council (GMC) code is recorded (a unique identification code used by the doctor's regulatory body, the GMC). Each record therefore has an identifiable consultant, as an individual, responsible for that patient's care episode; it does not assume the operating surgeon. However, radical cystectomy is of such complexity that it will at the very least be a consultant supervised operation. Episodes were linked into admissions using the patient's date of birth, sex and postcode, the hospital and date of admission, and admissions were linked together if the patient was transferred to another Trust. A Hospital Trust provides secondary healthcare services within the NHS, which can either be planned specialist medical care or surgery or emergency care. A Trust may comprise several hospital sites.

Admissions were excluded if they were emergencies, had invalid age, sex or length of stay, were day cases or did not have a primary diagnosis of cancer (ICD10 C66, C67, C68, D090).

Twenty four patients at the patient ID level had a total of 26 duplicate records. It was decided not to remove them on the basis that it was not possible to determine which was the 'real' record and the duplicates only made up 0.3% of the total case number. None of these records had a death recorded.

3.2.2 Assignment of institutional and surgeon volume bands

The number of cystectomies was counted by year and each of consultant code and Trust. To account for Trust mergers, Trust codes were unified to reflect their status as of April 2007. Trusts (institutions) were excluded due to very low volume if they either had fewer than three years of data or had an annual cystectomy rate of less than two. This was to try to capture institutions with genuine activity (i.e. not due to coding errors). If a provider had less than three years of data, it was considered unlikely that they were performing cystectomy

regularly during the time period analysed. This had to be balanced against not excluding providers (at the surgeon level) that either retired or were newly appointed during the analysed period and may therefore only appear for three or fewer years. We felt that this three year cut-off, in combination with an annual cystectomy rate of two, would identify cases more likely appearing as a result of coding error, while not inappropriately removing large numbers of low-volume providers, i.e. regularly performing between two and 10 operations per year at the institution level, or between one and five operations per year at the surgeon level. Thirteen institutions providing 41 cases (0.5% of total cases) were excluded on this basis. The remainder were put into three volume bands of roughly equal total operation numbers (Table 3.1).

Allocation of consultants to volume bands was performed in a similar fashion, although because a surgeon could operate at more than one institution, and hence more than one institutional volume band, this had to be performed in two stages. After excluding a single record with a missing consultant code, the surgeon's average annual cystectomy rate was calculated. Consultants were excluded if their average cystectomy rate was less than or equal to two cases per year and they had fewer than three years of data. Exclusion was performed without summating caseload of consultants operating at multiple institutions because coding errors occur at the level of the institution. One hundred and eighty nine consultants providing 313 operations (3.6% of total cases) were excluded. Following exclusion, the annual cystectomy rate for those surgeons operating at more than one institution was recalculated using the maximum number of years practised from whichever provider and total summation of operations. Consultants were then divided into three volume bands of roughly equal total operation numbers (Table 3.1).

	Institutional volume band (Annual cystectomy rate)		
Year	Low (>2 and < 10)	Medium (≥10 and < 16)	High (≥16)
2000/1	453	331	346
2001/2	437	383	359
2002/3	495	370	378
2003/4	461	392	398
2004/5	431	387	416
2005/6	375	425	463
2006/7	384	488	424
Total operations in volume band	3036	2776	2784
Number of institutions in volume band	84	31	19
	Surgeon volume band (Annual cystectomy rate)		
Year	Low (≥1 and < 5)	Medium (≥5 and ≤8)	High (>8)
2000/1	485	315	264
2001/2	511	306	303
2002/3	506	357	356
2003/4	480	388	359
2004/5	419	366	413
2005/6	334	405	490
2006/7	314	423	529
Total operations in volume band	3049	2560	2714
Number of surgeons in volume band	228	78	40

Table 3.1 Breakdown of number of operations for each institutional and surgeon volume bands generated prior to regression modelling

3.2.3 Model development

The data had a three-level structure that needed to be taken into account to avoid potentially biased estimates: surgeons treat a number of patients and each hospital has more than one surgeon (and some surgeons operate at more than one hospital). The final complex model was constructed in four stages. The first stage simply used logistic regression and the two volume variables (surgeon volume and institutional volume), with no adjustment for case-mix and ignoring the "clustering" (of patients within surgeons and surgeons within hospitals) (model one). The second added the case-mix variables (model two). The third took into account the clustering (model three) and the final model (model four) included the structural and process variables.

Volume-mortality relationships were presented as odds ratios using the low volume band as the reference group. The interaction effects between institutional and surgeon volumes were assessed.

3.2.3.1 Case-mix adjustment

Adjusted was made for gender, age, Carstairs deprivation quintile, Charlson comorbidity score [170] and year of operation. Carstairs deprivation scores for every output area (about 500 residents on average) in England were converted into quintiles with equal population in each and assigned to every record via the postcode using a look-up file from ONS Gridlink.[171] A sixth quintile was assigned to records with missing or invalid postcodes. These are the main case-mix variables available in HES for elective patients.

3.2.3.2 Multi-level modelling & adjustment for institutional volume and surgeon volume

The "clustering" was handled by fitting multilevel models with random effects for surgeons and hospitals, a very commonly used approach.[92] Both two- and three-level models were fitted. The two-level model recognises that some surgeons (around 5% in the sample) work at more than one hospital (surgeons and hospitals

are said to be "cross-classified"). The three-level model asserts that this proportion is small enough to ignore and, with continuing centralisation of services, is likely to decrease further in the future. Only 18 consultants of a total of 346 worked at more than one institution and at most two. The exact cross-classification of each surgeon with their respective hospital(s) in the models was not specified, but it is believed that this would not give different results from either the two- or the three-level approaches that were tried.

In practice, both sets of models gave very similar results (odds ratio point estimates typically differed by 0.05 at most and often by much less) and only the three-level model results are presented. Goodness of fit was assessed using the ratio of the generalised chi-square statistic to the degrees of freedom and was found to be close to 1 (i.e. very good fit) in all models.

3.2.3.3 Structural and process of care adjustment

Information for the following factors was obtained from The Information Centre for health and social care and Hospital Activity Statistics: average occupied acute bed rate, ratio of number of nurses to occupied bed count, ratio of number of critical care beds to total beds, ratio of registrars (all specialities) to occupied bed number, ratio of urology registrars to total urology episodes, and readmission rate following operation cancellation (percentage of patients not readmitted within 28 days following last minute cancellation). Total urology episodes and operation waiting times (period between operation booking date to operation date) were derived from HES. Trust (institutional) teaching status was derived from information available on the National Patient Safety Agency website: see Appendix A for further details. Selection of these variables was influenced by the existing volume-outcome literature, and therefore included staffing characteristics, teaching status and hospital capacity characteristics. The availability of data routinely collected within England did have some impact on the final variable definitions chosen. (Appendix A)

There were significant differences in the structural and process of care variables across the institutional volume groups Table 3.2.

3.2.4 Outcome measure

All-cause mortality within 30 days of the procedure (or admission if the procedure date was missing) was assessed in two ways: in-hospital mortality and mortality either in or out of hospital (total mortality). The latter was calculated using a date of death linked to HES by the Office of National Statistics. This was unavailable for 2006/7, but to avoid losing a year's records and to maintain statistical power, it was assumed that there were no out-of-hospital deaths in 2006/7; a sensitivity analysis was performed on the small shortfall.

3.2.5 Statistical Analysis

Within- and between-volume group variations in case-mix variables were assessed using Chi-squared tests. All logistic regression and other statistical analysis except for multi-level modelling (which was done using PROC GLIMMIX in SAS version 9.2, North Carolina, USA) were performed using the SPSS package version 14 for windows (SPSS, Inc., Chicago, IL).

	Low volume	Medium volume	High volume	Overall mean	One-way ANOVA
Ratio of number of nurses to occupied bed count	1.59 (1.12 to 2.60)	1.73 (1.31 to 2.76)	1.75 (1.36 to 2.14)	1.69	<0.01
Ratio of urology registrars to total urology episodes (x10,000)	4.1 (0.34 to 21)	4.3 (0.2 to 12)	6.3 (0.9 to 15)	4.9	<0.01
Operation waiting time	31.7 (1 to 394)	32.6 (1 to 339)	30.3 (1 to 251)	31.5	<0.01
Ratio of number of critical care beds to total beds	0.021 (0.01 to 0.062)	0.024 (0.009 to 0.055)	0.026 (0.013 to 0.048)	0.023	<0.01
Ratio of registrars (all specialities) to occupied bed number	0.09 (0.02 to 0.35)	0.13 (0.06 to 0.47)	0.15 (0.04 to 0.28)	0.12	<0.01
Trust teaching status (% within volume group)	9.7	19.1	41.2	22.9 (overall %)	<0.01*
Average occupied acute bed rate	0.85 (0.73 to 0.94)	0.84 (0.76 to 0.92)	0.85 (0.79 to 0.94)	0.85	<0.01
Readmission rate following operation cancellation	0.11 (0.00 to 0.42)	0.10 (0.00 to 0.35)	0.13 (0.01 to 0.33)	0.11	<0.01
Total beds, all sectors, available	5372 (2143 to 10441)	7356 (1686 to 16745)	9783 (1651 to 18044)	7441	<0.01
Total acute beds available	4056 (1723 to 8730)	5822 (1686 to 14285)	7942 (1651 to 13506)	5884	<0.01
Total urology episodes	25978 (8688 to 63818)	39640 (5132 to 83112)	49855 (3402 to 89385)	38123	<0.01

Table 3.2 Variation in the structural and process of care measures between the institutional volume groups. Minimum/maximum in parentheses

*** Chi squared p value**

3.3 Results

The number of radical cystectomies performed each year increased steadily from 1130 in 2000/01 to 1296 in 2006/2007 ($p = 0.005$ for linear trend), Figure 3.1. Table 3.1 shows the distribution by institutional/surgeon volume bands. There was a significant downward trend in the total mortality rate from 3.5% in 2000/01 to 2.1% in 2005/06 (out-of-hospital deaths not available for 2006/07), Table 3.3. Figure 3.2 and Figure 3.3 show observed (crude) mortality rates by institutional/surgeon volume tertiles.

3.3.1 Patient demographics and institutional characteristics

The proportion of treated patients aged 75 years and over increased from 19.8% in 2000/01 to 22.7% in 2006/07 ($p = 0.03$). There was a significantly greater proportion of males in 2000/01 and 2003/04 ($p = 0.03$). The proportion of patients in the highest Charlson category (six+) varied significantly year on year, but with no linear trend across the years. The median Charlson score and Carstairs quintile were two and three respectively for every year (Table 3.5).

Medium-volume institutions treated proportionally fewer patients in the most deprived Carstairs quintile (five) and a greater percentage of patients older than 75 years ($p<0.01$, Table 3.6). There was a statistically significant increase in the number of institutions with teaching status from the low-volume group through to the high-volume group ($p<0.01$). Medium and high-volume surgeons treated a greater proportion of males and fewer patients in the most deprived Carstairs quintile ($p<0.01$). Medium- and high-volume surgeons more frequently worked in institutions with teaching status than low-volume surgeons.

In the multi-level model, the 'ratio of number of critical care beds to total beds' was the only structural variable significantly ($p<0.05$) and independently related to both mortality outcome measures.

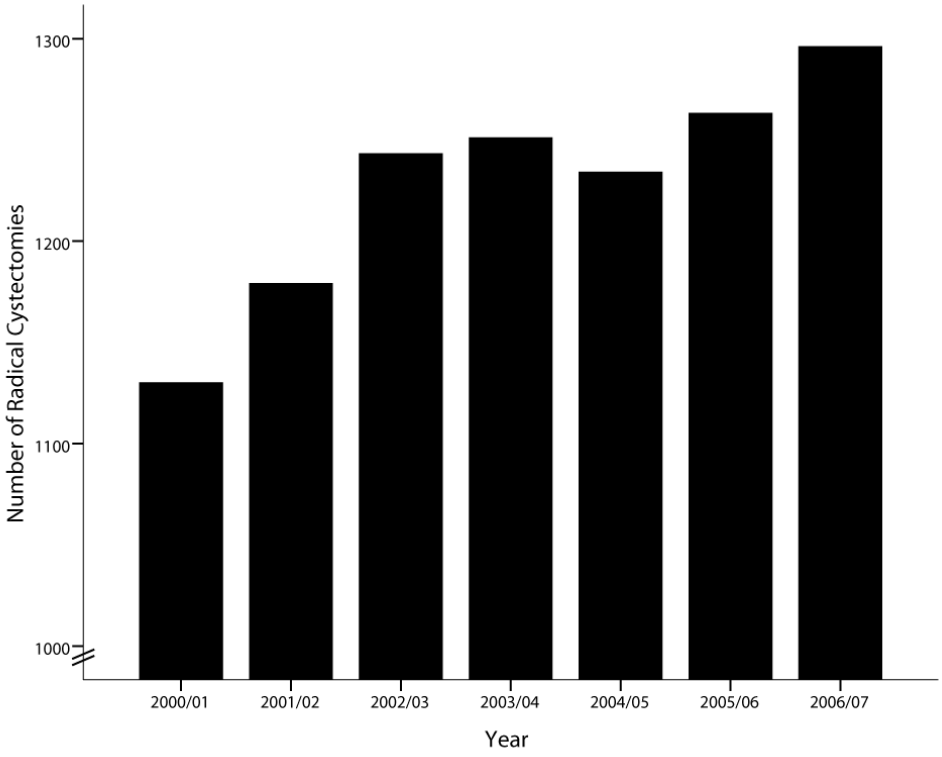


Figure 3.1 Number of radical cystectomies performed each year in England between 2000/01 and 2006/07 inclusive.

Financial year	In-hospital mortality (30 days) n (%)	Total mortality (30 days) n (%)	Proportion of in-hospital to total mortality
2000/1	32 (2.8)	39 (3.5)	0.82
2001/2	38 (3.2)	40 (3.4)	0.95
2002/3	34 (2.7)	36 (2.9)	0.94
2003/4	25 (2.0)	33 (2.6)	0.76
2004/5	40 (3.2)	42 (3.4)	0.95
2005/6	24 (1.9)	26 (2.1)	0.92
2006/7	30 (2.3)	30 (2.3)	1.00†
p value for linear trend across years	0.13	0.03	-

Table 3.3 Summary of mortality outcome measures by year for radical cystectomy (unadjusted data)

† 2006/7 uses Secondary User Survey to extract data. This does not include out of hospital mortality meaning in-patient hospital death rate equals total death rate.

	In-hospital mortality (30 days) (n)	Total mortality (30 days) (n)
Institutional volume band		
Low	85	90
Medium	82	90
High	56	66
Surgeon volume band		
Low	86	92
Medium	74	81
High	52	62

Table 3.4 Observed numbers of in-hospital and total mortality stratified by institutional/surgeon volume bands

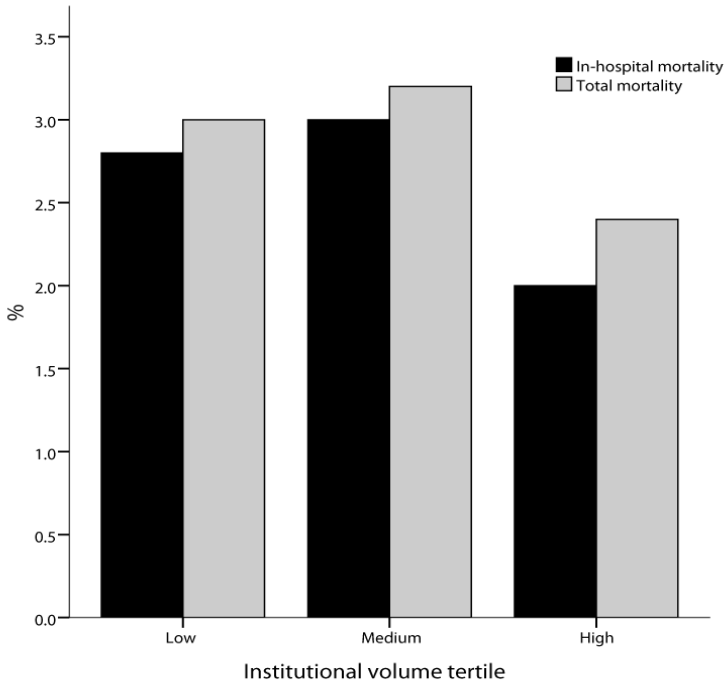


Figure 3.2 Crude mortality rates (%) stratified by institutional volume tertiles.

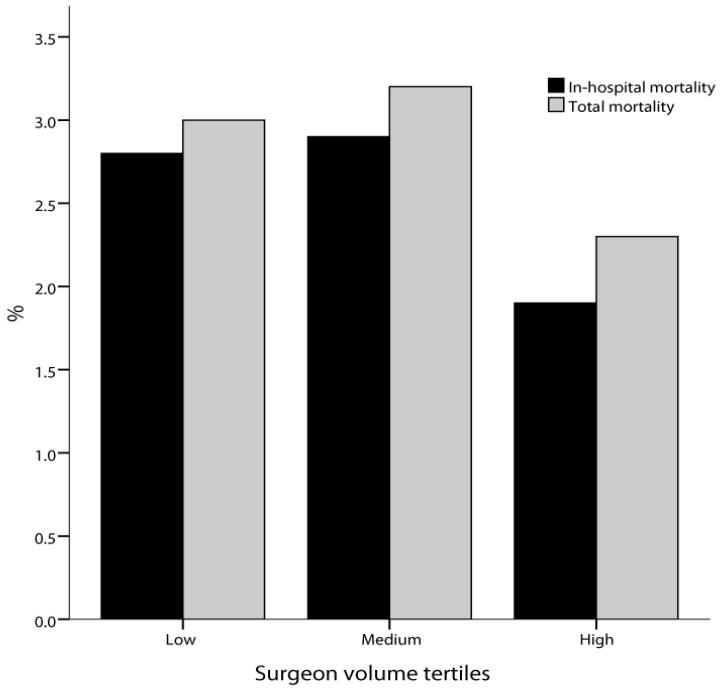


Figure 3.3 Crude mortality rates (%) stratified by surgeon volume tertiles

Financial year	Percentage of males	Charlson Score Percentage in category 6+	Age Percentage > 75 years	Carstairs Score Percentage in category 5
2000/1	78.0	1.9	19.8	15.7
2001/2	74.2	1.2	18.0	14.3
2002/3	75.2	1.7	19.2	16.3
2003/4	79.3	1.4	20	13.2
2004/5	75.9	2.3	20.7	14.5
2005/6	74.5	2.9	22.6	13.9
2006/7	75.2	1.9	22.7	13.0
Mean across years	**76.0**	**1.9**	**20.5**	**14.4**
Chi squared p value	**0.03**	**0.05**	**0.03**	**0.2**

Table 3.5 Patient demographics across each of the years analysed

Provider	Percentage of males	Charlson Score Percentage in category 6+	Age Percentage > 75 years	Carstairs Score Percentage in category 5	Teaching status Percentage within volume group
Institution					
Low volume	75.3	1.6	17.8	15.2	9.7
Medium volume	76.3	1.9	23.0	12.4	19.1
High volume	76.5	2.1	20.9	15.5	41.2
Chi squared p value	0.47	0.45	<0.01	<0.01	<0.01
Surgeon					
Low volume	74.2	1.6	19.4	16.0	19.4
Medium volume	76.3	2.0	21.2	13.9	25.4
High volume	78.2	2.3	21.0	12.6	24.2
Chi squared p value	<0.01	0.12	0.19	<0.01	<0.01

Table 3.6 Differences in patient/institution characteristics by provider volume group

3.3.1.2 Volume-mortality relationship

For in-hospital mortality, the OR for medium-volume institutions in model three was 1.16, but rose to 1.72 (1.00 to 2.98, p=0.05) in model four (which adjusted for the structural/process of care variables). Adjusting for the ratio of the number of nurses to occupied beds increased the OR to 1.31; this became 1.52 after adjusting for the ratio of urology registrars to total urology episodes and 1.62 after including operation waiting times. Adding the remaining structural/process of care variables led to the model four OR of 1.72. A similar pattern was seen for total mortality (model four OR=1.82, 1.08 to 3.06, p=0.02), (Table 3.7). There was no statistically significant difference seen using models one, two or three. The only significant result for high-volume institutions was a lower odds of in-hospital mortality in model two (OR 0.67 (0.47 to 0.97, p=0.03)) (Figure 3.4).

High-volume surgeons had a statistically significant lower odds of in-hospital mortality for models one and two (OR 0.67 (0.48 to 0.95, p=0.03) and OR 0.64 (0.44 to 0.91, p=0.01) respectively), but this became non-significant at the 5% level in models three and four. No statistically significant differences were seen for total mortality (Figure 3.5).

Table 3.8 displays the crude and adjusted odds ratios for in-hospital and total mortality across both institution and surgeon volume tertiles.

Figure 3.6 shows a summary of the adjusted probabilities for in-hospital and total mortality across both the institutional and surgeon volume bands. Adjusted probabilities were calculated using model four.

	Medium Volume Institutions		High Volume Institutions	
	In-hospital mortality	Total mortality	In-hospital mortality	Total mortality
No structural variable adjustment	1.16	1.18	0.80	0.86
Ratio of number of nurses to occupied bed count	1.31	1.34	0.91	0.98
Ratio of urology registrars to total urology episodes	1.52	1.58	1.02	1.10
Operation waiting time	1.62	1.67	1.07	1.15
All structural variables included in adjustment (model 4 OR)	1.72	1.82	1.09	1.19

Table 3.7 Change in ORs for medium and high-volume institutions with cumulative adjustment of each structural variable. Adding the remaining structural/process of care variables led to the final model four odds ratio.

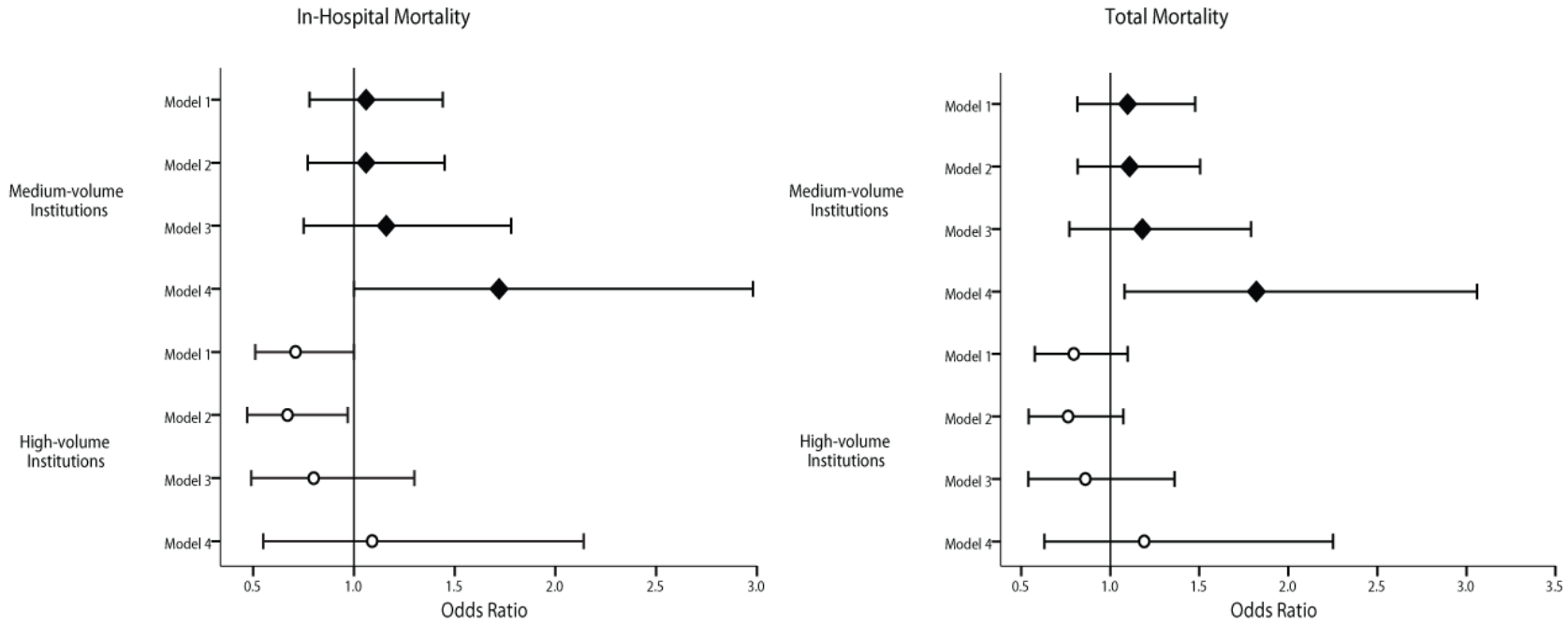


Figure 3.4 Summary of odds ratio and confidence intervals for institutional volume-mortality relationships across all four models

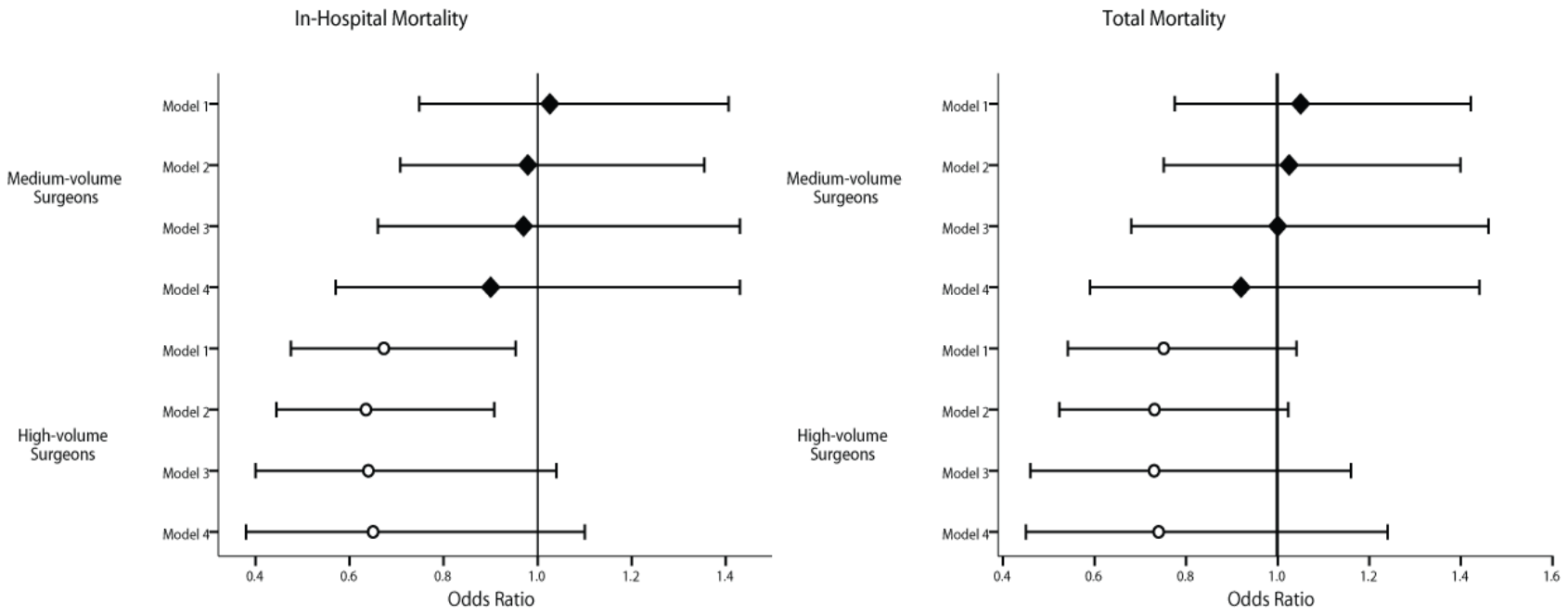


Figure 3.5 Summary of odds ratio and confidence intervals for surgeon volume-mortality relationships across all four models

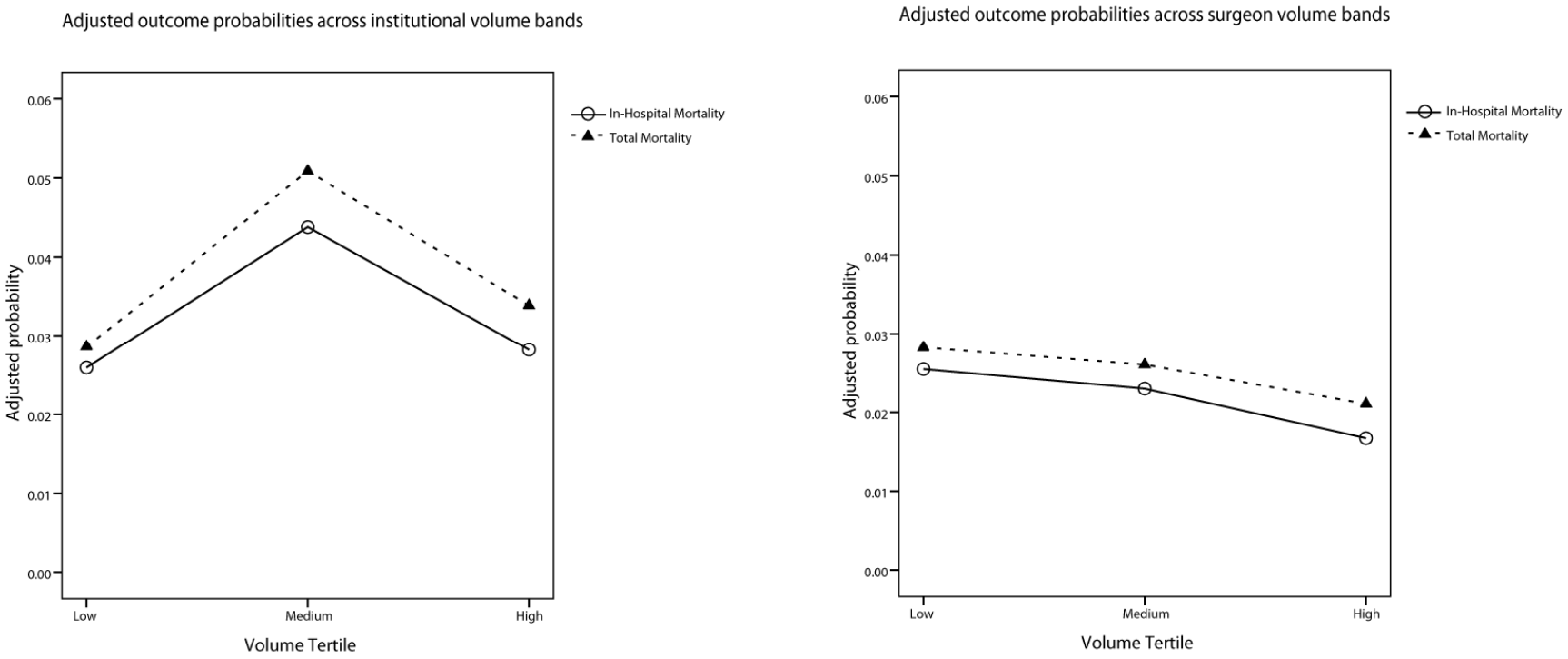


Figure 3.6 Summary of adjusted probabilities (derived from model 4) for in-hospital and total mortality across institutional and surgeon volume bands

3.3.2 Sensitivity analysis for total mortality

On average, there were four out-of-hospital deaths per year and 23 in total. For the sensitivity analysis, four surviving patients were randomly allocated to have a "total 30-day death": two in the low-volume and two in the high-volume institution groups. In clinical trials with mortality as an end-point, it is usual in sensitivity analyses to assume higher death rates in the drop-outs rather than assume that the deaths are missing at random. This allocation was chosen because of the finding that it was the medium-volume group that had the highest odds ratio from model four. The resulting odds ratio was a little lower at 1.66 (1.00 to 2.76, p=0.049).

3.3.2.1 Interaction effect

Testing for an interaction between the two volume variables resulted in a significant interaction term (p=0.035) but non-significant main effects (p>0.1), suggesting no evidence for a convincing interaction effect.

Outcome Measure									
	In-hospital mortality					Total mortality			
	Crude OR	Case-mix adjusted OR(CI, p value)	MLM OR	Structural variables OR (CI, p value)		Crude OR	Case-mix adjusted OR (CI, p value)	MLM OR	Structural variables OR (CI, p value)
Institution									
Low volume (ref)	1.00	1.00	1.00	1.00		1.00	1.00	1.00	1.00
Medium volume	1.06	1.06 (0.77-1.45, 0.73)	1.16 (0.75-1.78, 0.51)	1.72 (1.00-2.98, 0.05)		1.10	1.11 (0.82-1.51, 0.51)	1.18 (0.77-1.79, 0.45)	1.82 (1.08-3.06, 0.02)
High volume	0.71	0.67 (0.47-0.97, 0.03)	0.80 (0.49-1.30, 0.37)	1.09 (0.55-2.14, 0.81)		0.80	0.76 (0.54-1.07, 0.12)	0.86 (0.54-1.36, 0.52)	1.19 (0.63-2.25, 0.59)
Surgeon									
Low volume (ref)	1.00	1.00	1.00	1.00		1.00	1.00	1.00	1.00
Medium volume	1.03	0.98 (0.71-1.35, 0.90)	0.97 (0.66-1.43, 0.87)	0.90 (0.57-1.43, 0.67)		1.05	1.03 (0.75-1.40, 0.88)	1.00 (0.68-1.46, 0.98)	0.92 (0.59-1.44, 0.72)
High volume	0.67	0.64 (0.44-0.91, 0.01)	0.64 (0.40-1.04, 0.07)	0.65 (0.38-1.10, 0.11)		0.75	0.73 (0.52-1.02, 0.07)	0.73 (0.46-1.16, 0.19)	0.74 (0.45-1.24, 0.25)

Table 3.8 Differences in mortality across both institution and surgeon volume categories, displayed as crude and adjusted Odds Ratios (OR)

3.4 Discussion

Principal findings

When compared with low-volume institutions, patients undergoing radical cystectomy in medium-volume institutions had 72% and 82% greater odds of 30-day in-hospital and total mortality respectively. This increased risk in mortality was only seen after adjustment for institutional structural and process variables. The nursing and urology registrar staffing levels appeared to have the greatest influence, in keeping with previous studies [108 120].

Although not significant at the 5% level, there was weak evidence of reduced odds of in-hospital mortality (by 35%) for the high-volume surgeon. Surgeon volume also appeared to have some protective effect on in-hospital mortality at the institutional level. After patient case-mix adjustment alone, high-volume institutions displayed statistically significantly lower odds of in-hospital mortality. This effect disappeared after adjusting for surgeon volume.

There was relatively little difference in the crude OR and case-mix adjusted OR of mortality at both the institutional and surgeon level. This is interesting in itself, as surgeons often use case-mix variation to explain differences in performance. It may be that case-mix adjustment, using variables as incorporated in this study, is not always necessary for analysis of effects at a highly aggregated level, as opposed to risk prediction at an individual patient level. Hollenbeck et al.[107] also reported minimal difference in the unadjusted and adjusted (for patient demographics) OR of mortality for cystectomy between low-volume and high-volume institutions.

Strengths and weaknesses of the study

As described earlier in study one, the median IOM quality score for previous studies appraising the volume-outcome relationship for cystectomy was modest at 8.5 (range 8-11). Using the same scoring criteria, this current study would achieve

a higher score of 13 out of a possible 18. One mark is lost because, in this chapter, only the volume-mortality relationship is described. As outcomes other than mortality are presented in the next chapter, overall this improved methodology would score 14 out of 18. Four marks were not achievable because an administrative database was used and the appropriateness of patient selection for cystectomy could not therefore be determined and clinical data for risk-adjustment could not be used. Methods to circumvent this limitation could include linkage of administrative datasets with local or speciality society datasets, which would simultaneously serve to confirm coding accuracy. The need to centrally support such initiatives across the multiple available datasets has been acknowledged [172]. Although it was not possible to adjust data for cancer stage and grade, when this has been possible it appeared to have little impact on the reported findings [107 173].

Sub-analysis of mortality risk for patients who had either an ileal conduit diversion or continent diversion was not achievable because of under-reporting of continent diversions. 79% of patients were recorded as undergoing ileal conduit diversion, 0.5% as continent diversion and for 20% of patients it could not be determined. Similar difficulties have also been experienced with US administrative databases [108]. The reliability of coding within HES has been questioned by clinicians for some time. In England, training of coding staff, regular monitoring of data quality, and a commissioning system which financially reimburses healthcare providers according to case-mix adjusted activity has improved the quality of data completeness and accuracy since the early years of HES [174 175]. If institutional resources for coding are proportional to the overall workload of that institution, it is possible that miscoding rates are not systematically different by volume of hospital. This will have biased the estimated ORs towards 1 and therefore underestimated any volume-mortality relationship.

The mortality outcome definition only used a 30-day follow-up, as is usual [173]. Overall 90-day mortality is known to be higher than 30-day mortality following cystectomy [176], but whether this differs across providers is unclear; work on this, and longer-term cancer specific survival, should consider postoperative structural and process of care factors such as oncological ancillary services [120]. The only

previous study which used HES to assess mortality following cystectomy only included in-hospital mortality [106].

It was assumed that in 2006/7 there were no out-of-hospital deaths, rather than the four on average for the other years. Although this assumption is not expected to have been true in reality, the sensitivity analysis showed this was unlikely to have biased the results for this outcome.

The finding that medium-volume institutions have higher odds of mortality is surprising. Without further evaluation of potential causative factors, this is difficult to fully explain. This study has not assessed the relative contribution of the institution and surgeon to the volume-mortality relationship. It has also not attempted to define causal pathways of influence of the structural and process of care variables on mortality. Previous studies have demonstrated, for instance, that for colorectal resections medium-volume surgeons can achieve similar inpatient mortality results to high-volume surgeons when operating in medium- or high-volume hospitals but not low-volume hospitals [102]. Similarly, the contribution of individual institutional structural variables can be important. When low-volume institutions have as many residents/interns and registered nurses per 100 beds as high-volume institutions, overall mortality rates for paediatric heart surgery and heart transplants appeared equivalent [44].

From the summary of adjusted probabilities (Figure 3.6), medium-volume institutions do appear to have worse in-hospital and total mortality than high-volume institutions, which is consistent with previous studies. What is not clear is why low-volume institutions appear to have comparable outcomes to high-volume institutions. While it would be impossible to refute this finding, it may be artefactual. The division of volumes into tertiles does not assume a linear relation between volume and mortality, but rather allows for a non-linear one such as the middle volume having lower or higher mortality rates than the other two tertiles. It is possible that the volume cut-offs may not be optimal and that more complex functional forms (e.g. quadratics, splines) may perform better. The plotting of unit-level mortality rates against unit volume for the raw dataset did not reveal

any obvious relationships (Figure 3.7), and tertiles were chosen because of statistical power and because such an approach is commonly used. Large numbers of very low volume providers were not excluded which may have contaminated the low volume tertile if it contained patient-level records present only through coding errors.

Some of the structural and process of care factors used in the model may themselves be acting as proxies for volume. A number of the variables, such as staffing levels, critical care facilities and hospital capacity do increase across the institutional volume bands (Table 3.2). In extreme cases, this can lead to "collinearity", with multiple variables trying to describe the same quantity, but inspection of the standard errors showed no convincing evidence of this. However, the noticeable increase in the width of the confidence interval for the institution volume group with the addition of the structural and process variables suggests that the model is at the limit of acceptability regarding the number of variables included.

The use of an incremental modelling approach for volume-outcome research is in itself important for helping to decide whether volume should be defined at the institutional or surgeon level or both. Using the institution allows for the importance of overall teamwork on outcomes by factoring in institutional factors which cannot always be measured. It has been acknowledged that existing studies have rarely considered the relative effects of the unit of analysis and their possible interactions [177].

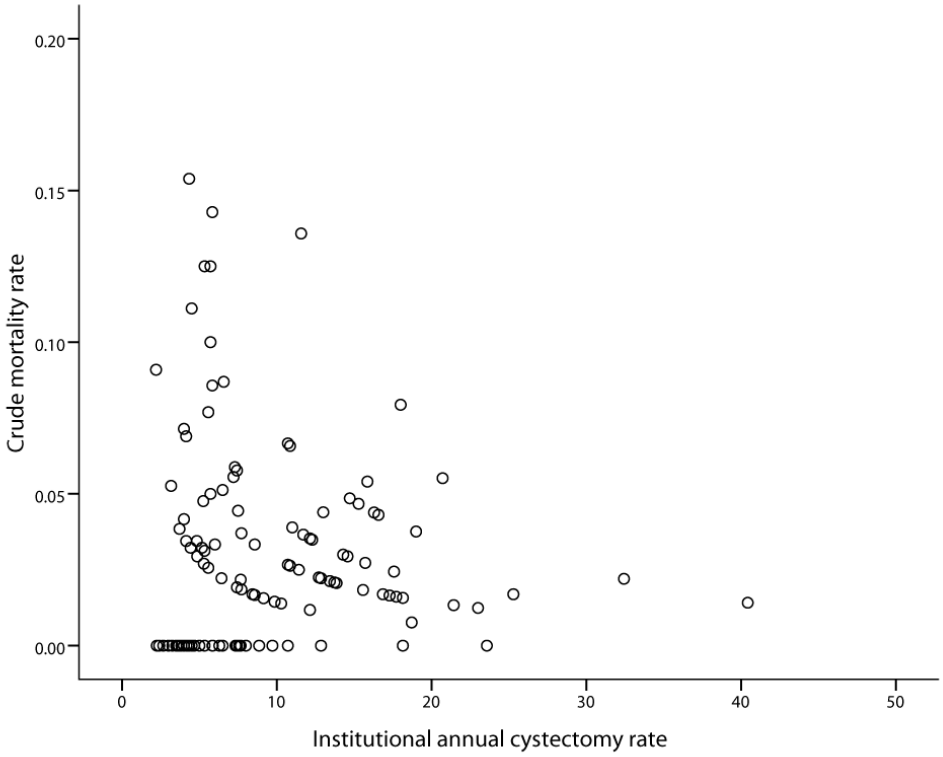


Figure 3.7 Scatter plot of Institutional crude mortality rate against institutional annual cystectomy rate (volume)

As defined earlier, each record in HES constitutes a finished consultant episode, which covers the continuous period during which a patient is under the care of the same consultant (identified by a consultant code defined by their General Medical Council code). This consultant code cannot in the strictest sense therefore be assumed to dictate who the lead operating surgeon was, and this is a limitation in definition of the surgeon level analysis. Radical cystectomy, however, has been and increasingly continues to be either a purely consultant performed operation, or at the very least an operation directly supervised by the consultant. Defining surgeon volume using the HES consultant code is therefore not unreasonable for this index procedure.

Comparisons with other studies

McCabe et al.[106] used only two volume categories in their HES-based analysis and defined an 'optimum' annual institutional caseload of 11 cases/year. When they shifted the case volume threshold to 16 cases/year, and therefore included more centres with optimum mortality outcomes in the group below the threshold (<16 cases/year), no significant difference in mortality rate between the groups was demonstrated.

The systematic review, using a random-effects analysis, described in study one found an 88% higher odds of mortality in low-volume than in higher-volume institutions for cystectomy. Only five studies could be included and their methodological quality was assessed as only modest at best. That the magnitude of difference was greater than in this current study can be partly explained by the limitation of meta-analysis when studies use different volume cut-off categories, adjustment methods, and amalgamate studies from different healthcare settings. US studies using the Health Care Utilization Project have demonstrated 3.2 times and 1.96 times the case-mix adjusted risk of dying postoperatively in low-volume institutions as compared with high-volume institutions (cystectomy rate 1-5/year versus ≥20/year [120] and cystectomy rate <1.5/year versus ≥2.75/year [105] respectively). The direction of effect is consistent with the results of this current

study when only considering the case-mix adjusted odds ratios, although their magnitude of effect was greater.

It could be suggested that there are only marginal differences in the annual caseloads between surgeon volume categories defined in this study. As compared with other studies however, the cut-offs are relatively far apart. Birkmeyer et al.[99] demonstrated that the adjusted odds ratio for operative death for patients with a low-volume surgeon (<2 cases/year) versus those with a high-volume surgeon (>3.5 cases/year) was 1.83 (CI 1.37 to 2.45). Conversely, Konety et al.[105] demonstrated no overall differences in in-hospital mortality across surgeon volume, but using only marginal volume cut-offs (low volume ≤ 1case/year and high volume > 1.5cases/year).

Few studies have investigated the effects of institutional structural variables and processes of care in the cystectomy treatment pathway. Some evidence from the US suggests that adjusting for processes of care before, during, and after radical cystectomy accounts for 23% of the volume-mortality effect [107]. Institutional structural variables can attenuate the volume-mortality relationship by up to 59% [120].

3.5 Conclusions

Interpretation of the institutional volume-mortality relationship reported in this study was only possible after adjustment for institutional structural and process of care variables. Adjusting for these confounders, in addition to case-mix, must be performed before considering the use of volume-outcome data to support centralisation of care to a few high-volume centres specialising in radical cystectomy for bladder cancer. This study has made no effort to determine causal relationships between individual structural and process of care measures and mortality following radical cystectomy. Mortality is only one of a number of outcomes that need to be examined when trying to improve the quality of care for patients and the volume-outcome relationship for outcomes other than mortality needs to be explored. Longer-term outcomes, including functional morbidity and

disease recurrence, may ultimately influence towards centralising care and so help inform future service design and reconfiguration.

3.6 Key points arising from study two

- A volume-outcome relationship for mortality following radical cystectomy exists at the institutional level, and to a lesser degree at the surgeon level

- Appropriate interpretation of the volume-mortality relationship was only possible after adjusting for institutional and surgeon volumes and structural and process of care confounders

- For radical cystectomy, centralising care to a few institutions should only be considered once the relationship between caseload volume and outcome (including outcomes other than mortality) has been properly adjusted for structural and process of care variables such as staffing levels of nurses and junior doctors, as well as case-mix.

Chapter 4

THE VOLUME-OUTCOME RELATIONSHIP FOR RADICAL CYSTECTOMY IN ENGLAND: THE IMPACT OF EXAMINING OUTCOMES OTHER THAN MORTALITY

The following chapter has been submitted as:

Mayer EK, Bottle A, Aylin P, Darzi AW, Athanasiou T, Vale JA. The volume-outcome relationship for radical cystectomy in England: An analysis of outcomes other than mortality. Submitted 2010.

4.1 Introduction

The previous study has presented the results of volume-mortality analysis for radical cystectomy in England. As discussed, mortality may not always be the most appropriate outcome measure to demonstrate differences in care between volume providers if the number of events is small. Mortality is a 'blunt' endpoint and can be the result of numerous other measurable outcomes in a patient's pathway of care.

Study two has demonstrated the importance of adjusting for structural and process of care confounders when examining the volume-mortality relationship. It can be postulated that some of these same confounders, such as staffing levels, may have an even greater effect on outcomes in the more immediate postoperative period such as complications and re-intervention rates.

The aim of this study is to expand the model developed in the previous study to include other outcome measures and so fully explore the volume-outcome relationship for radical cystectomy in England.

4.2 Methods

The data extraction, assignment of institutional and surgeon volume bands and model development is as previously described in study two, section 3.2.

4.2.1 Outcome measures

The additional outcome measures examined were postoperative re-intervention, postoperative complications and emergency readmission within 28 days for any reason to any hospital. Length of stay was also extracted from the dataset.

Two re-intervention variables were created by inspecting the list of procedures (coded using OPCS4) occurring within 30 days of the cystectomy: re-intervention within 14 days (i.e. between days 1 and 13 inclusive after the cystectomy) and re-

intervention within 30 days (i.e. between 1 and 29 days inclusive), even if performed following transfer or discharge from the institution at which the cystectomy had been performed (Appendix B). The secondary diagnosis fields were inspected for a list of potential postoperative complications as listed in Appendix B. This provided the outcome measure "complications". We also subdivided "complications" into procedure-specific (those deemed to be directly attributable to the cystectomy) and non-procedure specific (those that could occur after any type of surgery). Similarly, re-interventions were categorised into those related to managing a complication arising directly from the operation itself (operation-specific re-interventions) and those needed for other reasons (non-operation specific re-interventions). Categorisation of complications and re-interventions was performed independently by EM and JV without having knowledge of the volume data to avoid causing potential study bias. Any disagreement was resolved by consensus opinion. Complications and re-intervention within 14 days and 30 days were combined to give a further outcome measure ("any event").

The median length of stay was noted for each year but not included in the volume-outcome modelling. Length of stay was calculated in days (strictly speaking, in-patient nights) using the HES entry fields date of discharge minus date of admission.

4.3 Results

Re-intervention rates at 14 & 30 days and the outcome any event showed non-linear trends across the years (Table 4.1). There was also a significant, but small, decrease in average LOS (0.09 day each year).

Both medium and high-volume institutions have a higher unadjusted rate of re-interventions and any event. High-volume institutions have a greater rate of complications than both medium and low-volume institutions, Figure 4.1. Similarly at the surgeon level, medium and high-volume surgeons have a higher

unadjusted rate of re-interventions within 30 days, complications and any event, Figure 4.2. Absolute numbers of events are presented in Table 4.2.

For both institutions and surgeons there was little difference in the crude readmission rates across volume tertiles.

Statistically significant associations were seen between re-intervention and some of the structural confounders, Table 4.3.

4.3.1 Volume-outcome relationship

4.3.1.1 Re-intervention

Medium-volume institutions had statistically significant greater odds of re-intervention both within 14 and 30 days for all models. For model four, there was a 63% and 52% greater odds for re-intervention within 14 and 30 days respectively (OR 1.63 (1.15 to 2.32, p=0.01) and OR 1.52 (1.13 to 2.04, p=0.01)) (Figure 4.3). There were similar findings for high-volume institutions using models one, two and three, but for model four, the higher odds of re-intervention within 14 and 30 days decreased and became non-significant at the 5% level (OR 1.33 (0.86 to 2.04, p=0.20) and OR 1.26 (0.87 to 1.81, p=0.22) respectively) (Figure 4.3). After categorising re-interventions, medium-volume institutions continued to show a statistically significant greater odds of non-operation specific re-intervention (OR 1.46 (1.13 to 1.89, p<0.01)) and a trend towards higher odds of operation-specific re-intervention (OR 1.25 (0.96 to 1.63, p=0.10)) for model four (Figure 4.4).

Financial year	Re-interventions within 14 days n (%)	Re-interventions within 30 days n (%)	Complications n (%)	Any event n (%)	Readmission rate‡ n (%)	Median length of stay (days)
2000/1	109 (9.6)	164 (14.5)	210 (18.6)	320 (28.3)	170 (13.7)	17
2001/2	114 (9.7)	179 (15.2)	215 (18.2)	325 (27.6)	141 (11.4)	17
2002/3	128 (10.3)	180 (14.5)	253 (20.4)	358 (28.8)	174 (14.1)	17
2003/4	163 (13.0)	235 (18.8)	231 (18.5)	368 (29.4)	182 (14.7)	17
2004/5	155 (12.6)	239 (19.4)	249 (20.2)	388 (31.4)	194 (15.7)	17
2005/6	151 (12.0)	227 (18.0)	259 (20.5)	408 (32.3)	179 (14.5)	16
2006/7	132 (10.2)	191 (14.7)	264 (20.4)	372 (28.7)	197 (15.9)	16
p value for linear trend across years	0.09	0.06	0.11	0.06	0.24	<0.01*

Table 4.1 Summary of outcome measures by year for radical cystectomy (unadjusted data)

‡ Cases with in-hospital mortality excluded before determining readmission rates.

* Indicates significant, but small, decrease in average LOS each year (0.09 day) across analysed period after log transformation

	Re-interventions within 14 days (n)	Re-interventions within 30 days (n)	Complications (n)	Any event (n)	Readmission rate‡ (n)
Institutional volume band					
Low	274	408	559	801	430
Medium	336	505	487	818	422
High	342	502	635	920	385
Surgeon volume band					
Low	339	467	523	816	452
Medium	283	407	484	743	340
High	300	489	608	885	409

Table 4.2 Observed number of events for each outcome variable stratified by institutional/surgeon volume bands

‡ Cases with in-hospital mortality excluded before determining readmission rates.

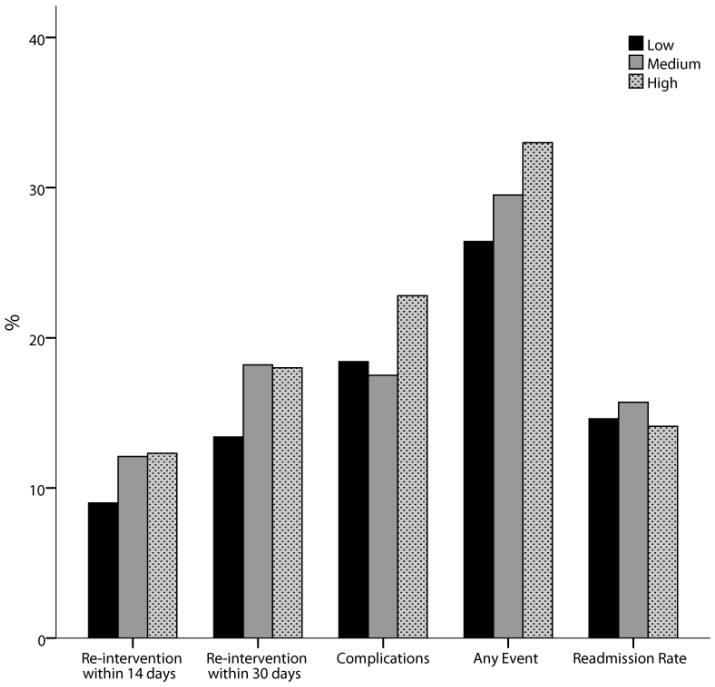


Figure 4.1 Crude outcome rates (%) stratified by institutional volume tertiles

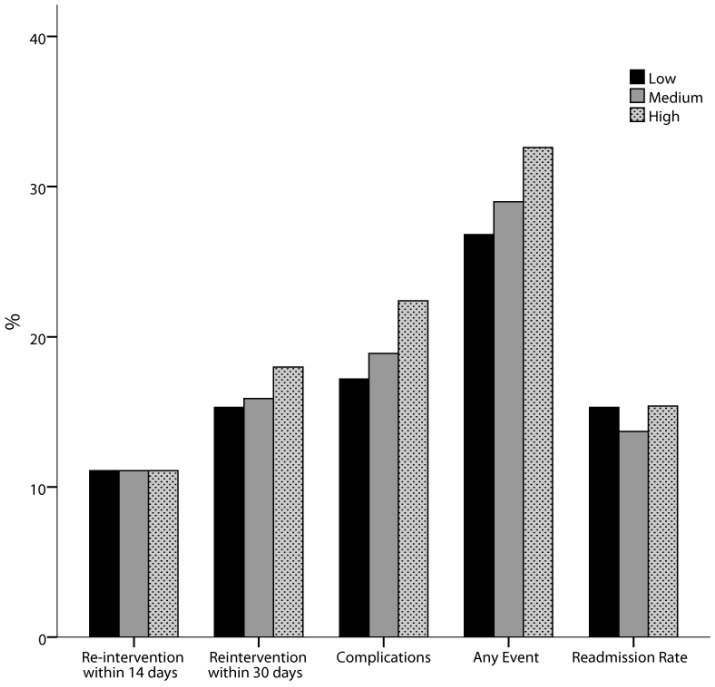


Figure 4.2 Crude outcome rates (%) stratified by surgeon volume tertiles

	Re-intervention within 14 days	Re-intervention within 30 days	Complications	Any event	Readmission
Trust teaching status	-	-	-	↑	-
Average occupied acute bed rate	↓	↓	-	↓	-
Ratio of number of nurses to occupied bed count	↓	-	-	-	-
Ratio of number of critical care beds to total beds	-	-	-	-	-
Ratio of registrars (all specialities) to occupied bed number	-	-	-	-	-
Ratio of urology registrars to total urology episodes	↑*	↑*	-	-	-
Operation waiting time	-	-	-	-	-
Readmission rate following operation cancellation	-	-	-	-	-
Total beds, all sectors, available	-	-	-	-	-
Total acute beds available	-	-	-	-	-
Total urology episodes	-	-	-	-	-

Table 4.3 Table summarising the statistically significant ($p<0.05$) associations between structural and process of care variables and outcome measures. Direction of association indicated by up (positive)/down (negative) arrow. (* Approached significance, $p=0.08$)

In contrast, high-volume surgeons had a statistically significant lower odds of re-intervention within 14 days for models three and four (OR 0.73 (0.55 to 0.96, p=0.02)) (OR 0.68 (0.51 to 0.91, p=0.01)) which was not apparent using models one and two (Figure 4.5). For re-intervention within 30 days, a statistically significant higher odds for models one and two (OR 1.22 (1.06 to 1.40, p<0.01) and OR 1.17 (1.01 to 1.34, p=0.04)) became non-significant at the 5% level for models three and four (Figure 4.5). There was no significant relationship between surgeon volume and operation-specific and non-operation specific re-interventions (Figure 4.6).

4.3.1.2 Complications

Both high-volume institutions and high-volume surgeons had higher odds of complications for models one and two. Model two gave an OR of 1.22 (1.06 to 1.39, p<0.01) and an OR of 1.34 (1.17 to 1.54, p<0.01) for institutions and surgeons respectively. These results became non-significant at the 5% level for models three and four. No significant differences were seen for medium-volume institutions or surgeons (Figure 4.7 & Figure 4.8). There was no statistically significant relationship between institutional or surgeon volume and procedure-specific and non-procedure specific complications. (Figure 4.7 & Figure 4.8)

4.3.1.3 Readmission

No statistically significant difference in odds of readmission within 28 days was seen across institutional or surgeon volume bands although for medium-volume surgeons a lower odds of readmission approached significance for model three (OR 0.86 (0.72-1.03, p=0.09)). (Figure 4.9 & Figure 4.10)

A summary of the adjusted probabilities for each outcome across both the institutional and surgeon volume bands is displayed in Figure 4.11. Adjusted probabilities were calculated using model four.

Table 4.4, Table 4.5 and Table 4.6 display the crude odds ratios and adjusted odds ratios for models two to four for each outcome assessed across both institution and surgeon volume tertiles.

4.3.2 Interaction effect

Only re-intervention within 14 days had a statistically significant relationship for both institutional and surgeon volume. This therefore allowed for testing of interaction effect between the two volume variables but this was not statistically significant (p=0.67).

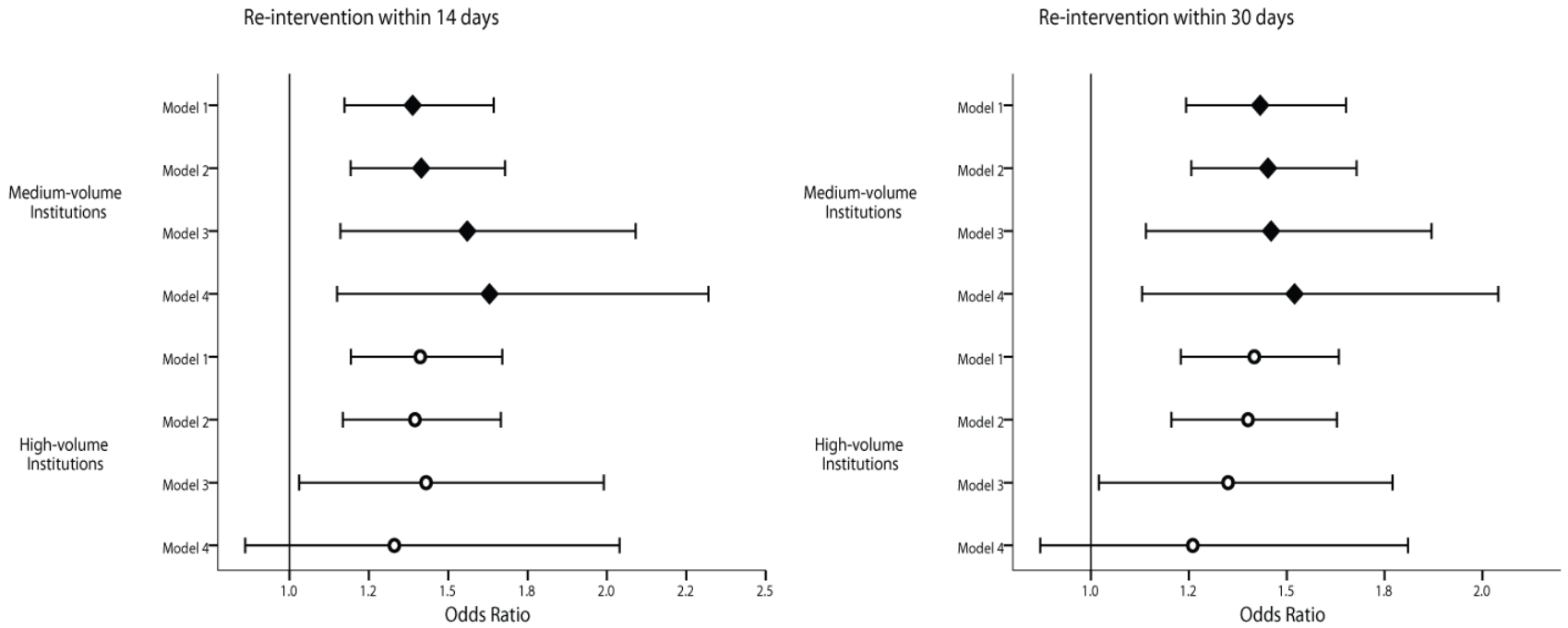


Figure 4.3 Summary of odds ratio and confidence intervals for institutional volume-re-intervention relationships across all four models

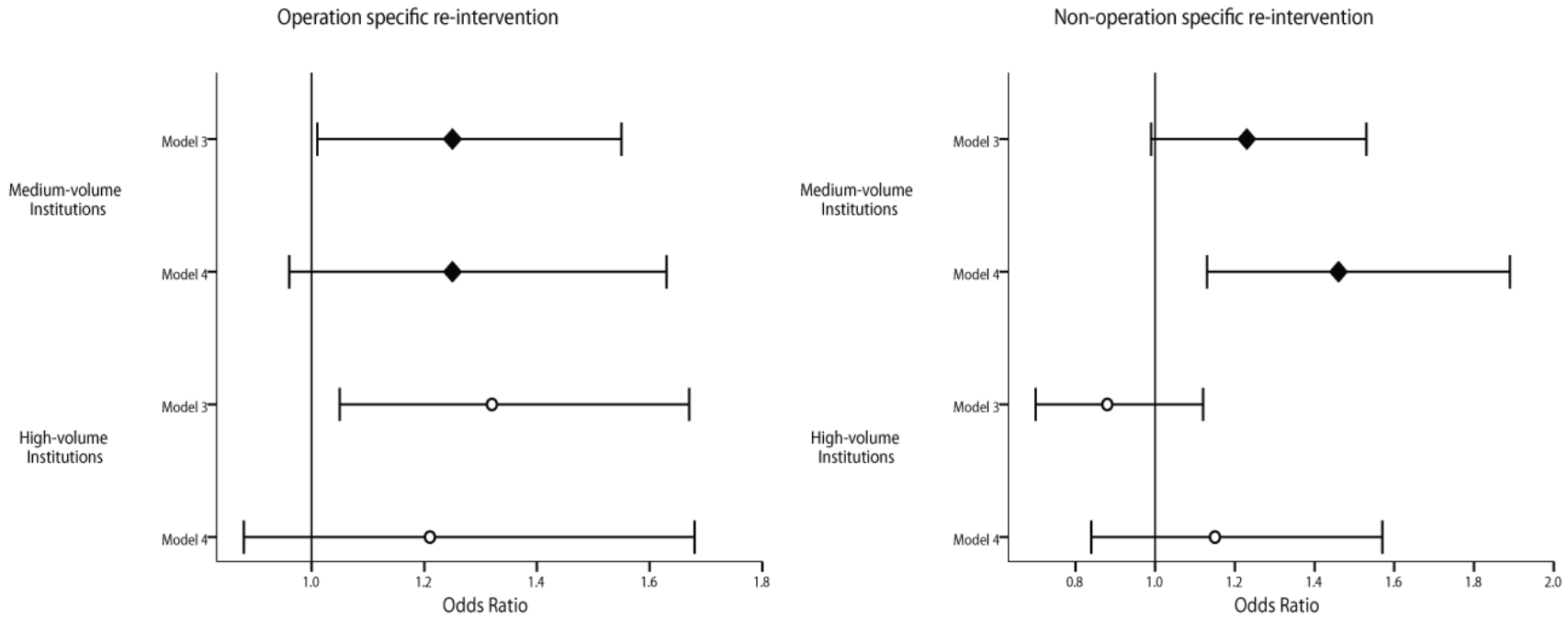


Figure 4.4 Summary of odds ratio and confidence intervals for institutional volume-re-intervention relationships across all four models, stratified by operation-specific and non-operation specific re-interventions

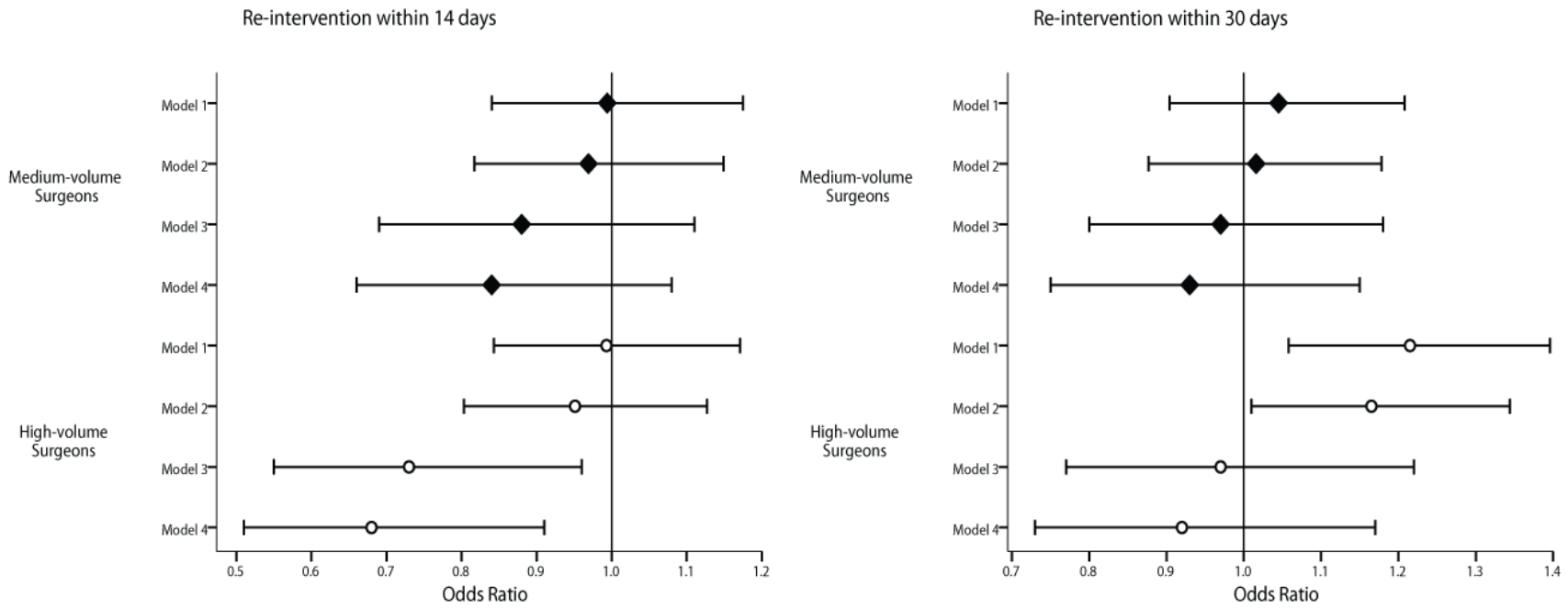


Figure 4.5 Summary of odds ratio and confidence intervals for surgeon volume-re-intervention relationships across all four models

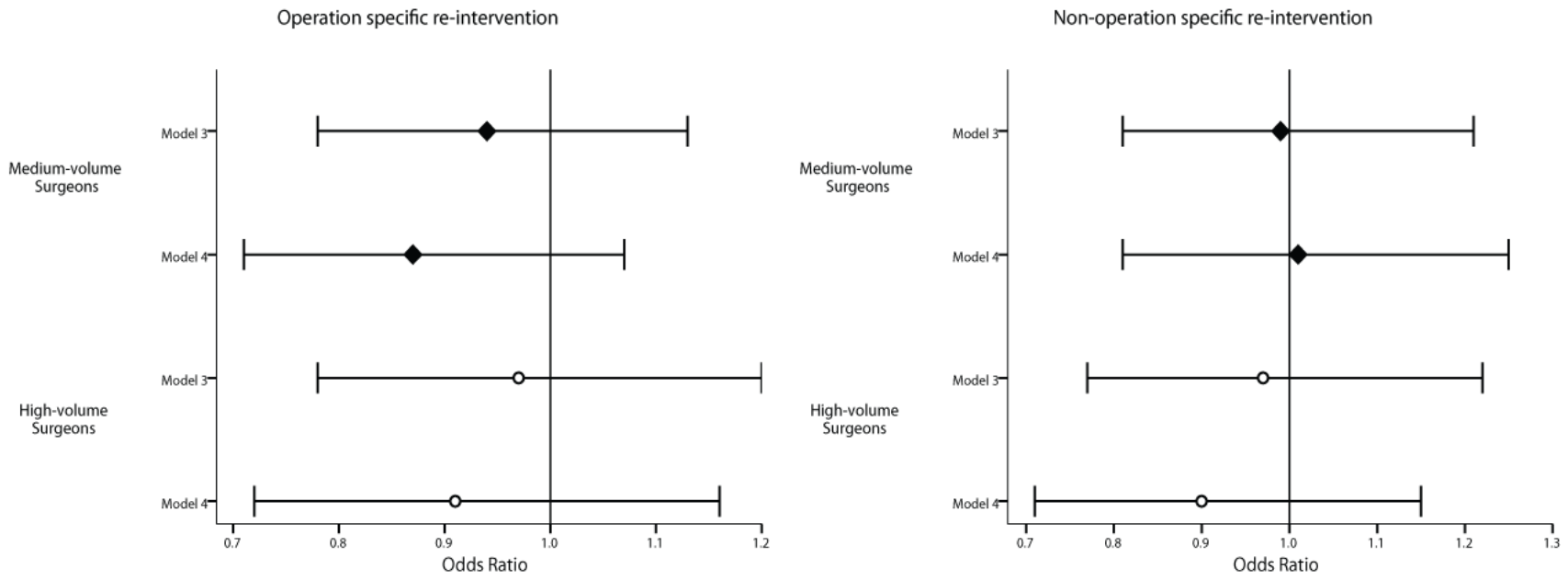


Figure 4.6 Summary of odds ratio and confidence intervals for surgeon volume-re-intervention relationships across all four models, stratified by operation-specific and non-operation specific re-interventions

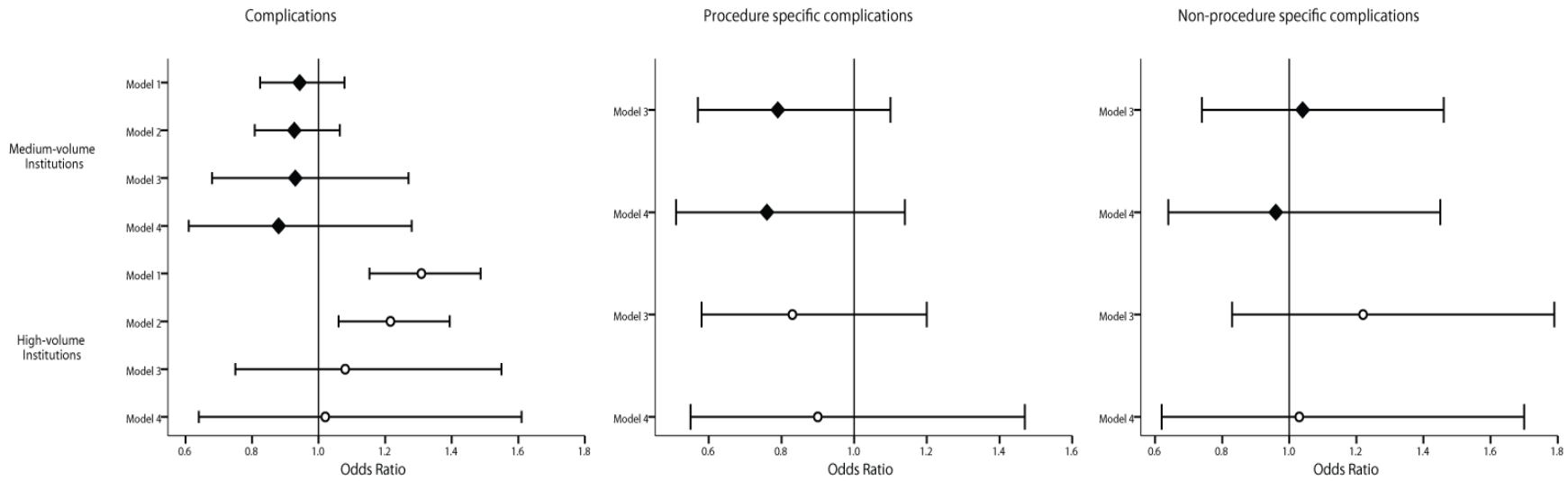


Figure 4.7 Summary of odds ratio and confidence intervals for institutional volume-complication relationships across all four models

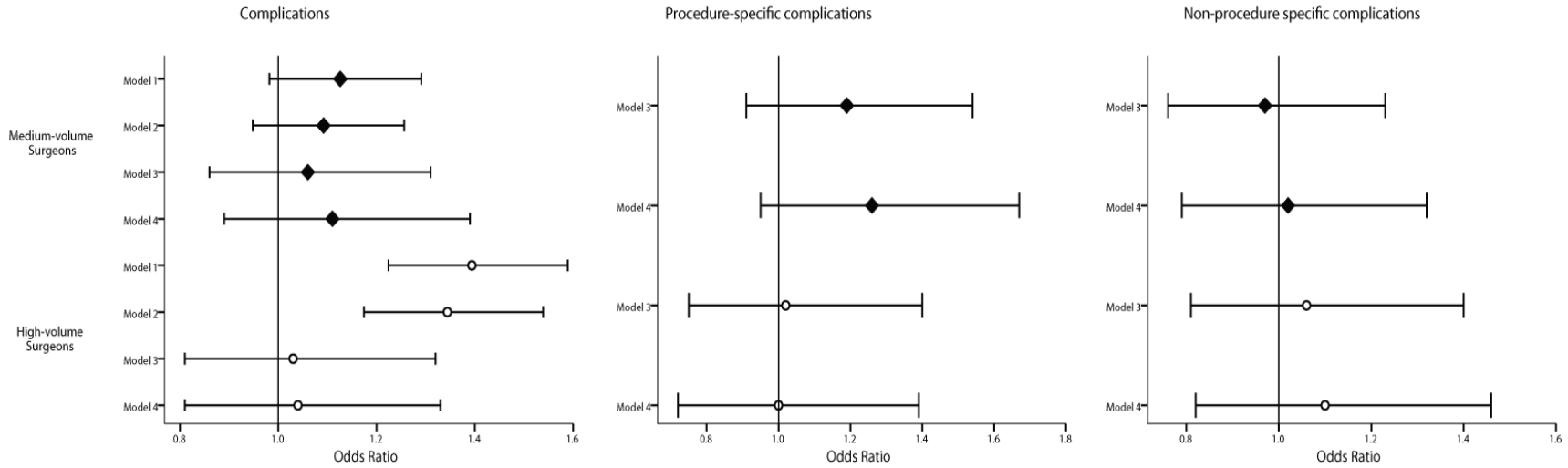


Figure 4.8 Summary of odds ratio and confidence intervals for surgeon volume-complication relationships across all four models

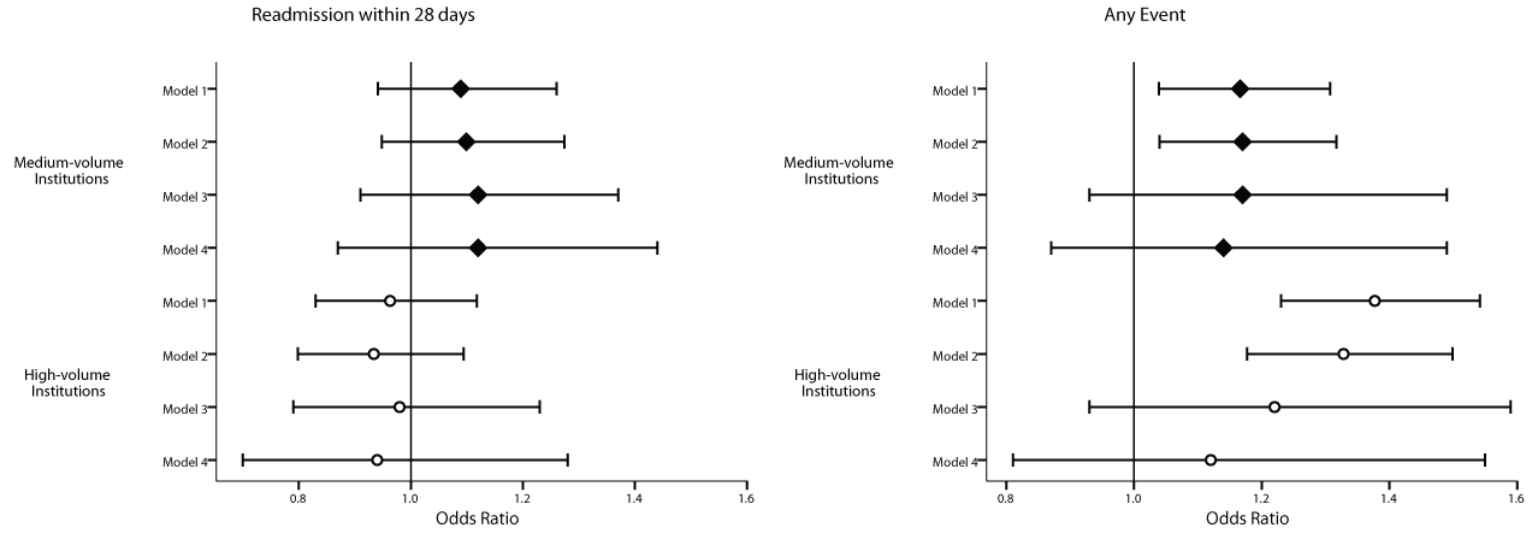


Figure 4.9 Summary of odds ratio and confidence intervals for institutional volume-readmission and any event relationships across all four models

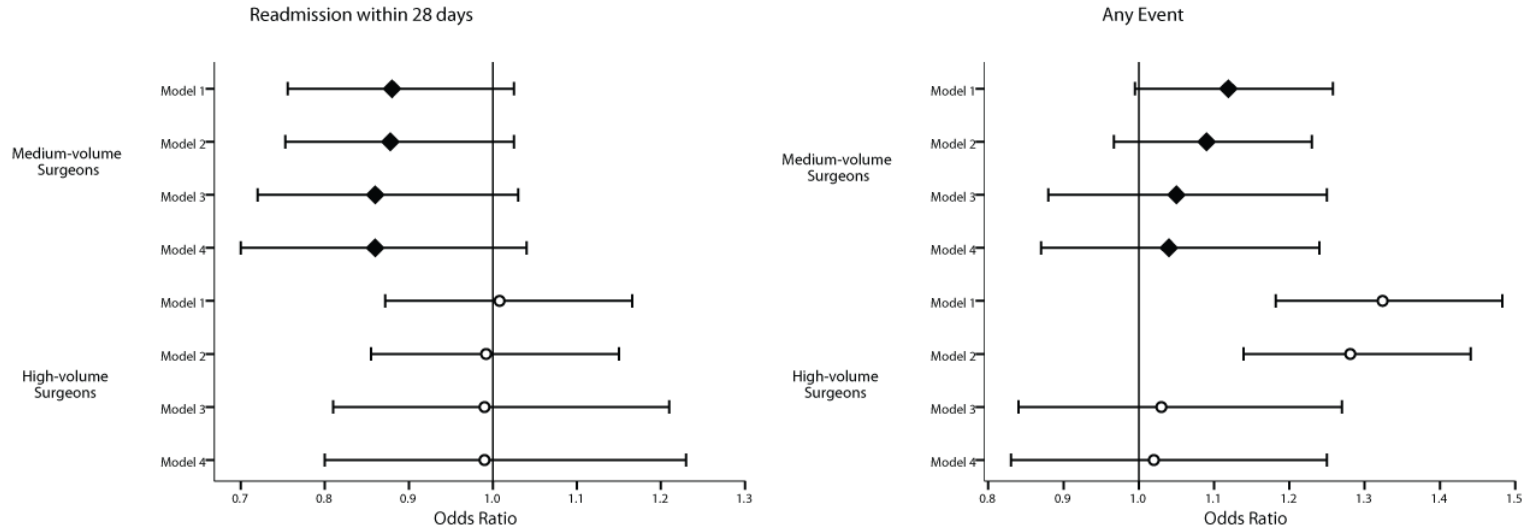


Figure 4.10 Summary of odds ratio and confidence intervals for surgeon volume-readmission and any event relationships across all four models

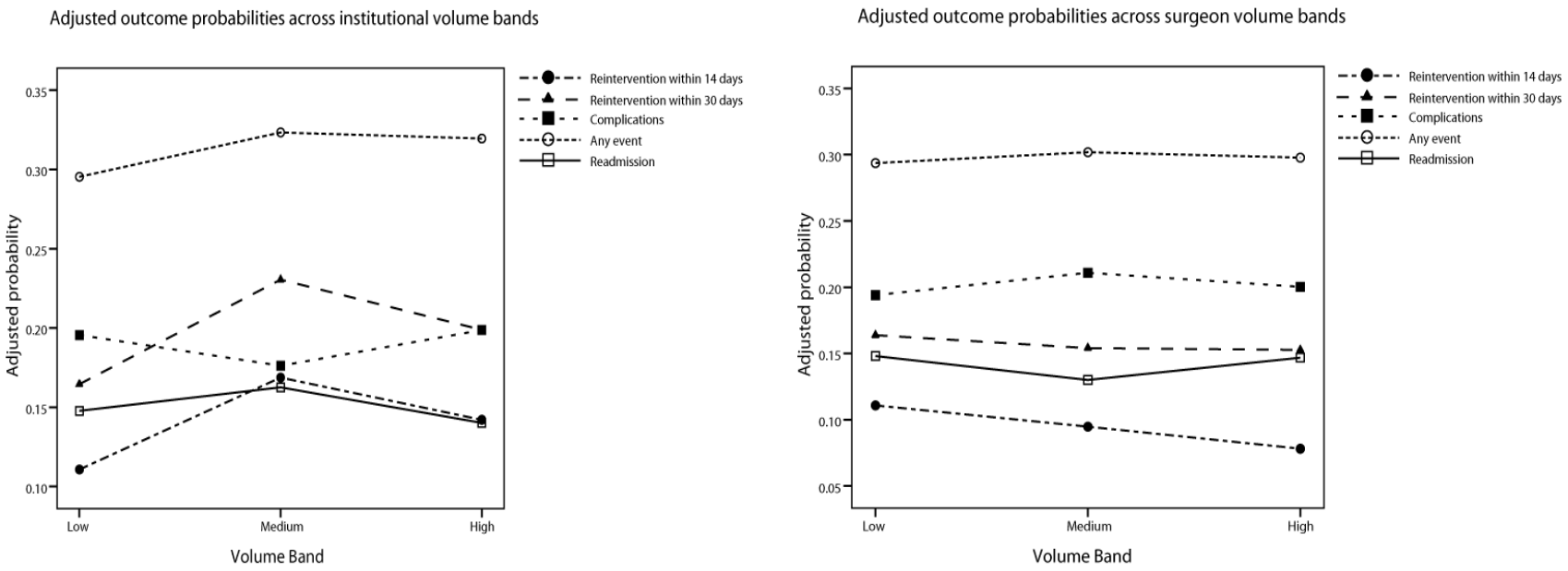


Figure 4.11 Summary of adjusted probabilities (derived from model 4) for each outcome variable across institutional and surgeon volume bands

Outcome Measure										
	Re-intervention within 14 days		Re-intervention within 30 days		Complications		Any event		Readmission	
	Crude OR	Case-mix adjusted OR (CI, p value)	Crude OR	Case-mix adjusted OR (CI, p value)	Crude OR	Case-mix adjusted OR (CI, p value)	Crude OR	Case-mix adjusted OR (CI, p value)	Crude OR	Case-mix adjusted OR (CI, p value)
Institution										
Low volume (ref)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Medium volume	1.39	1.42 (1.19-1.68, <0.01)	1.43	1.45 (1.26-1.68, <0.01)	0.94	0.93 (0.81-1.06, 0.28)	1.17	1.17 (1.04-1.32, <0.01)	1.09	1.10 (0.95-1.27, 0.21)
High volume	1.41	1.40 (1.17-1.67, <0.01)	1.42	1.40 (1.21-1.63, <0.01)	1.31	1.22 (1.06-1.39, <0.01)	1.38	1.33 (1.18-1.50, <0.01	0.96	0.93 (0.80-1.09, 0.40)
Surgeon										
Low volume (ref)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Medium volume	0.99	0.97 (0.82-1.15, 0.72)	1.05	1.02 (0.88-1.18, 0.83)	1.13	1.09 (0.95-1.26, 0.22)	1.12	1.09 (0.97-1.23, 0.16)	0.88	0.88 (0.75-1.03, 0.10)
High volume	0.99	0.95 (0.80-1.13, 0.56)	1.22	1.17 (1.01-1.34, 0.04)	1.39	1.34 (1.17-1.54, <0.01)	1.32	1.28 (1.14-1.44, <0.01)	1.01	0.99 (0.86-1.15, 0.91)

Table 4.4 Differences in outcome measures across both institution and surgeon volume categories, displayed as crude (model one) and case-mix adjusted (model two) odds ratios (OR)

	Outcome Measure									
	Re-intervention within 14 days		Re-intervention within 30 days		Complications		Any event		Readmission	
	MLM OR	Structural variables OR (CI, p value)	MLM OR	Structural variables OR (CI, p value)	MLM OR	Structural variables OR (CI, p value)	MLM OR	Structural variables OR (CI, p value)	MLM OR	Structural variables OR (CI, p value)
Institution										
Low volume (ref)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Medium volume	1.56 (1.16-2.09, <0.01)	1.63 (1.15-2.32, 0.01)	1.46 (1.14-1.87, <0.01)	1.52 (1.13-2.04, 0.01)	0.93 (0.68-1.27, 0.64)	0.88 (0.61-1.28, 0.52)	1.17 (0.93-1.49, 0.19)	1.14 (0.87-1.49, 0.33)	1.12 (0.91-1.37, 0.28)	1.12 (0.87-1.44, 0.37)
High volume	1.43 (1.03-1.99, 0.03)	1.33 (0.86-2.04, 0.20)	1.35 (1.02-1.77, 0.03)	1.26 (0.87-1.81, 0.22)	1.08 (0.75-1.55, 0.70)	1.02 (0.64-1.61, 0.95)	1.22 (0.93-1.59, 0.15)	1.12 (0.81-1.55, 0.50	0.98 (0.79-1.23, 0.88)	0.94 (0.70-1.28, 0.71)
Surgeon										
Low volume (ref)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Medium volume	0.88 (0.69-1.11, 0.27)	0.84 (0.66-1.08, 0.18)	0.97 (0.80-1.18, 0.78)	0.93 (0.75-1.15, 0.50)	1.06 (0.86-1.31, 0.58)	1.11 (0.89-1.39, 0.33)	1.05 (0.88-1.25, 0.57)	1.04 (0.87-1.24, 0.69)	0.86 (0.72-1.03, 0.09)	0.86 (0.70-1.04, 0.12)
High volume	0.73 (0.55-0.96, 0.02)	0.68 (0.51-0.91, 0.01)	0.97 (0.77-1.22, 0.78)	0.92 (0.73-1.17, 0.49)	1.03 (0.81-1.32, 0.81)	1.04 (0.81-1.33, 0.77)	1.03 (0.84-1.27, 0.75)	1.02 (0.83-1.25, 0.87)	0.99 (0.81-1.21, 0.95)	0.99 (0.80-1.23, 0.95)

Table 4.5 Differences in outcome measures across both institution and surgeon volume categories, displayed as odds ratios (OR) following multi-level modelling (model three) and adjustment for structural variables (model four)

	Outcome Measure													
	All complications		Procedure-specific complications		Non-procedure specific complications		All re-intervention within 14 days		All re-intervention within 30 days		Operation-specific re-intervention within 30 days		Non-operation specific re-intervention within 30 days	
	MLM OR	Structural variables OR (CI, p value)	MLM OR	Structural variables OR (CI, p value)	MLM OR	Structural variables OR (CI, p value)	MLM OR	Structural variables OR (CI, p value)	MLM OR	Structural variables OR (CI, p value)	MLM OR	Structural variables OR (CI, p value)	MLM OR	Structural variables OR (CI, p value)
Institution														
Low volume (ref)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Medium volume	0.93 (0.68-1.27 , 0.64)	0.88 (0.61-1.28 , 0.52)	0.79 (0.57-1.10 , 0.17)	0.76 (0.51-1.14 , 0.19)	1.04 (0.74-1.46 , 0.82)	0.96 (0.64-1.45 , 0.86)	1.56 (1.16-2.09 , <0.01)	1.63 (1.15-2.32 , 0.01)	1.46 (1.14-1.87 , <0.01)	1.52 (1.13-2.04 , 0.01)	1.25 (1.01-1.55 , 0.04)	1.25 (0.96-1.63 , 0.10)	1.23 (0.99-1.53 , 0.06)	1.46 (1.13-1.89 , <0.01)
High volume	1.08 (0.75-1.55 , 0.70)	1.02 (0.64-1.61 , 0.95)	0.83 (0.58-1.20 , 0.33)	0.90 (0.55-1.47 , 0.69)	1.22 (0.83-1.79 , 0.32)	1.03 (0.62-1.70 , 0.92)	1.43 (1.03-1.99 , 0.03)	1.33 (0.86-2.04 , 0.20)	1.35 (1.02-1.77 , 0.03)	1.26 (0.87-1.81 , 0.22)	1.32 (1.05-1.67 , 0.02)	1.21 (0.88-1.68 , 0.24)	0.88 (0.70-1.12 , 0.30)	1.15 (0.84-1.57 , 0.39)
Surgeon														
Low volume (ref)	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Medium volume	1.06 (0.86-1.31 , 0.58)	1.11 (0.89-1.39 , 0.33)	1.19 (0.91-1.54 , 0.20)	1.26 (0.95-1.67 , 0.12)	0.97 (0.76-1.23 , 0.79)	1.02 (0.79-1.32 , 0.89)	0.88 (0.69-1.11 , 0.27)	0.84 (0.66-1.08 , 0.18)	0.97 (0.80-1.18 , 0.78)	0.93 (0.75-1.15 , 0.50)	0.94 (0.78-1.13 , 0.53)	0.87 (0.71-1.07 , 0.20)	0.99 (0.81-1.21 , 0.93)	1.01 (0.81-1.25 , 0.95)
High volume	1.03 (0.81-1.32 , 0.81)	1.04 (0.81-1.33 , 0.77)	1.02 (0.75-1.40 , 0.88)	1.00 (0.72-1.39 , 1.00)	1.06 (0.81-1.40 , 0.66)	1.10 (0.82-1.46 , 0.53)	0.73 (0.55-0.96 , 0.02)	0.68 (0.51-0.91 , 0.01)	0.97 (0.77-1.22 , 0.78)	0.92 (0.73-1.17 , 0.49)	0.97 (0.78-1.20 , 0.75)	0.91 (0.72-1.16 , 0.45)	0.97 (0.77-1.22 , 0.80)	0.90 (0.71-1.15 , 0.39)

Table 4.6 Differences in operation and non-operation specific complication and re-intervention rates across both institution and surgeon volume categories, displayed as odds ratios (OR) following multi-level modelling (model three) and adjustment for structural variables (model four)

4.4 Discussion

4.4.1 Principal findings

The odds of re-intervention within 14 and 30 days of operation, for patients undergoing radical cystectomy in medium-volume institutions when compared with low-volume institutions, were found to be 63% and 52% higher respectively. This did not translate to a correspondingly higher readmission rate; this could be due to the fact that re-intervention may have itself reduced readmission rates, the complication having become manifest before discharge. Patients had 46% greater odds of non-operation specific re-interventions using the final complex model. Adjustment for the same variables had no effect on operation-specific re-intervention rates. This suggests that the structural and process of care factors had an important protective role at the medium-volume institutional level on non-operation specific re-interventions.

High-volume surgeons (but not medium volume) were associated with a reduced odds (32%) of early re-intervention (within 14 days) compared with low-volume surgeons. Operation-specific re-interventions are generally needed within a shorter timeframe following the index procedure and this could explain why the lower odds of overall re-intervention seen for high-volume surgeons within 14 days is no longer present at 30 days due to dilution by non-operation specific re-interventions, which occur later. Categorisation of re-interventions within a 30-day period into operation specific and non-operation specific re-interventions, however, did not reveal any noticeable relationships with surgeon volume.

There was no statistically significant relationship between volume and the risk of complications at either the institutional or surgeon level. Recording of complications will vary between institutions and the finding may be artefactual. An argument against this is that a statistically significant higher rate of complications was seen for both high-volume surgeons and institutions after case-mix adjustment, and this only became non-significant after adjusting for

clustering. The implication may be that complications can be directly influenced by both the surgeon and institution (some more surgically related such as anastomotic leaks, wound haematomas and some more medically related to post-operative care, such as deep vein thrombosis and chest infection). The results did not reveal any definitive patterns after categorisation of complications into procedure-specific and non-procedure specific.

As seen for mortality, there was relatively little difference in the crude OR and case-mix adjusted OR of each outcome at both the institutional and surgeon level. This provides further supporting evidence that case-mix adjustment may not be necessary for high level performance analysis, such as used in this study.

4.4.2 Strengths and weaknesses of the study

As highlighted in the previous study, it was not possible to perform sub-analysis of outcomes on the basis of whether an ileal conduit diversion or continent diversion had been performed. This limitation was probably less important when examining mortality, but differentiating between continent and incontinent diversion will have more of an impact on complication and re-intervention rates and indeed even possibly readmission rates.

Mortality is a definitive endpoint and avoids some of the concerns of coding errors in administrative databases for outcomes such as re-interventions and complications, which require a greater degree of subjectivity when interpreting clinical notes for the purpose of coding. A coding audit carried out by the Audit Commission identified that the vast majority of errors of diagnostic and procedure codes were caused by coders rather than non-coders because of suboptimal standards of source documentation [178]. Procedure (re-intervention) coding was demonstrated to be more accurate than diagnosis (complication) coding [174]. The Payment by Results national benchmarker has been introduced to target areas for clinical coding audits as part of an assurance framework to further improve data quality at the local level [178]. The average clinical coding error rate of 16.5% across primary and secondary diagnosis and procedure coding in 2007/08 improved in 2008/09 to 12.8% with a reduction also seen in the interquartile ranges for both

procedure and diagnosis coding across acute NHS Trusts in England [174]. If the recording of complications/re-interventions in HES is very dependent on volume, there is the possibility that differences seen across volume bands could be at least partly due to artefact, but it is impossible to prove this one way or the other with the available data.

This study has adjusted for both institutional and surgeon volume simultaneously to begin to assess the influence of each across a number of outcomes. It can be difficult to distinguish the independent effects of the institution/surgeon on outcomes because high-volume surgeons operate at high-volume institutions and vice versa. This study demonstrated a non-significant interaction effect for institution/surgeon volume on re-intervention rates within 14 days, although statistical power to detect an interaction effect was limited. Of those operations performed in a high-volume institution, 16%, 32% and 50% were performed by low, medium and high-volume surgeons respectively. The 2% of operations that did not have an attributable surgeon volume arose because these surgeons were excluded from the surgeon level analysis for performing, on average, ≤2 cystectomies per year and having fewer than three years of data.

The variable 're-intervention' rather than return to theatre was used as an outcome measure because contemporary approaches to the management of post-operative complications are more conservative and often include radiological techniques. As a result, transfusion episodes were incorporated as a re-intervention. However, the method by which the re-intervention was generated means that only transfusions given in the days following cystectomy, and not therefore required intra-operatively, were included and as such reflect a 'true' re-intervention. It was not possible to determine the number of units transfused for each patient.

The outcome measure 'any event' was created to explore the impact of assessing the volume-outcome relationship using a composite outcome measure as opposed to more granular measures, i.e. re-intervention and complications separately. In the final complex model, neither at the institutional level or surgeon level was

there a volume-any event relationship when one existed for one of its component elements, re-intervention.

The previous study discussed the issue of overfitting of the model (collinearity), such that if the variables used in our final complex model were highly correlated with volume this would have the effect of widening the confidence intervals and reducing the statistical significance of any differences seen; there would not be any biasing of the results. The noticeable increase in the width of the confidence interval seen for the institution volume group with the addition of the structural and process variables when examining mortality was not convincingly repeated for the other outcomes. There is therefore no convincing evidence of overfitting of the model. Including 11 outcomes in the analysis affords more opportunity for false positives than if only a single outcome measure was used, but the finding that five of these outcomes had significant (at the conventional 5% level and in fact with $p<0.02$) volume relations is more than would be expected by chance.

4.4.3 Comparisons with other studies

The volume-complications relationship following cystectomy has been infrequently considered. Elting et al.[108] demonstrated that the risk of developing a complication was 47% lower if the patient was treated in a high-volume hospital (>10 cases/year). In contrast Konety et al.[118] only showed a statistically significant lower risk in high-volume hospitals (> 3 cases/year) for primary complications (directly attributable to cystectomy), OR 0.81 (0.66–0.99); no differences were seen for their list of secondary complications, OR 0.96 (0.75–1.24). This current study did not corroborate the finding of a lower risk of primary complications, but this is not surprising as Konety et al. defined high-volume as > 3 cases per year as compared to > 16 cases per year (low-volume > 2 < 10 cases per year) used in this study. Although a higher case-mix adjusted risk of complications for both high-volume institutions and surgeons was demonstrated, the loss of this relationship after adjusting for clustering and categorisation of complications indicates the importance of this analysis approach. Another study reported no significant association between hospital volume and complications [179]. It has also been shown that there can be volume-complication relationships for specific complications,

where none exists for composite complications measures [121]. Absolute differences in the risk of complications for high-volume hospitals in this study as compared with the others described could be explained by the different volume cut-offs used. No study has specifically evaluated the re-intervention rates following cystectomy, although inevitably some of the complications reported will require surgical intervention. Re-intervention, like mortality, forms 'hard' measurable endpoints and can be associated with significant morbidity for the patient.

There was no relationship between volume and readmission rates. Measuring readmission rates has been shown to be useful as a performance measure when assessing a relatively homogenous patient population with respect to case-mix, and the severity of the condition of interest displays little variance, such as in the day surgery setting [98].

The only other urological study to investigate complications and readmission rates using the HES database assessed these outcomes in relation to radical prostatectomy [121]. They reported that there was an inverse relationship between in-hospital complications within 30 days and hospital volume. This association however disappeared after case-mix adjustment. They did confirm that readmission rates within a year were lower in higher volume institutions but again after case-mix adjustment, there was a non-statistical difference across institutional volume of an order similar to our study. Other studies have shown no relationship between volume and 30-day readmission rate following cystectomy [97]. There appears, therefore, to be little additional benefit in measuring readmission beyond 30 days.

4.5 Conclusions

This study, in combination with the evidence from study two, implies that when using currently available administrative data for risk-adjustment, measures of mortality, complications and re-intervention rates can be used as outcome measures to discern differences across institutional or surgeon volume providers when analysed independently of each other. However, when the institutional and

surgeon volume are co-examined and adjustment for structural and process of care variables performed, volume-outcome relationships become less clear, although mortality and re-intervention rates continue to demonstrate differences. Institutions and surgeons are not mutually exclusive and treatment outcomes are dependent on the interaction between them. Mortality is a definitive endpoint and avoids some of the concerns of coding errors in administrative databases for outcomes such as re-interventions and complications, which require a greater degree of subjectivity when interpreting clinical notes for the purpose of coding.

While this study has suggested correlation between individual structural and process of care measures and outcomes of radical cystectomy, it has made no effort to determine causal relationships and this will need to be the subject of further research to help inform future service design and reconfiguration. The influence of the surgeon, on-call surgical and nursing cover and availability of ancillary services such as interventional radiology can all influence the detection and management of complications that may lead to re-intervention and/or death; all of these factors need to be measured and considered together.

4.6 Key points arising from study three

- A volume-outcome relationship for re-intervention after radical cystectomy existed at the institutional and surgeon level.

- Re-intervention rates can be used as outcome measures to discern differences across institutional or surgeon volume providers when the institutional and surgeon volume are co-examined and adjustment for structural and process of care confounders performed.

- The surgeon volume-re-intervention relationship only became apparent after adjusting for the influence of the institution and provided the strongest evidence for the influence that the surgeon/institution interaction can have on patient outcomes.

- The absence of a volume-outcome relationship for the outcome measure 'any event' when one was present for re-intervention, demonstrated the importance of not using composite outcome measures.

- Further understanding of the interplay between the volume-complication, re-intervention and mortality relationships will help to determine the causal relationships that may exist between structural and process of care confounders and patient outcomes.

Chapter 5

RISK-ADJUSTED FUNNEL PLOT ANALYSIS OF RADICAL CYSTECTOMY OUTCOMES ACROSS ENGLISH NHS TRUSTS

The following chapter has been submitted as:

Mayer EK, Bottle A, Aylin P, Darzi AW, Vale JA, Athanasiou T. What is the role of risk-adjusted funnel plots in the analysis of radical cystectomy volume-outcome relationships? Submitted 2010.

5.1 Introduction

As highlighted in the introductory chapter, ranking in surgical performance has significant limitations, risks and implications which can be addressed by the use of statistical control charts. CUSUM are ideal for prospective assessment once 'control' has been confirmed. Funnel plots are good for identifying which processes are not yet in control and are practical tools, easy to use, and offer the opportunity to statistically define control limits around measurable outcomes and allow for adjustment of unmeasured factors, but should not be used to overcome significant inadequacies in data quality. Displaying performance data using funnel plots has advantages over conventional bar graphs and caterpillar plots [89].

The incremental model developed in studies two and three demonstrated the benefits of adjusting volume-outcome relationship data for structural and process of care factors and considering the hierarchical nature by co-examining institutional and surgeon volume. For this aggregated, 'high-level' analysis, there appeared to be little impact of case-mix adjustment using currently available variables from administrative data.

Risk-adjusted funnel plots can be produced for each step of the incremental statistical model and display risk-adjusted outcomes for each provider unit. Analysis at a more granular level might reveal volume-outcome relationships which are obscured by aggregated data analysis and consequently forms an important component of a broader methodological framework for volume-outcome relationship analysis.

The first aim of this study was to identify whether risk-adjusted funnel plots were useful as an addition to aggregated cross-sectional volume-outcome data analysis. The second aim was to increase the understanding, at the institutional level, of the impact of analysing volume-outcome data by combining funnel plots with an incremental statistical model approach.

5.2 Methods

5.2.1 Data extraction and model development

The data extraction, assignment of institutional volume bands, model development and outcome definitions are as previously described in studies two and three, sections 3.2 and 4.2.

Risk-adjusted funnel plots were produced for mortality and re-intervention rate as these two outcomes were shown, in studies two and three, to have a statistically significant institutional volume-outcome relationship following radical cystectomy. The definitions for total mortality and re-intervention within 30 days, as previously described, were used.

5.2.2 Creating risk-adjusted funnel plots

Funnel plots were produced in Excel using templates, freely available from the Eastern Region Public Health Observatory [180]. These funnel plots for proportions use exact binomial limits (based on the F-distribution). The templates were adapted to create risk-adjusted funnel plots and so that each Trust was assigned a different colour data point depending on whether it was low-, medium- or high-volume. A single funnel plot was created for each of models one to four, using the corresponding Trust-level crude outcome rates for model one or predicted probabilities for the other models.

The adjusted outcome rates, at the Trust level, were calculated by the formula:

(observed number of events for the Trust / total predicted probability for the Trust) x (total observed number of events in England / total number of operations in England)

Predicted probabilities at the patient level for model two were calculated using logistic regression in SPSS version 16. Predicted probabilities for models three and four were derived from three-level models in SAS v9.2 that allowed for the

clustering of patients within surgeons and of surgeons within hospitals. Predicted probabilities at the patient level were aggregated for each model to give the total predicted probability for each outcome by Trust.

Using Trust-level crude mortality rates, a bar graph and caterpillar plot were created to allow direct comparison of data presentation with an unadjusted funnel plot. 95% confidence intervals for the caterpillar plot were calculated in the standard fashion using the normal approximation:

95% CI = mortality rate +/- (1.96 x standard error)

Calculating the mean predicted probabilities by volume tertile for total mortality confirmed that these mean rates reflected the trends of the odds ratios derived from study two. Medium volume Trusts demonstrated a higher mean predicted probability as compared to low-volume Trusts and high-volume Trusts. A similar correlation of trends was seen between the predicted probabilities and odds ratios for re-intervention within 30 days, Table 5.1.

	Mortality		Re-intervention	
	Odds ratio	Mean predicted probability	Odds ratio	Mean predicted probability
Low-volume Trusts	1.00	0.0234	1.00	0.1373
Medium-volume Trusts	1.82	0.0324	1.52	0.1812
High-volume Trusts	1.19	0.0232	1.26	0.1795

Table 5.1 Summary table showing trends across Trust volume tertiles for odds ratios and mean predicted probabilities for total mortality and re-intervention within 30 days.

5.3 Results

5.3.1 Bar graph and Caterpillar plot analysis

Conventional performance data display as in Figure 5.1 does not consider volume of procedures performed and spuriously ranks Trusts against one another. Trusts 82 to 134 had a mortality rate greater than the average of 2.9%. Figure 5.2 is similar to Figure 5.1 in its ranking of Trusts but in addition plots the 95% confidence intervals. Trust A and Trust B are the only two that have their lower confidence interval above the average and so it can be stated that their observed mortality rate was greater than the national average. Two Trusts were observed to have lower than average rates (Trusts 41 and 46).

5.3.2 Funnel plot analysis

Using the same crude mortality rates, Figure 5.3 (model one), is an unadjusted funnel plot that demonstrates that no Trusts had truly divergent high mortality rates (outside the 3 standard deviation control limits). Trust A and Trust B, can be seen to be lying between the upper 2 standard deviation and upper 3 standard deviation control limits. Compared with medium and high-volume Trusts, considerably more low-volume Trusts had a recorded mortality rate of zero. Comparing model one with model two (Figure 5.3), there appeared to be relatively little impact at the individual Trust level of adjusting for case-mix. Adjusting for the influence of the surgeon (model three, Figure 5.4), demonstrated, in particular for Trusts A and B, that there was a considerable decrease in the adjusted mortality rates. In the final complex model, which adjusted for structural and process of care confounders, all Trusts, apart from two high-volume providers who recorded zero mortalities, simply show the natural variation inherent in any stable process (or, in SPC terminology, ‘common cause variation’ as opposed to ‘special-cause variation’).

For re-intervention within 30 days of operation, the unadjusted data identified a single high-volume Trust (Trust C) that had a truly divergent high rate of re-intervention, Figure 5.5, model one. A single medium-volume Trust was lying on

the upper 3 standard deviation control limit and a number of low-, medium- and high-volume Trusts lie between the upper 2 and 3 standard deviation control limits. As for mortality, there appeared to be relatively little influence of case-mix adjustment at the individual Trust level when comparing model one's funnel plot with that of model two, Figure 5.5. Adjusting for the influence of the surgeon caused the adjusted re-intervention rate of Trust C to return to sit within the control limits. Figure 5.6. In the final complex model all Trusts, apart from three low-volume Trusts who recorded zero re-intervention rates, were acting under common cause variation. For several medium-volume Trusts, however, adjusting for structural and process of care confounders did cause an increase in the adjusted re-intervention rate and a single Trust came to lie between the upper 2 and 3 standard deviation control limits.

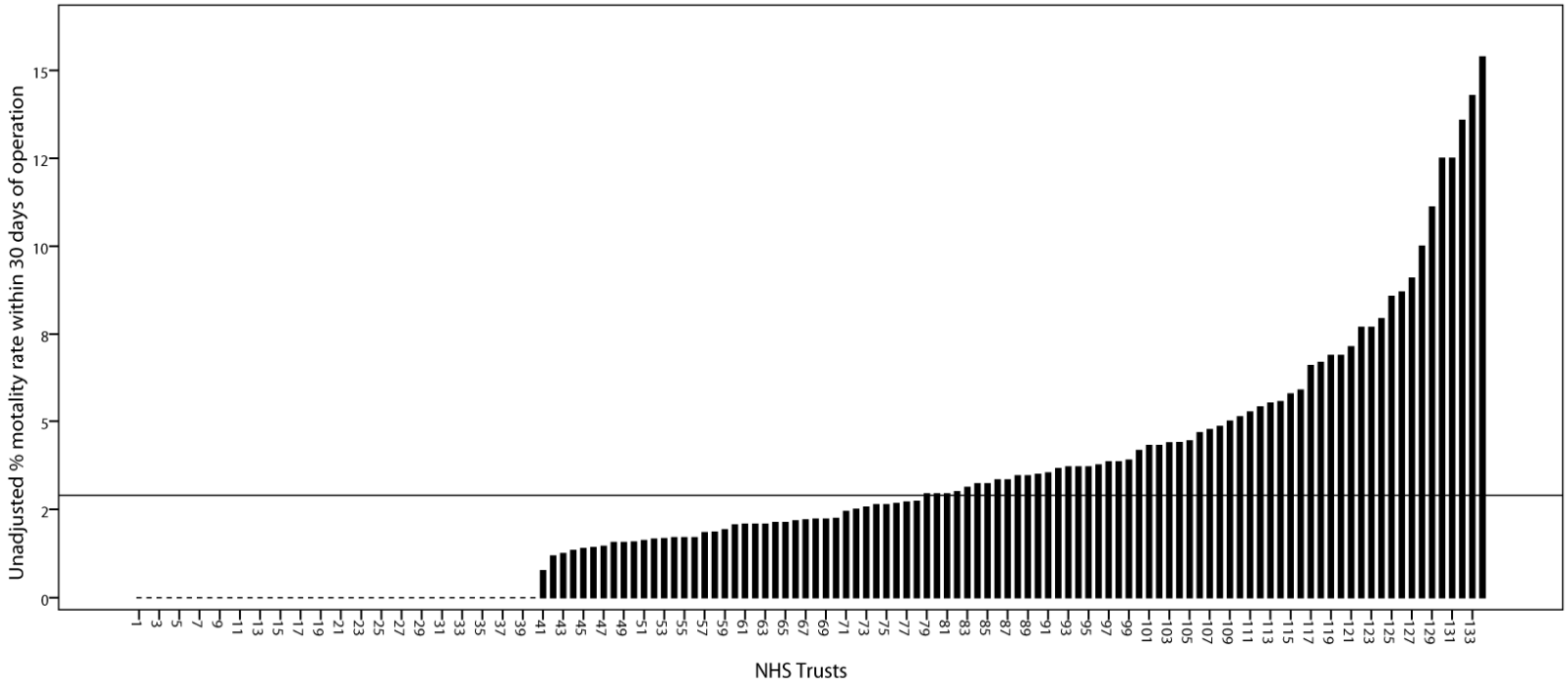


Figure 5.1 Bar graph of crude mortality rate within 30 days of operation across NHS Trusts in England. Horizontal intercept line represents the national average rate of 2.9%.

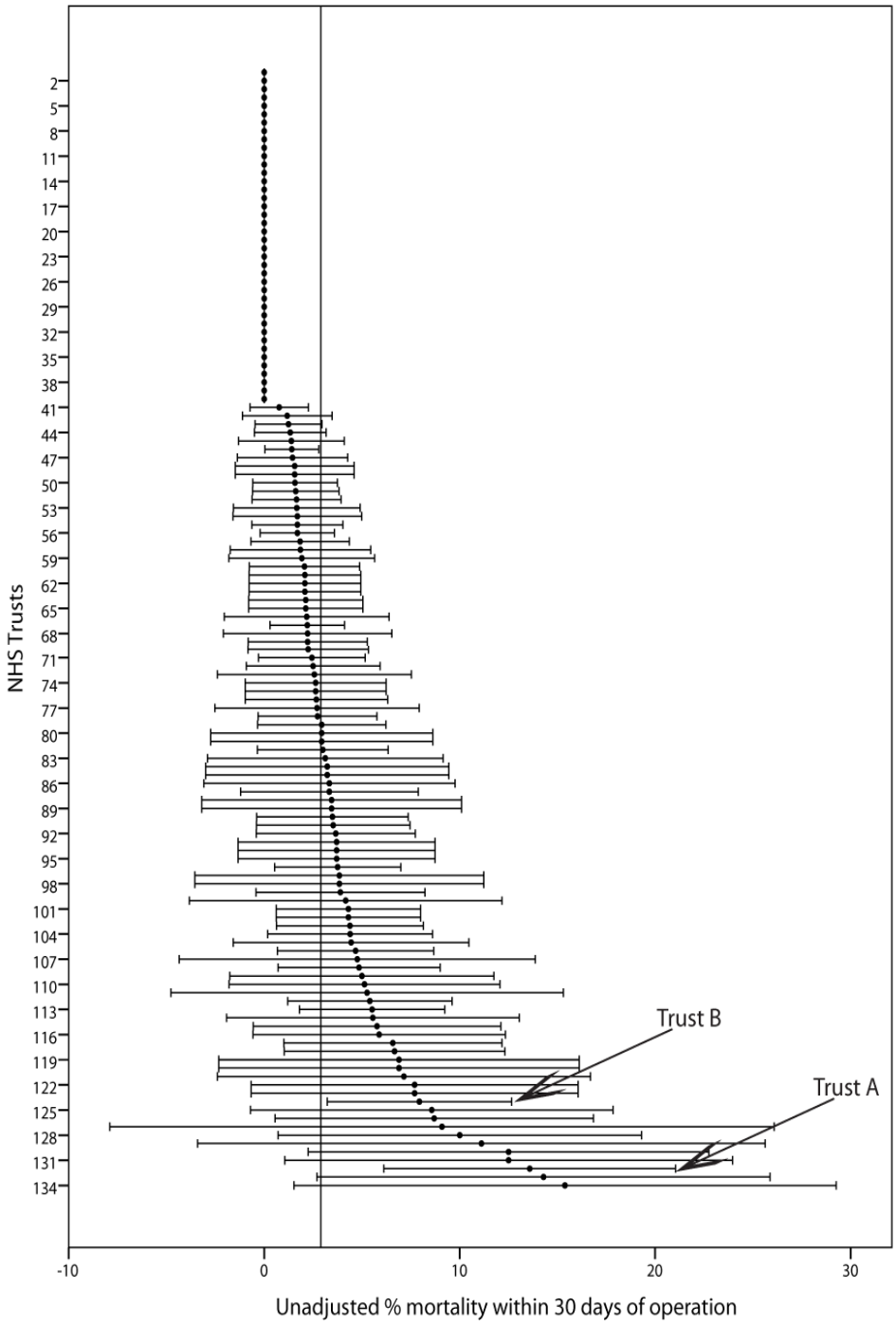


Figure 5.2 Caterpillar plot of crude mortality within 30 days of radical cystectomy for NHS Trusts in England. 95% confidence intervals are plotted and compared with the overall proportion of 2.9%

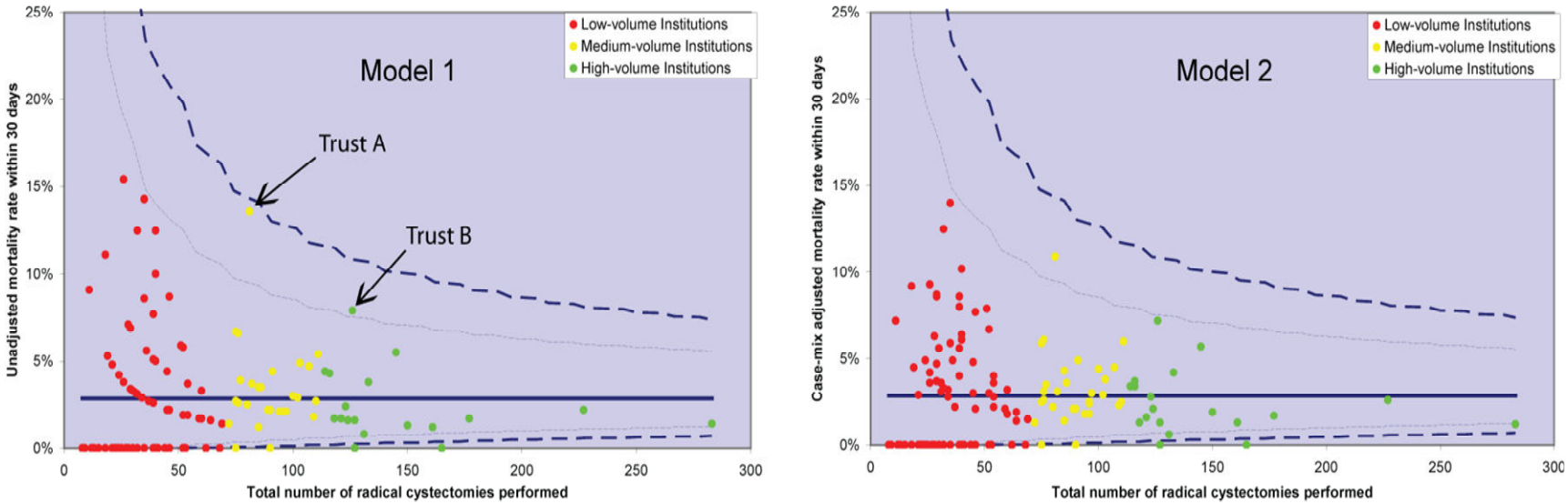


Figure 5.3 Risk-adjusted funnel plots of mortality within 30 days of operation at the Trust level for models one and two

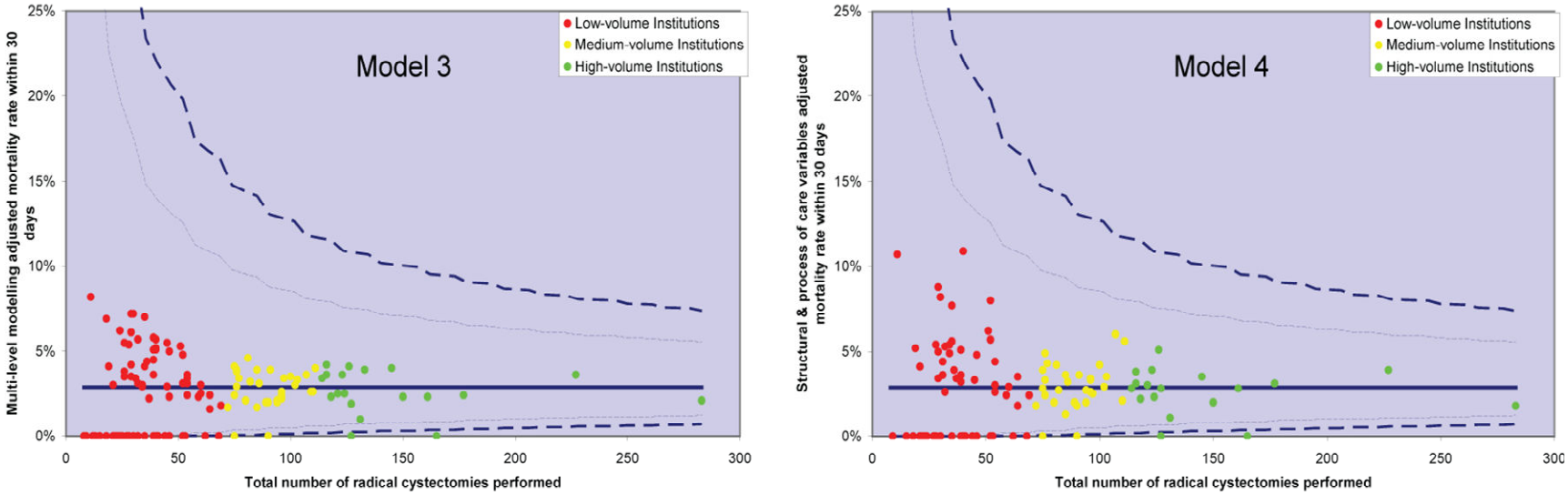


Figure 5.4 Risk-adjusted funnel plots of mortality within 30 days of operation at the Trust level for models three and four

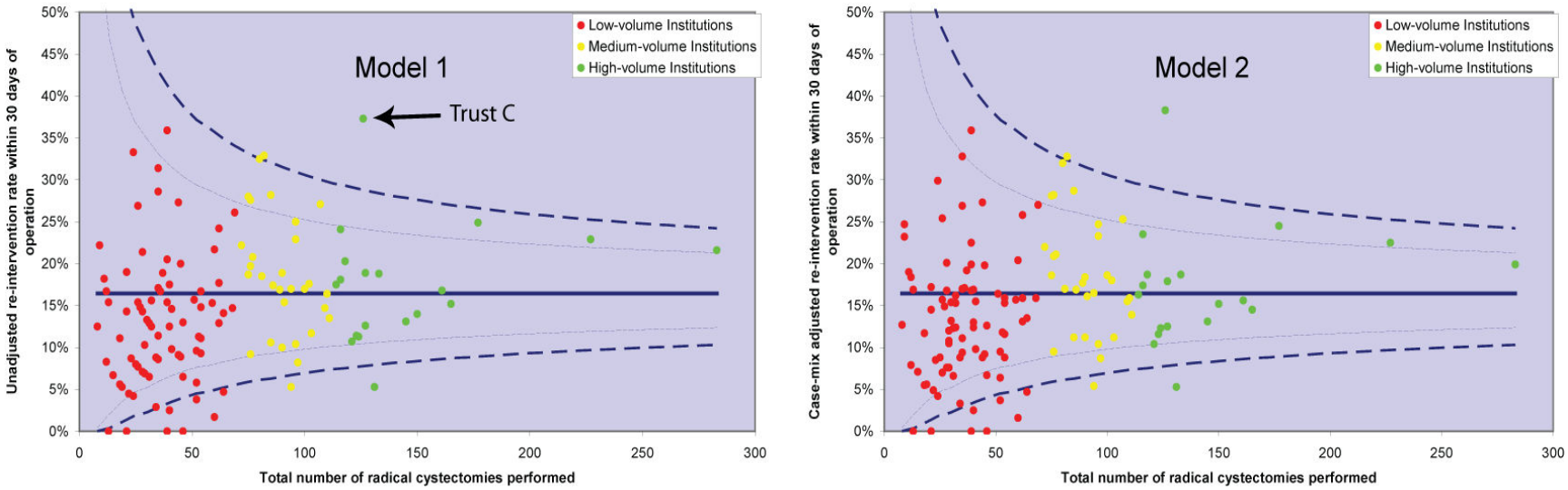


Figure 5.5 Risk-adjusted funnel plots of re-intervention within 30 days of operation at the Trust level for models one and two

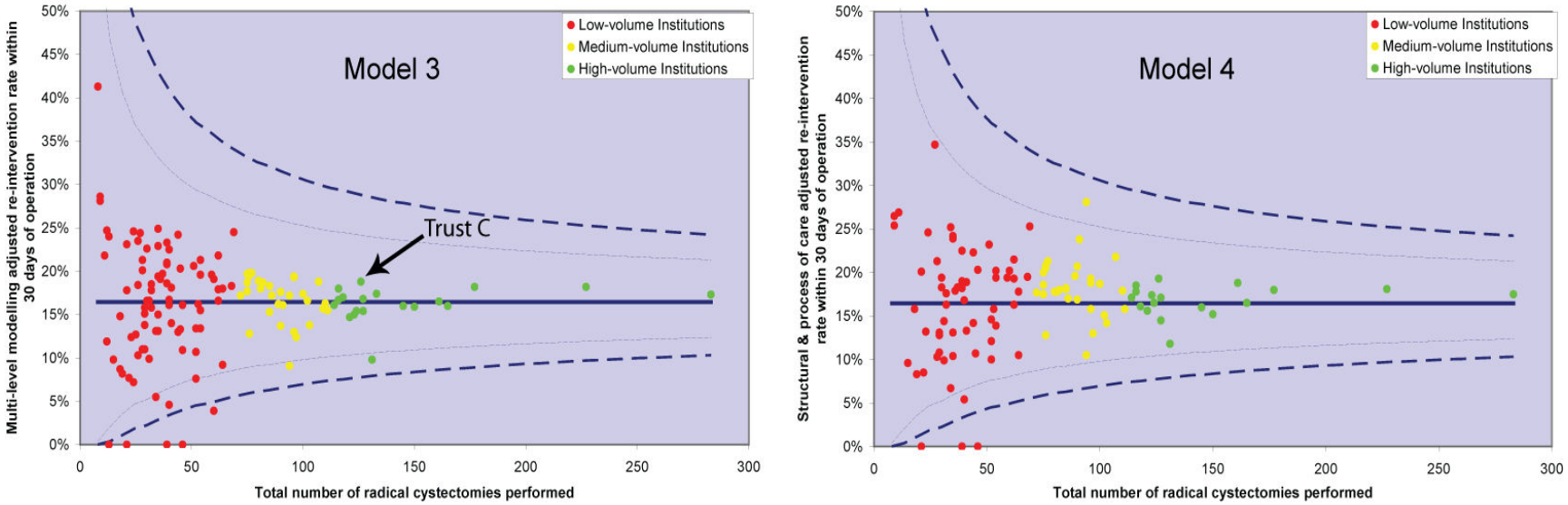


Figure 5.6 Risk-adjusted funnel plots of re-intervention within 30 days of operation at the Trust level for models three and four

5.4 Discussion

5.4.1 Principal findings

Ranking Trusts based on their mortality rates using a conventional technique identified that over a third of Trusts had a rate that was worse than the national average and which could be construed, in particular by the lay person, to have poor (divergent) performance when in reality this was not true. The addition of 95% confidence intervals as seen in the caterpillar plot did correctly identify that for all except two of these Trusts, the observed rate cannot be said, with any statistical certainty, to be different from the mean. Plotting the same data using unadjusted funnel plot methodology demonstrated that no Trust had truly divergently high mortality rates, although the same two Trusts highlighted by the caterpillar plot did lie between the upper two and three standard deviation control limits which might warrant further investigation.

No Trust in the final complex model risk-adjusted funnel plots exhibited special-cause variation with abnormally high mortality or re-intervention rates. Special-cause (non-random) variation signals the effects of external factors upon a process which requires further investigation. It was reported in studies two and three that when compared with low-volume Trusts, medium-volume Trusts had 82% greater odds of 30-day total mortality and 52% greater odds of re-intervention within 30 days. There were no outliers at the individual Trust level which satisfactorily explained this higher aggregated mortality odds ratio for medium-volume Trusts. Two high-volume Trusts in the final complex model, each having performed more than 100 cases, had a recorded zero mortality rate and were acting under special-cause variation with a truly divergent low mortality rate. Further investigation could well identify important lessons to be learnt.

For both mortality and re-intervention rate, adjusting for influence of the surgeon had the greatest impact. This was particularly true for re-intervention rates, where a single high-volume Trust that displayed special-cause variation with a high re-intervention rate was brought back ‘under statistical control’ (i.e. within the

control limits). This control was maintained when adjusting for structural and process of care confounders and confirms that further investigation at the surgeon level could identify important issues. A similar trend was seen for a single medium-volume and high-volume Trust with respect to mortality, although their performance was not truly divergent before adjusting for the surgeon influence.

5.4.2 Strengths and weaknesses of the study

Funnel plot methodology has the advantage of allowing for the simultaneous interpretation of performance with volume of procedures performed, a considerable advantage over methods such as caterpillar plots. As previously described, funnel plots can be used for the initial assessment of any volume-outcome relationship [89]. They cannot, however, overcome basic inadequacies in data quality such as 'coding errors'. For the volume tertiles generated in studies two and three, large numbers of very low-volume providers were not excluded which may have contaminated the low-volume tertile if it contained patient-level records present only through coding errors. It can be seen from the mortality funnel plots that a relatively large proportion of low-volume Trusts, as compared to medium and high-volume Trusts, have a mortality rate recorded as zero. It is impossible to say if this is from miscoding or is a 'true' zero reflecting the low overall operation count. This proportion of 'zero Trusts' would have influenced the aggregated odds ratios generated in study two thereby diluting any potential volume-outcome relationship. It is not possible to say anything more about these low-volume Trusts in terms of their mortality rates from the current data, but further investigation could look to see if they are really carrying out these procedures. The plotting of individual provider level data for funnel plots, therefore, has advantages when interpreted alongside the more traditional aggregated volume-outcome relationship research.

Plotting risk-adjusted funnel plots as an adjunct to aggregated volume-outcome analysis also allows for a degree of model development validation. As can be seen from this study, colour coding the Trusts in each volume tertile demonstrated that the annual cystectomy rate cut-offs used for the volume tertiles correlated well with overall number of procedures performed. This means that at the Trust level,

volume categorisation was reflective of both total procedures performed and an annual rate of procedures performed. This is maybe not surprising at the Trust level, but may not hold true at the surgeon level, where newly appointed consultants and those retiring during the time period analysed will influence the correlation between experience gained from having performed a certain total number of procedures and a critical number of procedures performed on an ongoing basis to maintain performance levels. The importance of overall number of procedures performed and quality of training on ongoing performance in uro-oncological surgery has been previously highlighted [181] [182].

Funnel plot control limits can be generated to ask a number of different questions, but this study has only used them to assess which Trusts' rates are not compatible with the overall average. As no Trust in the funnel plot using the final complex model for risk-adjustment, for both mortality and re-intervention, can be confidently said to have a performance worse than the national average, constructing different control limits to further explore which Trusts are probably extraordinary performers was not necessary [183].

5.4.3 Comparison with other studies

To the best of my knowledge, no study has previously published or made publicly available a similar funnel plot analysis of outcomes for radical cystectomy either in England or elsewhere. Direct comparison with other studies is therefore not possible. The advantages of funnel plots as a graphical aid for comparing institutional performance and avoiding spurious ranking of institutions into 'league tables' has been reported [89]. The case for a more widespread application of funnel plots in surgery in order to help to overcome some of the existing difficulties in assessing and reporting performance was made in chapter 1, section 1.3. The use of funnel plot methodology in combination with a more statistically robust methodological framework for volume-outcome relationship assessment has not been previously reported.

From studies two and three, it was concluded that medium-volume Trusts had a statistically higher odds of mortality and re-intervention within 30 days of radical

cystectomy as compared with low-volume Trusts. High-volume Trusts had statistically no difference in odds as compared with low-volume Trusts. Although this incremental funnel plots analysis cannot refute this finding, it has provided additional information as to potential reasons for this finding. It was stated in study two that compared with low-volume Trusts, patients undergoing radical cystectomy in medium-volume Trusts had 82% greater odds of total mortality following their radical cystectomy. From Figure 5.4 (model four), it is difficult to see such a clear difference in performance between low-volume and medium-volume Trusts that cannot be attributed to common-cause variation alone. Furthermore, the greater odds of total mortality for medium-volume Trusts only became apparent after adjustment for structural and process of care confounders, i.e. model four vs. model three. By comparing the mortality funnel plots for models three and four (Figure 5.4) it can be seen that although there is relatively little difference in their appearance, two medium-volume Trusts of relatively higher volume (> 100 total cases) increase their mortality rates from 3.6% to 6.0% and 4.0% to 5.6%. This change, in two relatively high-volume Trusts within the medium-volume tertile, may have been instrumental in determining the statistically significant higher odds ratios calculated for medium-volume Trusts in the final complex model (study 2). The benefits of both visualising provider-level performance data and displaying them in combination with volume are therefore clearly evident. In study three, it was demonstrated that the magnitude of the greater odds of re-intervention in medium-volume Trusts compared with low-volume Trusts was similar for all models. By comparing the re-intervention rate funnel plots for models two and three (Figure 5.5 and Figure 5.6), it can be seen, however, that there was considerable impact of adjusting for the influence of the surgeon on the individual Trust level outcomes.

From the aggregated data analysis, it was postulated that case-mix adjustment, using variables as incorporated in studies two and three, was not always necessary for analysis of effects at a highly aggregated level. This appears to have been confirmed by the funnel plot analysis, which does not identify any significant changes, at the individual Trust level, between models one and two for either mortality or re-intervention rates.

5.5 Conclusions

Unadjusted and risk adjusted funnel plots as produced in this study do have a useful role to play as a component of a methodological framework for investigating the volume-outcome relationship at the institutional level. They help with model validation by displaying disaggregated outcomes at the provider level and account for unmeasured confounders, so reducing the opportunity for spurious labelling of outliers. They have the advantage over conventional bar graphs and caterpillar plots of accounting for volume.

Although funnel plots show something about the distribution of outcomes by volume of cases and should theoretically be able to demonstrate a volume-outcome effect, in this analysis, the aggregated volume-mortality and volume-re-intervention relationships previously reported in studies two and three are not obvious graphically. The interpretation of risk-adjusted funnel plots alongside more conventional aggregated volume-outcome relationship data can identify important areas for further evaluation and a combined approach to this question is therefore recommended.

There is a need for funnel plots to play a more dynamic role as quality control tools in future surgical performance assessment and expand their use as an adjunct within a methodological framework for investigating the volume-outcome relationship which could influence future healthcare delivery.

5.6 Key points arising from study four

- In the final complex model, no Trust in England demonstrated a truly divergent high rate of mortality or re-intervention within 30 days following radical cystectomy.

- Risk-adjusted funnel plots do have a useful role to play as a component of a methodological framework for investigating the volume-outcome relationship.

- Analysing volume-outcome data using risk-adjusted funnel plots can potentially influence the overall conclusions made from aggregated data analysis alone and therefore act as a useful adjunct for assessing the volume-outcome relationship.

- Applying risk-adjusted funnel plots to an incremental statistical model can help direct further investigation of confounding factors at the individual Trust level.

Chapter 6

PROVISION OF RADICAL PELVIC UROLOGICAL SURGERY IN ENGLAND AND COMPLIANCE WITH IMPROVING OUTCOMES GUIDANCE

The following chapter was published as:

Mayer EK, Bottle A, Darzi AW, Athanasiou T, Vale JA. Provision of radical pelvic urological surgery in England, and compliance with improving outcomes guidance. BJU Int. 2009;104(10):1446-51.

6.1 Introduction

The importance of the volume-outcome relationship as a means to improve the quality of care that patients receive has been acknowledged in England and Wales with health policy centralising oncological services across a number of specialities [43]. Two hypotheses have been used traditionally to explain higher volume-better outcome relationships reported by numerous studies [96]: "practice makes perfect" and the "selective-referral" hypothesis. The latter describes the phenomenon by which institutions or surgeons that display better outcomes receive more referrals and as a result treat greater volumes of patients [82]. Several volume-outcome studies have suggested a minimum number of operations that should be achieved to improve patient outcomes. Pertinent to England, McCabe et al. suggested that a volume of 11 cystectomies per year was associated with the lowest mortality rates [106]. Judge et al. demonstrated a 'U-shaped' association between 30-day in-hospital mortality and the mean annual volume of prostatectomies categorised into quintiles (1-14 through to 46-93) [121].

In 2002, NICE published IOG for urological cancer services, which was one supporting element for the implementation of the NHS cancer plan [153] and therefore served to improve treatment for cancer patients. Importance was placed on a regionalised multidisciplinary approach for the management of a number of urological cancers including radical pelvic surgery which considered radical cystectomy and radical prostatectomy in combination. The latest guidance for radical pelvic urological surgery indicated that a service should serve a catchment population of at least one million and consequently provide a minimum of 50 radical cystectomy/prostatectomy procedures in total per annum [161]. Although the idea of a catchment population has been introduced more recently, the minimum standard of 50 cases per annum had been a feature since the original guidelines.

This study had three aims: firstly to ascertain the compliance of institutions to IOG guidelines for radical pelvic surgery, secondly to explore, in more detail, the pattern of service provision for radical cystectomy and radical prostatectomy both

before and after the introduction of IOG, and thirdly to investigate the referral patterns between Trusts for radical cystectomy service provision.

6.2 Methods

6.2.1 Data extraction

Data extraction follows the principles described in section 3.2.1, but is again summarised here.

Data for inpatient elective cystectomies (OPCS4 code M34 occurring in any procedure field in any episode) were taken from Hospital Episode Statistics (HES) for the six financial years 2000/1 to 2005/6 and from the Secondary Users Service for 2006/7. This was repeated using the OPCS4 code M61 to extract elective prostatectomy. HES covers all inpatient and day case admissions in NHS hospitals in England. Each record in HES constitutes a finished consultant episode, which covers the continuous period during which a patient is under the care of the same consultant, whose GMC code is recorded. Episodes were linked into admissions using the patient's hospital number, the Trust and date of admission, and admissions were linked together if the patient was transferred to another Trust.

Admissions were excluded if they were emergencies, had invalid age, sex or length of stay or were day cases or did not have a primary diagnosis of cancer (ICD10 C66, C67, C68, D090 for cystectomy, and C61 for prostatectomy). The proportion of cystectomies and prostatectomies that had such a diagnosis was calculated for each year.

6.2.2 Assignment of radical cystectomy and radical prostatectomy volume bands

The assignment of Trusts into volume bands for radical cystectomy and radical prostatectomy was necessarily independent of each other, although the approach and exclusion criteria used were identical. The numbers of radical cystectomies /radical prostatectomies were counted by year and by NHS hospital Trust. To

account for Trust mergers, Trust codes were unified to reflect their status as of April 2007. Trusts were excluded due to very low volume if they either had fewer than three years of data or had an annual rate of less than two radical cystectomies/radical prostatectomies per year. There is no agreed technique for dealing with Trusts which have low numbers of radical cystectomies/radical prostatectomies recorded, either as a result of coding errors or not, and the exclusion criteria were based on an assumption that less than three years' data entry and mean annual radical cystectomy/radical prostatectomy rate less than two is unlikely to reflect true activity. Thirteen Trusts contributing 41 records from the radical cystectomy dataset and 17 Trusts contributing 73 records from the radical prostatectomy dataset were excluded on this basis. These excluded cases made up only 0.5% and 0.4% of the original datasets respectively. The remaining Trusts were put into ascending order of annual radical cystectomy/radical prostatectomy rate and divided into three groups of roughly equal volumes. The Trust volume bands defined for both radical cystectomy and radical prostatectomy are summarised in Table 6.1.

6.2.3 Assignment of IOG volume bands

After exclusion of very low volume Trusts as described above, the numbers of radical cystectomy and radical prostatectomy performed by each Trust were combined for each year to give a total annual IOG activity per Trust. Where there was no recorded activity for either radical cystectomy or radical prostatectomy within a Trust in any one year, a figure of zero was assumed and the IOG activity for that year was calculated using only the number of either radical cystectomy or radical prostatectomy performed (whichever was present). Trusts were then categorised into three groups of unequal size according to their IOG activity for each year: Group 1 = <25 procedures, Group 2 = 25-50 procedures, Group 3 = >50 procedures.

	Annual procedure rate	Total operations in volume band	Number of institutions in volume band
Radical Cystectomy volume bands			
Low	**>2 and < 10**	**3036**	**84**
Medium	**≥10 and < 16**	**2776**	**31**
High	**≥16**	**2784**	**19**
Radical Prostatectomy volume bands			
Low	**≥2 and <22**	**5956**	**69**
Medium	**≥22 and <35**	**6403**	**32**
High	**≥35**	**5979**	**18**

Table 6.1 Summary table indicating allocation of Trust/institutional volume bands for radical cystectomy and radical prostatectomy

6.2.4 IOG catchment populations

Each radical cystectomy and radical prostatectomy spell was attributed to a Middle Layer Super Output Area (MSOA) via the patient postcode. MSOAs are geographical units with a minimum population of 5000. A Trust catchment population for radical cystectomy and radical prostatectomy was then calculated using the proportionate flow method. This is described in more detail elsewhere [184]. In brief, the same proportion of the population within each area (MSOA) is assigned to a Trust provider as the proportion of operative spells (radical cystectomy and radical prostatectomy) that originated from that area. The assigned populations from each area are then added together to give an overall catchment population specific for the operation(s) of interest.

6.2.5 Radical cystectomy referral patterns

In order to explore the referral activity of patients for radical cystectomy, patient IDs of those who had undergone radical cystectomy were used to identify records for preceding endoscopic extirpation of lesion of bladder (M42) or diagnostic endoscopic examination of bladder with biopsy of lesion of bladder (M45.1, M45.8, M45.9) as an indication of where the diagnosis and decision for RC had been made - “final endoscopic bladder procedure”. For each patient, the OPCS4 episode occurring immediately prior to radical cystectomy was identified and the time period between date of admission for this final endoscopic bladder procedure and radical cystectomy noted. Differences in the recorded provider codes for radical cystectomy and preceding endoscopic bladder procedure were also identified.

6.2.6 Data validation

In the radical cystectomy dataset 24 patient IDs (combinations of date of birth, sex and postcode) had a total of 26 duplicate records. It was decided not to remove them as it was not always possible to determine which was the ‘real’ record and the duplicates only made up 0.3% of the total case number. In the radical prostatectomy dataset there were 120 patients each with a single duplicate record making up 0.7% of the total prostatectomy number.

Seven hundred and two patients (8.2%) in the radical cystectomy dataset did not have an identifiable, linkable record prior to their radical cystectomy admission record, and these were excluded from the referral pattern analysis only.

Some of the patient records did not have an assigned postcode which precluded them from the catchment population analysis. Missing postcodes made up 0.9% of the radical cystectomy dataset and 0.6% of the radical prostatectomy dataset.

Statistical analysis was performed using Statistical Package for Social Scientists (SPSS, version 16, Chicago, USA). Chi-square tests, logistic regression, linear regression and analysis of variance were used with significance defined at the 5% level.

6.3 Results

6.3.1 Improving Outcome Guidance adherence

When combining, at the Trust level, the number of radical cystectomies and radical prostatectomies performed each year, it can be seen that those Trusts performing less than 25 procedures per annum (Group 1) decreased over the studied period and the decrease was occurring prior to the introduction of Improving Outcomes Guidance. The number of Trusts performing greater than 50 procedures per year (Group 3) increased and, similarly, seems to have occurred irrespective of Improving Outcomes Guidance. The greatest effect of Improving Outcomes Guidance seems to have been on those Trusts performing 25 to 50 procedures per annum (Group 2) where numbers have fallen after 2004/05 (Figure 6.1). Group 1 had higher numbers than expected in 2000/01, but within the expected range for all other years; group 3 had lower than expected counts in 2000/01 and then higher in 2005/06 and 2006/07.

The absolute number of Trusts achieving the guidance of 50 or greater procedures per year increased from eight in 2000/01 to 39 in 2006/2007. This means that 34.2% of Trusts in 2006/07 were adhering to Improving Outcomes Guidance; this

compares with only 13.4% in 2002/03 – the year Improving Outcomes Guidance was introduced. There was a significant increase, during the time period analysed, in the percentage of Trusts achieving the recommended minimal case volume (36% rise in odds per year, P<0.0005 (OR 1.36, CI 1.24 to 1.50)) (Figure 6.2).

In 2006/07, only a single Trust achieved the recommended Improving Outcomes Guidance catchment population of greater than 1 million. The other 38 Trusts that achieved the cut-off of 50 cases per annum had a mean catchment population of 472,577 (range 336,714 to 920,464), Figure 6.3.

6.3.2 Pattern of service provision for radical cystectomy and radical prostatectomy

8596 records of cystectomy performed for cancer were extracted for the time period analysed. The number of cystectomies performed each year increased steadily from 1366 in 2000/01 to 1564 in 2006/07. The proportion of these assigned a diagnosis of cancer (mean 83.3%) did not vary significantly across the years (Chi square p value = 0.63). The number of prostatectomy operations performed for cancer over the time period studied increased significantly from 1,601 to 3,083 (p value of trend across years <0.01); in total 18338 radical prostatectomy records were extracted. The proportion of radical prostatectomies to overall prostatectomy numbers similarly increased, 79.4% in 2000/01 to 90% in 2006/07.

The percentage of patients undergoing radical cystectomy under 65 years of age was relatively unchanged year on year with a mean of 36.9% (Figure 6.4). The mean percentage of patients who underwent radical prostatectomy under the age of 65 was 59.4%, Figure 6.5.

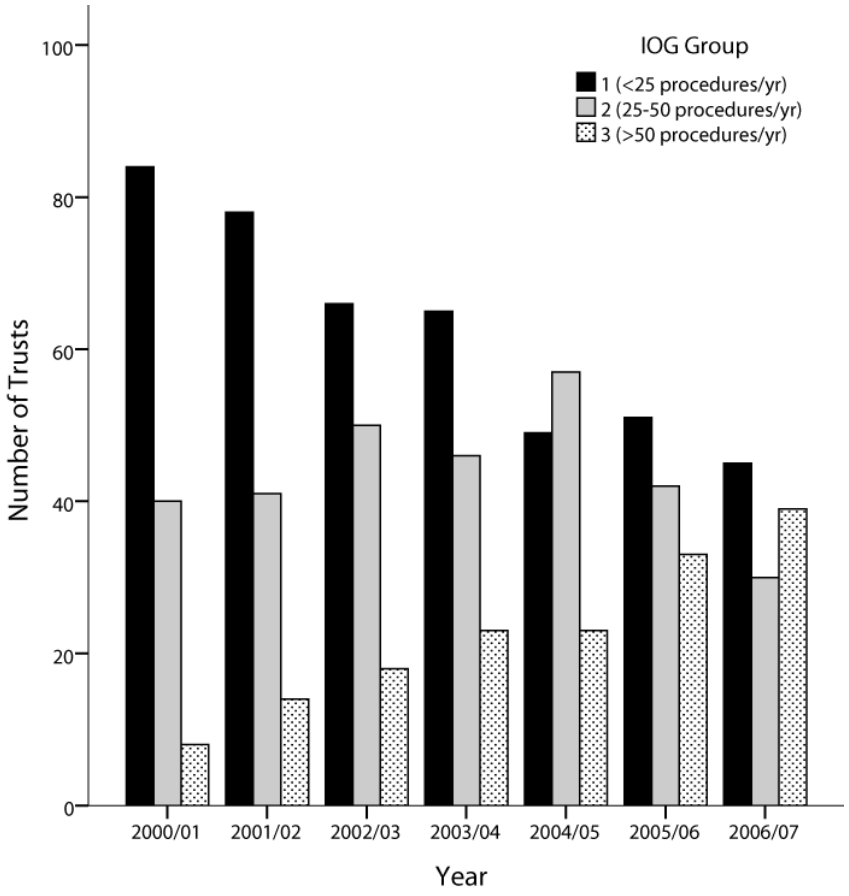


Figure 6.1 Histogram showing the variation in number of Trusts by IOG volume category over the time period studied

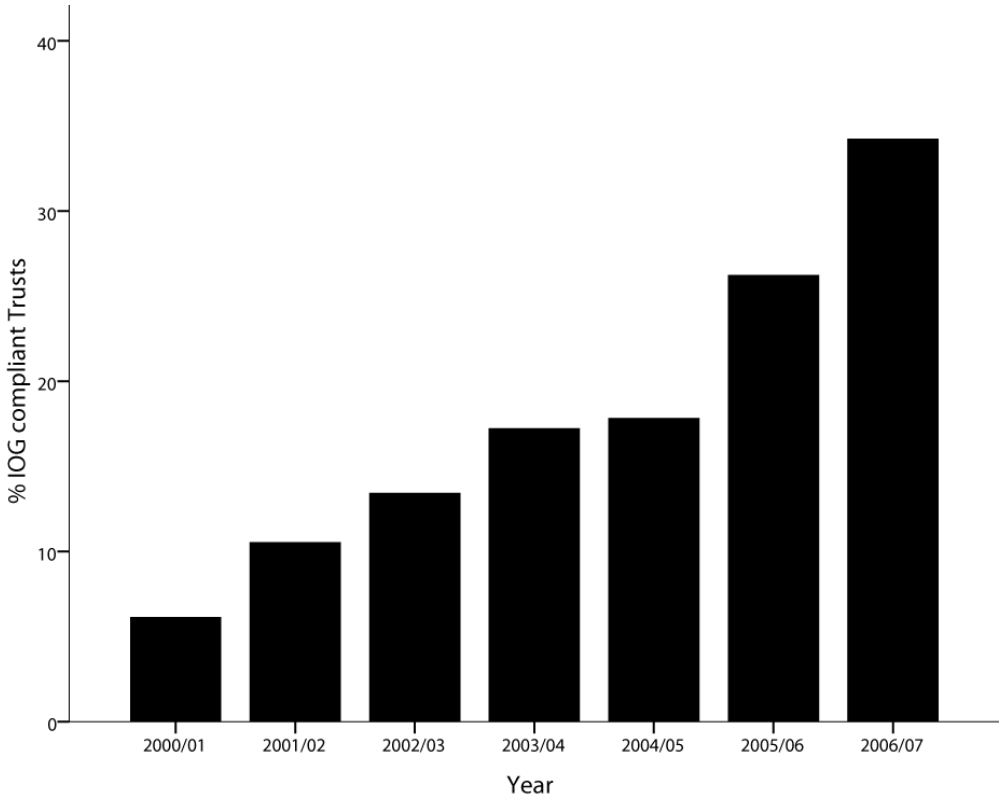


Figure 6.2 Figure demonstrating percentage of Trusts performing greater than 50 radical cystectomies and radical prostatectomies per year and therefore achieving IOG compliance

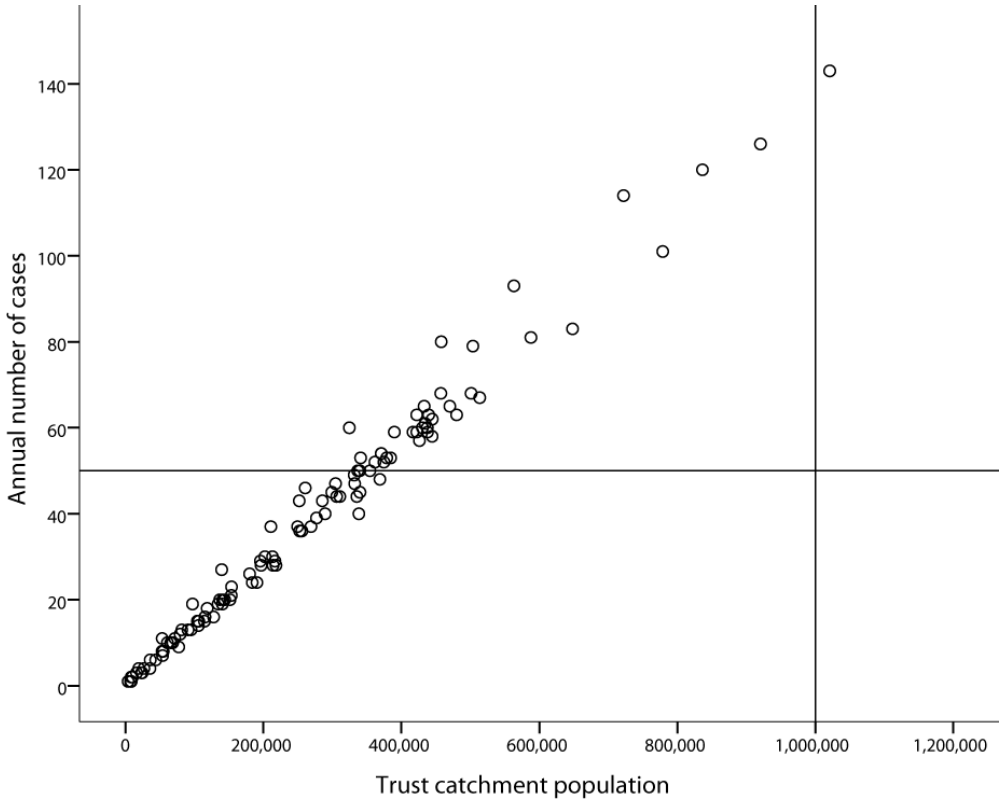


Figure 6.3 Scatter plot of catchment population versus annual number of combined cystectomies and prostatectomies at the Trust level. The horizontal intercept line indicates the cut-off of 50 cases per year and the vertical intercept line indicates the recommended one million catchment population per Trust

During the studied period, 30 Trusts were involved in a merger resulting in a change of their provider code, accounting for 575 operations (6.7% of total). No mergers took place after 2003. Mergers resulted in 12 providers (as defined by 2007 codes) changing volume band during the seven-year period (five Trusts moved from volume band two to three and five Trusts moved from volume band one to two; two Trusts would have been excluded).

The total number of Trusts performing radical cystectomy significantly decreased over the years (linear regression p value = 0.03), whereas the total number of Trusts performing radical prostatectomy did not significantly decrease over the years (linear regression p value = 0.6) (Figure 6.6). When categorised into volume groups, it is the low-volume Trusts both for radical cystectomy and radical prostatectomy that decreased in number over the years, although the decrease was not more than expected by chance. The numbers of medium- and high-volume Trusts remained static. The proportion of radical cystectomies performed by low-volume Trusts was less than expected in 2005/06 and 2006/2007. The proportion performed by high-volume and medium-volume Trusts was more than expected in 2005/06 and 2006/07 respectively (Table 6.2). For radical prostatectomy, the proportion of operations performed by low-volume Trusts was lower than expected in 2005/06 and 2006/07 having been more than expected from 2002/03 to 2004/05. Medium-volume Trusts performed more than expected procedures in 2005/06 and 2006/07 (Table 6.2).

Only 47% of Trusts that were high volume for radical cystectomy were also high-volume providers for radical prostatectomy, Table 6.3. 50% of radical prostatectomy high-volume providers were also high-volume radical cystectomy providers (data not shown). The R-squared linear regression coefficient for correlation between annual cystectomy rate and annual prostatectomy rate is 0.4, (Figure 6.7).

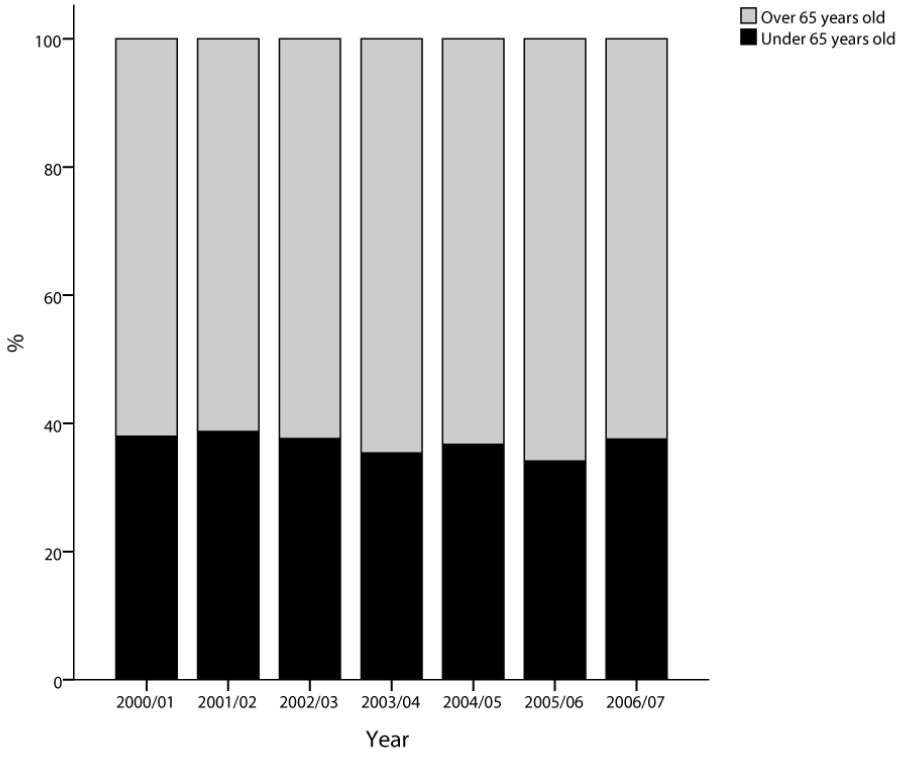


Figure 6.4 Percentage of patients, who underwent radical cystectomy, who are over or under the age of 65 years.

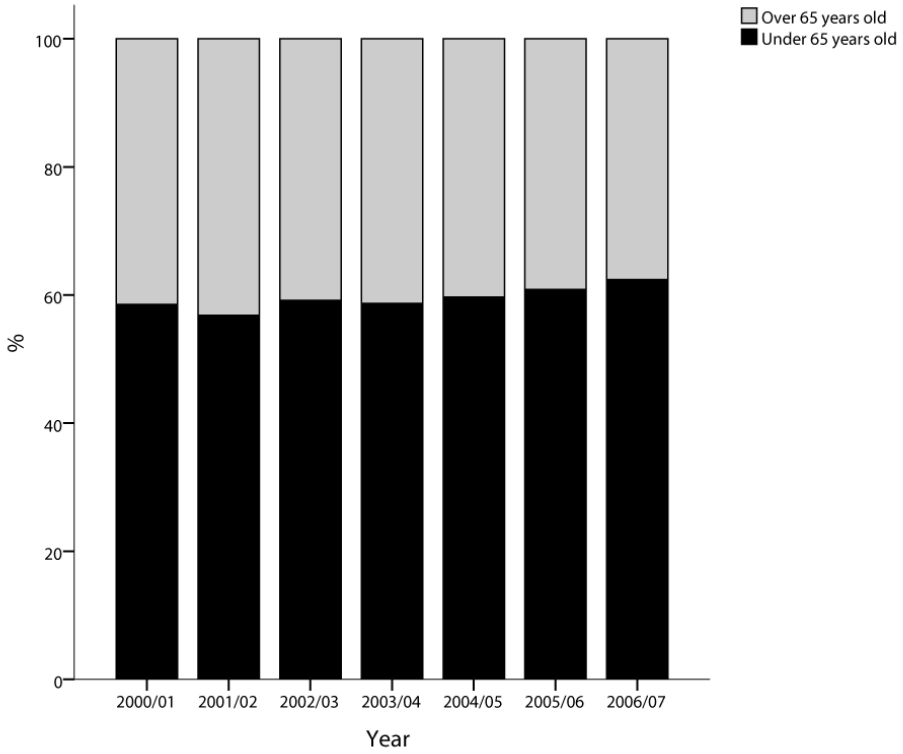


Figure 6.5 Percentage of patients, who underwent radical prostatectomy, who are over or under the age of 65 years

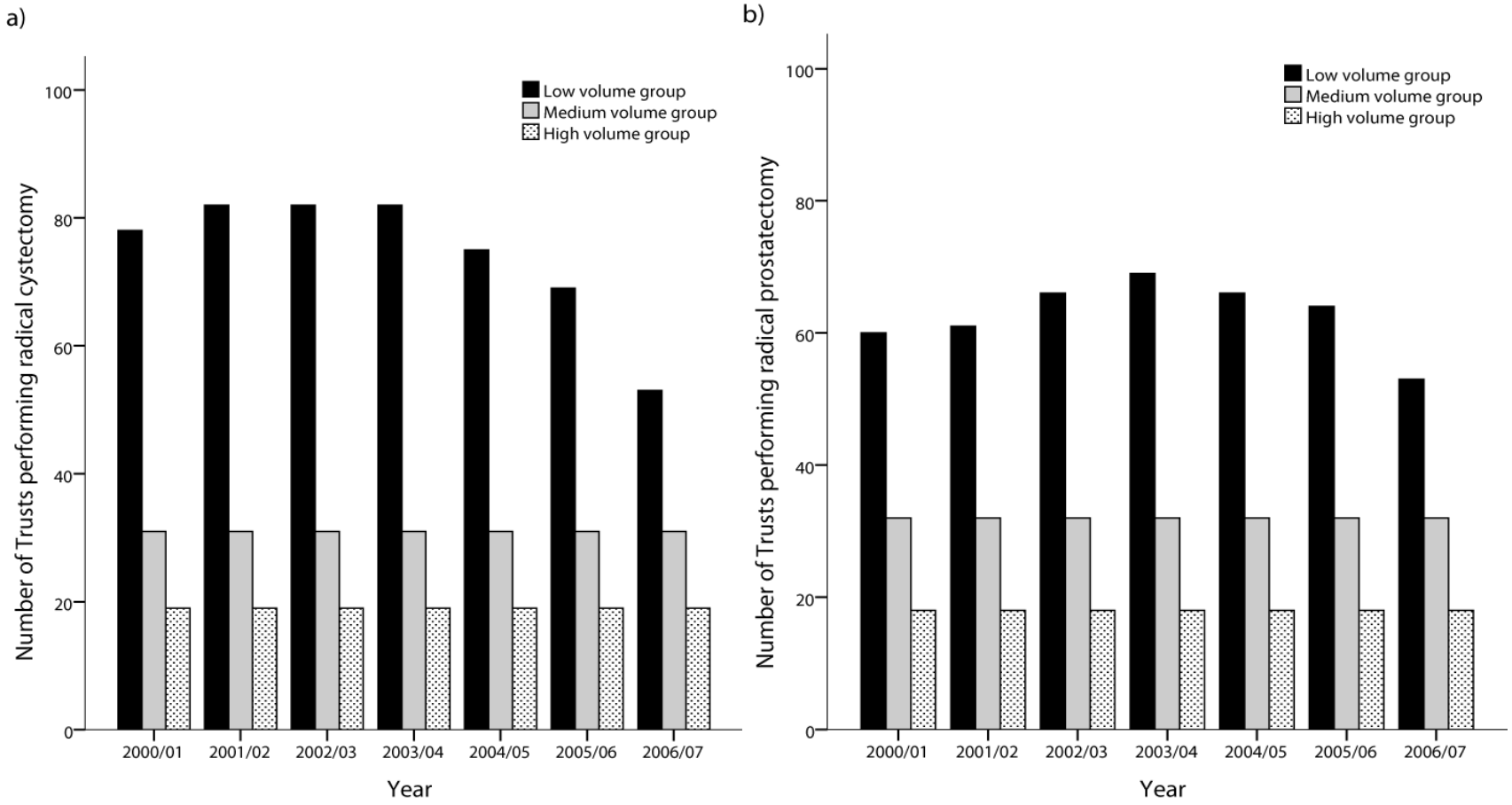


Figure 6.6 Clustered histograms demonstrating the breakdown of the number of Trusts performing a) radical cystectomy and b) radical prostatectomy by volume group across the years

Radical Cystectomy	Low-volume institutions	Medium-volume institutions	High-volume institutions
2000/2001	40.1 +	29.3	30.6
2001/2002	37.1	32.5	30.4
2002/2003	39.8 +	29.8	30.4
2003/2004	36.9	31.3	31.8
2004/2005	34.9	31.4	33.7
2005/2006	29.7 -	33.7	36.7 +
2006/2007	29.6 -	37.7 +	32.7
Radical Prostatectomy			
2000/2001	31.9	34.8	33.3
2001/2002	32.8	32.6	34.6
2002/2003	35.7 +	32.3 -	32.0
2003/2004	34.8 +	33.0	32.2
2004/2005	34.5 +	33.5	32.0
2005/2006	29.8 -	38.5 +	31.7
2006/2007	28.6 -	38.0 +	33.4

Table 6.2 Table summarising the percentage of the annual total operations (radical cystectomy and radical prostatectomy) performed by each of the Trust volume groups.

+ percentages that are more than expected by chance

- percentages that are less than expected by chance

		Trust volume tertile for radical prostatectomy			
		No volume group	**Low volume**	**Medium volume**	**High volume**
Trust volume tertile for radical cystectomy	**No volume group***	0	2 (100%)	0	0
	Low volume	17 (20.2%)	53 (63.1%)	10 (11.9%)	4 (4.8%)
	Medium volume	0	11 (35.5%)	15 (48.4%)	5 (16.1%)
	High volume	0	3 (15.8%)	7 (36.8%)	9 (47.4%)

Table 6.3 Comparison of Trust's volume tertile for radical cystectomy versus radical prostatectomy. Percentages calculated using total radical cystectomy number for each volume group as denominator

'No volume group' indicates either a Trust that had been excluded during the assignment of cystectomy/prostatectomy volume bands process or a provider that was not recorded as performing cystectomy/prostatectomy

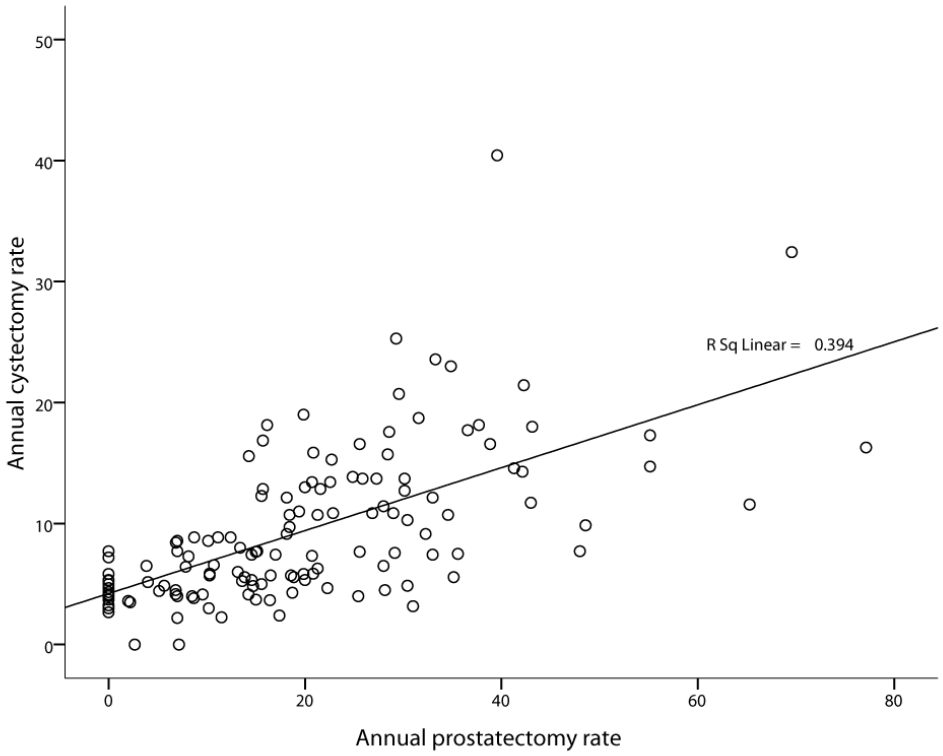


Figure 6.7 Correlation between annual cystectomy rate and annual prostatectomy rate at the Trust level.

6.3.3 Referral patterns for radical cystectomy

7081 patients (82.4%) underwent the "final endoscopic bladder procedure" at the same Trust as the radical cystectomy. 813 (9.5%) patients were referred to another Trust for their radical cystectomy. Of these, 61% were transferred to a high-volume radical cystectomy provider from a lower-volume Trust or a provider that had no volume band assigned (either a Trust that had been excluded during the assignment of volume bands process or a provider that was not recorded as performing radical cystectomy), Table 6.4.

Over the years there was a significant increase in the percentage of patients referred to another provider for their radical cystectomy, from 5.5% in 2000/01 to 19.6% in 2006/07 – 28% rise in odds per year, P<0.0005 (OR 1.28, CI 1.23 to 1.33). There were less than expected patients referred in the years 2000/01 to 2003/04 and more than expected in 2005/06 and 2006/07. For the 267 records referred from Trusts in the 'no volume band assigned group', 171 (64%) were from small acute Trusts, 27 from medium acute Trusts, 25 from a single large acute Trust, six from a single acute teaching Trust and one from an acute specialist Trust. 32 records were recorded as being referred from a Primary Care Trust (12%) and five from mental health Trusts.

The average time between admission for final endoscopic bladder procedure and radical cystectomy was significantly longer for patients who had different providers for the two procedures compared with those who did not (157 days versus 114 days, ANOVA p<0.0005). This difference remained across the time period assessed (Linear regression p<0.0005). To confirm significance of the difference in lag times between the two groups, trim points were generated for the 5^{th} and 95^{th} percentile to exclude lag time outliers (775 records excluded) which could have arisen from limitations of the methods used or date coding errors (the endoscopic procedure being repeated on day of cystectomy (62 records) or apparent extremely long waits). The mean lag times were reduced to 110 days and 91 days for those referred and those not referred respectively (also p<0.0005).

		Trust volume group - Cystectomy provider			
		1	2	3	Total
Trust volume group – Endoscopic bladder procedure provider	0	20	2	245	267
	1	86	156	225	467
	2	12	23	24	59
	3	6	3	11	20
	Total	124	184	505	813

Table 6.4 Breakdown of the 813 patients referred for radical cystectomy to a Trust differing from that at which their final endoscopic bladder procedure performed. The institutional volume grouping was based-on radical cystectomy provision only

Trust volume groups defined using annual cystectomy rate: >2 and < 10 (Group 1), ≥10 and < 16 (Group 2) and ≥16 (Group 3). Group 0 indicates either a Trust that had been excluded during the assignment of cystectomy volume bands process or a provider that was not recorded as performing cystectomy

6.4 Discussion

6.4.1 Principal findings

Improving Outcomes Guidance appears to have had a positive effect on the centralisation of radical pelvic surgery. The decrease in low-volume providers prior to Improving Outcomes Guidance implementation (approximately 20 low-volume Trusts) is likely to have reflected the natural regional reorganisation of services as a result of Trust mergers. After 2003/04 (when no mergers occurred), there was a continual decrease in low-volume providers and furthermore a decrease in medium-volume providers was seen after 2004/05 when previously numbers had been relatively stable. Although there had been a continued increase in the number of Trusts performing more than 50 cases during the introduction of Improving Outcomes Guidance in 2002/03, there was a greater increase from 2005/06. Correspondingly there was a greater then expected increase in the number of patients referred to another provider for their radical cystectomy during 2005/06 and 2006/07 and this is therefore likely to have represented a true Improving Outcomes Guidance phenomenon.

It is plausible that the Improving Outcomes Guidance took time to take effect and as such studying change over a wider time period was necessary to reflect this. Of those Trusts compliant with Improving Outcomes Guidance in 2006/07, all but one did not have a catchment population greater than one million and the majority had a population less than half a million.

The subject of volume-outcome relationships in surgery is a contentious one and this may explain in part the relatively slow uptake of Improving Outcomes Guidance and the number of Trusts that remained non-compliant in 2006/07. This may reflect sufficient doubt of clinicians as to the quality of existing evidence upon which Improving Outcomes Guidance has been built and therefore a reluctance to change local service arrangements. It was acknowledged in the Improving Outcomes Guidance for Urological Cancers research evidence [185] that

"Most of the published research on urological cancers focuses on clinical evaluation of treatment; relatively little direct research has been carried out on the organisation and delivery of services. In addition, for many service delivery issues, RCTs (categorised here as the highest quality evidence) are not feasible. Therefore, research designs which are regarded as of relatively poor quality for evaluating a clinical intervention may be the most reliable available for assessing the effectiveness of service delivery".

The authors concluded that based on US data, having acknowledged its limitations, for radical prostatectomy, concentration to a small number of professionals appeared to lead to an increase in the effectiveness of that intervention. Evidence from a single English study did not support a similar finding for bladder cancer survival. The authors also acknowledged that increased volume alone would not necessarily lead to an improvement in the quality of care offered to patients. It is therefore maybe not surprising that engagement of clinicians with the centralisation policy is not forthcoming.

Other inhibiting reasons for slow uptake of Improving Outcomes Guidance could also include local commissioning processes, new patient-choice reforms, public health factors such as patient willingness to travel and existing local referral networks amongst clinical colleagues.

Improving Outcomes Guidance for radical pelvic urological surgery assesses radical cystectomy in combination with radical prostatectomy and therefore considers important the impact of different, but related, procedures (non-index procedures) on outcomes of the procedure of interest (index procedure). Even if there are different surgeons performing these operations, the institutional set-up and processes of care will influence the patient treatment episode irrespective of the index procedure. This is also reflected pre-operatively in the multi-disciplinary team environment and making available to the patient the most appropriate, evidence-based, management options. Gilbert et al. demonstrated that the relationship between surgical volume and mortality was reduced by 20% for radical prostatectomy and 60% for radical cystectomy after adjusting for surgical

volume of 'non-index' operations (radical nephrectomy and radical cystectomy and radical nephrectomy and radical prostatectomy respectively) [186]. This current study has shown that there is correlation between the numbers of radical cystectomies and prostatectomies performed on a yearly basis at the NHS Trust level, although there is a fair amount of spread, (Figure 6.7). In study two, adjusting for the institutional total urology episodes, although differing significantly across institutional volume groups, did not affect the volume-mortality relationship (Table 3.7). This is in keeping with the findings of Fradet et al. who demonstrated that a higher hospital radical cystectomy volume appeared to lower the risk of complications after other common urological oncological procedures (radical prostatectomy and nephrectomy), but not after non-oncological urology procedures [187].

The finding that about 37% and 59% of patients each year, who underwent radical cystectomy and radical prostatectomy respectively, were under the age of 65 years is further evidence why translating US based data to the UK setting can have limitations. Some of the US databases, as a result of the insurance schemes, are not representative of the entire treated population by only collating data on those over 65 years.

6.4.2 Strengths and weaknesses of study

There is no agreed process for dealing with potential coding errors when handling HES data for studies of this nature. During the assignment of volume groups for radical cystectomy it was decided to exclude very low volume Trusts (fewer than three years of data or an annual rate of <2 cystectomies per year) on the basis that it was not certain if these records reflected true activity. Justification for our exclusion criteria is provided by the referral pattern data. Of the 13 Trusts excluded during volume group assignment, 11 were subsequently identified as routinely referring their patients to another provider for radical cystectomy, contributing 193 of the total 813 (24%) referred patient population. This substantiates that the very low-volume radical cystectomy activity identified for these 11 Trusts was correctly excluded as coding errors. Similarly, this finding provides supporting evidence for the approach taken to exclude very low volume

Trusts in the volume-outcome relationship model presented in studies two and three.

It was not possible to track the lag time between diagnosis of prostate cancer and radical prostatectomy in a similar fashion to that done for radical cystectomy. This is because outpatient procedures, such as prostate biopsy, are generally still not recorded in the HES database. This could potentially be overcome by combining national cancer registry data with HES data but, as demonstrated by a recent study, discrepancies when combining datasets can result in loss of a large number of the study population [173]. It is possible that the percentage of patients identified as having been referred to another Trust for their radical cystectomy is under estimated because some institutions may routinely repeat the "final endoscopic bladder procedure" prior to radical cystectomy when taking a patient from elsewhere.

6.4.3 Comparison with other studies

The most recent study that commented on Improving Outcomes Guidance for radical pelvic surgery looked at volume-outcome relationship for radical prostatectomy in England [121]. Having categorised mean annual radical prostatectomy volumes over an eight year period into quintiles they calculated potential Improving Outcomes Guidance adherence by adding a mean Trust cystectomy rate in 2004/05 to the quintile volumes for radical prostatectomy ("on average Trusts do 10 radical cystectomies per year in 2004/05"). On that basis they suggest that about 20% of Trusts would achieve the Improving Outcomes Guidance 50 cases per year. This current study suggests that this figure may be even lower at 17.8% for 2004/05 using the technique of combining radical cystectomy and radical prostatectomy at the Trust level. Furthermore, there was significant year-on-year variation in this figure. In addition to the different methods used, a further possible explanation for the difference in results between studies is that Judge et al. included all prostatectomy cases and not just those performed for cancer, i.e. radical prostatectomy. Moreover, Judge et al. quoted a median annual hospital Trust volume of 35 in 2004/05 as compared with a median of only 25 from this analysis. The group code M61 incorporates procedures such

as transvesical prostatectomy. Irrespective of the differences discussed, Judge et al. showed that no more than a fifth of institutions performing radical prostatectomy approximately two years after introduction of Improving Outcomes Guidance were compliant. From this current study, this figure has only risen to approximately one-third compliance in 2006/07.

6.5 Conclusion

Improving Outcomes Guidance appears to have had a positive effect on the centralisation of radical pelvic surgery although it took time to happen. However, four years after its introduction only a third of English NHS Trusts achieved the minimum standard of 50 procedures. All but one of these did not have a catchment population of one million. The impact of this centralisation on patient outcomes and quality of care for radical cystectomy and radical prostatectomy in England has yet to be fully ascertained, but could be further explored by longitudinal data analysis. An analysis of this sort would need to incorporate the methodological improvements described in studies two and three and also incorporate risk-adjusted statistical control charts as part of an improved methodological framework for assessing quality of care.

6.6 Key points arising from study five

- In 2006/07, only about one third of NHS Trusts in England were compliant with Improving Outcomes Guidance for volume of radical cystectomies and radical prostatectomies performed.

- Of those Trusts who performed more than 50 cases per annum, only a single Trust served the recommended catchment population of at least one million.

- Improving Outcomes Guidance does appear to have had some positive effect on centralisation of services with a decrease in low and medium-volume providers in recent years.

Chapter 7

CONCLUSIONS & FUTURE WORK

Elements of the following chapter were published as part of:

Mayer EK, Purkayastha S, Athanasiou T, Darzi A, Vale JA. Assessing the quality of the volume-outcome relationship in uro-oncology. BJU Int. 2009;103(3):341-9.

Mayer EK, Chow A, Vale JA, Athanasiou T. Appraising the Quality of Care in Surgery. World J Surg. 2009;33(8):1584-93.

Mayer EK, Bottle A, Darzi AW, Athanasiou T, Vale JA. The volume-mortality relation for radical cystectomy in England: retrospective analysis of hospital episode statistics. BMJ 2010 Mar 19;340:c1128. doi: 10.1136/bmj.c1128.

7.1 Conclusions

This thesis has evaluated methods to assess the volume-outcome relationship as applied to radical cystectomy in English NHS Trusts and provided justification for future research in this area to be conducted under the guidance of a sound methodological framework.

Such a methodological quality framework was originally proposed by the IOM in 2000. It proposed clear guidance for future volume-outcome research, having extensively evaluated research up to that time, and incorporates two areas specific to methodological quality improvement (Table 7.1). Despite this guidance, study one demonstrated that the quality of existing volume-outcome research as applied to uro-oncology was only modest at best and this was consistent with previously studies reporting on similar analysis [85] [91]. The recommendations of the IOM therefore appear to have had a limited impact and yet the application of a validated scoring system, which is readily useable and can be interpreted by healthcare professionals who may not have an in-depth knowledge of the area, has the potential to enhance interpretation of the volume-outcome literature. This in turn would improve translation to clinical practice and the practice of evidence-based surgery by addressing several fundamental questions [86].

The paucity of volume-outcome research relevant to uro-oncology and originating from a UK healthcare setting was highlighted as was the potential for more methodological robust statistical techniques to challenge the conclusions of previous research with the secondary implications for existing healthcare policy, which has advocated centralisation of services.

Methodological Development

- Examine outcomes other than mortality (e.g., functional status, quality of life), including longer-term outcomes
- Include potential intervening variables (e.g., processes and systems of care) in volume– outcome studies
- Integrate social science methods (e.g., sociology, medical anthropology, systems analysis) into volume–outcome research
- In studies with mortality as an end point, examine the time of death to help explain the cause of death
- Develop procedure-condition-specific risk-adjustment tools.

Health Services Research Data Infrastructure

- Develop condition-procedure-specific, prospective, population-based clinical databases and registries (e.g., New York's cardiovascular surgery database)
- Develop chronic disease databases that cover both hospital and outpatient care
- When clinical databases are available to serve as a "gold standard," evaluate the sensitivity and specificity of volume as a quality indicator
- Selectively add key clinical risk factors and process data into administrative databases
- Include more reliable identifiers of physicians in federal and state administrative databases.

Table 7.1 Guidance as developed by the Institute of Medicine for improving the quality of volume-outcome research

The volume-outcome relationship for radical cystectomy in the English healthcare setting was evaluated, using an improved statistical/methodological framework that considered the hierarchical structure of the institutional and surgeon influences (multi-level modelling), appropriate handling of the provider volume variable, adjustment for structural/process of care measures, and measured outcomes other than mortality. A volume-outcome relationship for mortality and re-intervention following radical cystectomy was evident across English NHS Trusts. At the institutional level, medium-volume institutions were found to have statistically significant greater odds of mortality. Appropriate interpretation of the volume-mortality relationship was only possible after adjusting for institutional and surgeon volumes and structural and process of care confounders. The finding that medium-volume institutions, as compared with low-volume institutions, had higher odds of mortality was surprising. More consistent with previous studies was the finding that medium-volume institutions appeared to have worse in-hospital and total mortality than high-volume institutions. What is not clear is why low-volume institutions appeared to have comparable outcomes to high-volume institutions.

For re-intervention, a volume-outcome relationship existed at the institutional and surgeon level. Again, as for mortality, the odds of re-intervention within 14 and 30 days of operation, were found to be greater in medium-volume institutions when compared with low-volume institutions. High-volume surgeons (but not medium volume) were associated with a reduced odds of early re-intervention (within 14 days) compared with low-volume surgeons. The surgeon volume-re-intervention relationship only became apparent after adjusting for the influence of the institution and provided the strongest evidence for the influence that the surgeon/institution interaction can have on patient outcomes.

Adjustment for structural and process of care variables appeared to have the most impact at the medium-volume institutional level. The difficulties and dangers of trying to define an absolute minimum caseload are evident. More work is needed to define ‘medium-volume’ institutions and also the surgeon’s volume that makes it protective for institutional outcomes. Centralising care to specified institutions

without accounting for the inherent structural and process factors that result in optimum care is meaningless. Equally, the outcomes a surgeon can achieve are dependent on the support network provided at the institutional level. The relative contribution of all these factors needs to be considered when reconfiguring service delivery to improve the quality of care that a patient receives [177].

Analysing the same volume-outcome data using risk-adjusted funnel plots potentially influences the overall conclusions made from the aggregated data analysis alone and therefore act as a useful adjunct for assessing the volume-outcome relationship. Using the final complex model, no Trusts in England demonstrated a truly divergent high rate of mortality or re-intervention within 30 days following radical cystectomy. It cannot be said with any certainty therefore that any Trust has a performance worse than the national average for these outcomes. As risk-adjusted funnel plots also display disaggregated provider level data, applying them to an incremental statistical model can help direct further investigation of confounding factors at the individual Trust level and compliment methods of data validation.

The volumes of other related operations that may influence mortality risk for radical cystectomy were not included in the statistical model. The impact of overall uro-oncological operative volumes has been highlighted and, for cystectomy, adjustment for institutional volumes of nephrectomies and prostatectomies attenuated the institutional cystectomy volume-mortality relationship by 60% [186]. This is particularly important for English healthcare and Improving Outcomes Guidance for urological cancer services, as the minimum caseload guidance for radical pelvic surgery of 50 cases/year considers cystectomy and prostatectomy in combination. As demonstrated in study five, although there was some correlation between the numbers of radical cystectomies and prostatectomies performed on a yearly basis at the NHS Trust level, in 2006/07 only about one third of NHS Trusts in England were compliant with Improving Outcomes Guidance. There appeared to have been some positive effect on centralisation of services with a decrease in low and medium-volume providers in recent years, suggesting at least some degree of engagement, at the regional

level, with the volume-outcome agenda to improve quality of care. A survey of consultant urological surgeons in the UK suggested that most supported the principle of setting minimum volume thresholds for urological cancer operations, including radical cystectomy [188].

Longitudinal data analysis is important for assessing causality and causal direction, as it indicates whether outcomes improve if annual volume of activity increases over time. As previously reported [189], studies which have demonstrated volume-outcome relationships on cross-sectional analysis have subsequently found no relationship between changes in volume and outcome over time. The absence of reliable longitudinal data in volume-outcome research has been highlighted as a concern when trying to predict, with any certainty, the success of centralising services to improve outcomes [177].

7.2 Unanswered questions and future research

The volume-outcome relationship investigated in this thesis is only one component of a much broader future quality improvement framework that must combine other integral factors such as adherence to process of care measures encompassing the entire patient treatment pathway, other outcomes of surgery including longer-term clinical outcomes (cancer recurrence rates and cancer-specific survival), patient-reported measures including quality of life assessment and patient satisfaction/experience assessment.

Further volume-outcome research needs to explore the relative contributions of each of the factors, the institution, the surgeon and individual structural and process of care characteristics and explore the relationship between caseload volume and operative outcomes other than mortality. This, in combination with further efforts to improve the methods used for volume-outcome research, such as longitudinal data analysis to help establish causality, will be important for informing future healthcare service changes, such as centralising care for relatively infrequent operations like radical cystectomy and improving the quality of care delivered.

A novel quality framework will need to address the challenges of weighting, standardising, and combining a number of structural, process, outcome and patient-reported measures to produce a composite score. The Quality Measurement Task Force of The Society of Thoracic Surgeons compared a number of statistical methodological approaches to develop such a composite score for 11 quality indicators across four domains for coronary artery bypass grafting [77] (Table 7.2). The Task Force used nine principles to guide the selection of the quality indicators (Table 7.2). This work is the most comprehensive quality measurement programme to date and uses a scientifically rigorous approach to evaluate difficulties such as combining multiple measures into a single score, while maintaining the ability to isolate the individual components to allow for performance assessment of specific areas. A detailed description of this work has been published [78], but in summary recommends the use of Bayesian random-effects analyses, all or none scoring for 'within-domain' composite scoring and rescaling for combining across-domain scores. The latter two approaches avoid the subjective weighting of the individual quality indicators. All or none scoring ensures consistent directionality between 'quality' measures and risk-adjusted performance measures; the numerators of the risk-adjusted performance measures are defined as the number of patients who avoid the 'non-desired' end point.

Larger values of the numerator are therefore favourable for the chosen 'quality' measures. Rescaling considers that 'quality' measures have different scales of measurement. The Bayesian random-effects approach accounts for small sample size and incorporates a risk-adjustment for performance [78].

It is clearly impractical, in terms of time and resource, to collect and collate hundreds of 'quality measures'. Using the technique of data mining with factor analysis or structural equation modelling, pilot studies can be used to identify correlation among the measures. This establishes whether any of the measures can be eliminated if they provide redundant rather than complementary information. Furthermore, to confirm that each quality indicator is contributing to, but not dominating, the final composite score, the individual indicator to total score correlation can be calculated [78].

The Quality indicators were selected using the following nine principles:

1. Quality assessment should be at the level of the program or hospital rather than the individual surgeon
2. Initial quality reports should focus on coronary artery bypass grafting surgery
3. Quality measures should be chosen from among those endorsed by the National Quality Forum
4. Quality measure selection should be consistent with the principles and criteria recommended in the 2006 Institute of Medicine report "Performance Measurement: Accelerating Improvement"
5. Quality measures should be available as data elements within The Society of Thoracic Surgeons National Adult Cardiac Surgery Database
6. Quality scores should consider structure, process, and outcomes
7. Quality scores should assess three temporal domains: preoperative, operative, and postoperative
8. Quality scores should satisfy multiple criteria for validity
9. Quality scores should be interpretable and actionable by providers.

Eleven individual measures of coronary artery bypass grafting quality across four domains were selected:

Domain	Description	Quality indicator
1. Perioperative medical care	A process bundle of four medications*	Preoperative β-blockade
		Discharge aspirin, β-blockade, and lipid-lowering agents
2. Operative care	A single process measure	Use of at least one internal mammary artery
3. Risk-adjusted operative mortality	A single outcome measure	Mortality within 30-days of operation
4. Postoperative risk-adjusted major morbidity	The absence of a bundle of five complications*	Renal failure
		Deep sternal wound infection
		Reexploration
		Stroke
		Prolonged ventilation/intubation

Table 7.2 Details of the nine principles that the Quality Measurement Task Force used to select eleven indicators of coronary artery bypass grafting quality. Table drawn from information presented in reference [77].

* The bundle measures use the All-or None phenomenon – i.e. compliance has to be with all four medications and the patient has to be discharged without the occurrence of any of the five serious postoperative complications.

Reported findings from this thesis and the approach described above by the Society of Thoracic Surgeons highlight the need and importance for a high-level quality framework to be further adapted within specific surgical specialities, and even named procedures, to maximise the benefit and relevance to the quality of care delivered for particular cohorts of patients. Quality control can only be done for institutions and surgeons using prospectively collected procedure-specific outcome measures such as cancer control data, intra- and postoperative parameters and complications and short, medium and long-term morbidity including quality of life assessment. This will be an important move away from the frequent use of mortality as a surrogate measure of outcome. The generation of hybrid administrative and clinical datasets at a Trust level may allow us to evaluate, with much greater complexity, underlying determinants within the volume-outcome relationship including structural and process differences between Trusts, departments and surgeons and their impact on the quality of care that our patients receive. The framework will need to incorporate an appropriate statistical methodology such as multilevel modelling to ensure that the data is handled in the correct manner.

Although the public reporting of performance data stimulates quality improvement activity at the provider level [4], this does not always lead a patient to choose the most appropriate provider for him/herself [4]. 'Packaging' of this information for the public is therefore crucial and inconsistent presentation of information leads to variable results [3]. One key element relies on the ease with which the information can be interpreted by the public. Hibbard et al. suggest that it should "be immediately obvious who the top and bottom performers are, to stimulate quality improvement efforts" [190]. Yet the presentation of information on quality can create resistance and disengagement from quality improvement programmes. Providers must be able to trust the data (this can be achieved by giving them ownership), and trust the methods of presentation, such that there is not spurious ranking and unfair discrimination as a result of unmeasured or immeasurable factors. This is particularly important when quality improvement programmes are evolving and datasets may not reflect true case-mix adjustment, for example.

Although funnel plots are limited to displaying only one 'quality measure' at a time, multiple plots have been used, in combination, to provide an overall assessment of provider performance across a range of indicators [191]. The number of quality measures can be minimised by using techniques such as data mining, as described previously, to eliminate 'redundant' measures. This means that a smaller number of funnel plots can be produced displaying a bundle of variables providing complementary information.

Existing initiatives of quality appraisal have been important, but there is still the need for a universally accepted and validated quality framework which incorporates a multi-dimensional approach to measuring both the clinical and patient-orientated measures of quality of care. The assessment tools for patient-reported measures need to be agreed using an evidence-base and then made freely available through a quality forum to facilitate both national and international benchmarking and continuous improvements through feedback mechanisms.

APPENDIX A

Structural variables

Teaching status – Trusts were assigned a flag to say whether they had teaching hospital status according to the National Patient Safety Agency website.

Data source - http://www.npsa.nhs.uk/nrls/patient-safety-incident-data/feedback-reports/sample-reports-benchmarking/

Total available beds, all sectors – Calculated using summation of available beds, all sectors, across all seven years.

Data source Hospital Activity Statistics – http://www.performance.doh.gov.uk/hospitalactivity/data_requests/beds_open_overnight.htm

Total available acute beds – Calculated using summation of available acute beds across all seven years.

Data source Hospital Activity Statistics - http://www.performance.doh.gov.uk/hospitalactivity/data_requests/beds_open_overnight.htm

Total urology episodes – Calculated using summation of total urology episodes, defined as either the specialty code for the episode or a urology OPCS code ('M' and 'N' chapters). Data source – Hospital Episode Statistics

Average acute occupied bed rate – Calculated using total (across all seven years) occupied acute beds divided by total available acute beds by provider.

Data source Hospital Activity Statistics – http://www.performance.doh.gov.uk/hospitalactivity/data_requests/beds_open_overnight.htm

Average nurse to occupied bed ratio – Calculated using summation of all nurses by provider across all seven years divided by total occupied beds, all sectors, by provider. Data source of nursing workforce – The Information Centre Workforce Census (obtained by request).

Average critical care to total bed ratio – Calculated using total available critical care bed counts for six years available (2001/02 – 2006/07) divided by total available beds, all sectors, across corresponding six years.

Data source Hospital Activity Statistics - http://www.performance.doh.gov.uk/hospitalactivity/data_requests/critical_care_beds.htm

and

http://www.performance.doh.gov.uk/hospitalactivity/data_requests/beds_open_overnight.htm

Average registrars (all clinicians) to occupied beds ratio – Calculated using summation of registrars counts, whole time equivalents, (all specialities) across all seven years divided by total occupied bed counts, all sectors, across all seven

years. A single Trust had clinician counts only available for six years and corresponding years for bed counts were therefore used to calculate the average.

Data Source of registrar workforce - The Information Centre for Health and Social Care (obtained by request).

Urology registrars to total urology episodes – Calculated using summation of urology registrar counts across years available and divided by total urology episodes across years available.

Data Source of registrar workforce - The Information Centre for Health and Social Care (obtained by request).

Process of care variables

Average operation cancellation rates (Rate not admitted within 28 days following cancellation – Average calculated across six years for which data available (2001/02 – 2006/07).

Data source Hospital Activity Statistics - http://www.performance.doh.gov.uk/hospitalactivity/data_requests/cancelled_operations.htm

Waiting times for surgery (days) – Calculated at the patient ID level using the HES fields "Elecdate" (booking date) minus "admidate" (admission date for cystectomy). Data source – Hospital Episode Statistics

APPENDIX B

ICD-10 codes used to define potential complications resulting from surgery:

Procedure-specific complications

N328, T810, T811, K661, R571, R58X Haemorrhage/haematoma

A40, A41, A49 Sepsis

R31 Unspecified haematuria

S366 Rectal Injury (+/- formation of colostomy)

N995 Stoma complication (ischaemia, necrosis, infection, bleeding, prolapse)

R390 Urine leakage (anastomtic breakdown)

N393, N394, R32 Incontinence of urine

N484 Impotency (erectile dysfunction)

R33 Retention of urine

S346, S348, S741, S742, S747, S748, S749, T133 Neuropraxia

S365 Bowel injury

S360 Splenic injury

S361 Liver injury

S350, S351, S352, S353, S355 Aorta/IVC injury + other major vessel injury

S362 Pancreas injury

Non-procedure specific complications

N390, T835 Urinary tract infection

J12, J13, J14, J15, J16, J17, J18, J22 Chest infection

I26 Pulmonary embolus

I60, I61, I62, I63, I64 Stroke

I801, I802, I822 Deep vein thrombosis

I21, I22, I249, I256 Myocardial infarction

N17, N19 Acute renal failure

A09 Infective diarrhoea

K920, K921, K922, K250, K252, K260, K262, K270, K272, K280, K282, K290 Gastrointestinal bleed

T814 Wound infection

K560, K567 Ileus

S273, S276 Lung/pleura injury

S270, J938, J939 Pneumothorax

A list of the codes considered to be a re-intervention:

Operation-specific re-intervention

X339, X332, T283, T302, M478, M136, M193, T309, S424, T301, T413, G693, M459, T303, X338, S423, T453, T414, T462, M198, S571, G699, G731, M062, M498, T452, T458, T554, X333, G784, H151, M132, M195, M292, S624, T308, T315, X342, M162, M479, T318, T463, T468, T556, G712, G758, H159, M215, M418, M472, T342, T343, T418, G738, G742, G788, H335, H468, H628, M151, M203, P298, T253, T289, T488, G633, G698, G702, G714, G723, G728, G741, G743, G755, H059, H113, H152, H158, H161, H176, M022, M063, M069, M118, M119, M135, M138, M168, M182, M221, M258, M264, M302, M373, M378, M382, M388, M768, M832, T282, T288, T316, T348, T361, T374, T384, T451, T459, T461, T469, T549, T838

Non-operation specific re-intervention

L912, L913, G459, E423, G451, L911, G478, L133, T124, X408, G434, H259, S472, X344, X403, X501, Y903, G603, G608, H221, H289, K608, L714, E498, E499, G344, G521, G522, K601, L728, L951, S422, S579, T123, W901, X402, A559, A843, C623, E051, E315, E368, E429, E506, E594, G289, G348, G383, G432, G433, G435, G448, G479, G601, H012, H029, H229, H248, K491, K633, L161, L463, L631, L634, L722, L741, L791, L918, L938, N039, N242, N262, N304, N306, P092, P172, Q552, S421, S428, S429, S432, S476, S478, S578, S608, T121, T122, T223, T278, T393, V331, W471, X291, X318, X368, X388, X502, X508, X558

REFERENCES

1. Haynes AB, Weiser TG, Berry WR, Lipsitz SR, Breizat AH, Dellinger EP, et al. A surgical safety checklist to reduce morbidity and mortality in a global population. *N Engl J Med* 2009;360(5):491-9.

2. Berwick DM, James B, Coye MJ. Connections between quality measurement and improvement. *Med Care* 2003;41(1 Suppl):I30-8.

3. Hibbard JH. What can we say about the impact of public reporting? Inconsistent execution yields variable results. *Ann Intern Med* 2008;148(2):160-1.

4. Fung CH, Lim YW, Mattke S, Damberg C, Shekelle PG. Systematic review: the evidence that publishing patient care performance data improves quality of care. *Ann Intern Med* 2008;148(2):111-23.

5. Department of Health. High quality care for all: NHS Next Stage Review final report. 2008. Available at: http://www.dh.gov.uk/en/Publicationsandstatistics/Publications/PublicationsPolicyAndGuidance/DH_085825. Accessed July 2008.

6. Hibbard JH, Stockard J, Tusler M. It isn't just about choice: the potential of a public performance report to affect the public image of hospitals. *Med Care Res Rev* 2005;62(3):358-71.

7. Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, et al. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med* 2006;355(4):375-84.

8. Rosenthal MB, Landon BE, Normand SL, Frank RG, Epstein AM. Pay for performance in commercial HMOs. *N Engl J Med* 2006;355(18):1895-902.

9. Rosenthal MB, Dudley RA. Pay-for-performance: will the latest payment trend improve care? *JAMA* 2007;297(7):740-4.

10. Lindenauer PK, Remus D, Roman S, Rothberg MB, Benjamin EM, Ma A, et al. Public reporting and pay for performance in hospital quality improvement. *N Engl J Med* 2007;356(5):486-96.

11. Chassin MR, Galvin RW. The urgent need to improve health care quality. Institute of Medicine National Roundtable on Health Care Quality. *Jama* 1998;280(11):1000-5.

12. Quality of care. Council on Medical Service. *Jama* 1986;256(8):1032-4.

13. BUPA Hospitals UK. BUPA Hospitals. Available at: http://www.bupa-hospitals.co.uk/asp/patientcare/quality.asp#2. Accessed July 2007.

14. Donabedian A. Evaluating the Quality of Medical Care. *The Milbank Memorial Fund Quarterly* 1966;44(3):166-203.

15. Donabedian A. The seven pillars of quality. *Arch Pathol Lab Med* 1990;114(11):1115-8.

16. Schiff GD, Rucker TD. Beyond structure-process-outcome: Donabedian's seven pillars and eleven buttresses of quality. *Jt Comm J Qual Improv* 2001;27(3):169-74.

17. National Academies Press. Crossing the Quality Chasm: A New Health System for the 21st Century. 2001. Available at: http://www.iom.edu/Object.File/Master/27/184/Chasm-8pager.pdf Accessed June 2008.

18. Institute for Healthcare Improvement. Whole System Measures. 2007. Available at: www.IHI.org Accessed June 2008.

19. Leatherman ST, Hibbard JH, McGlynn EA. A research agenda to advance quality measurement and improvement. *Med Care* 2003;41(1 Suppl):I80-6.

20. Professor the Lord Darzi of Denham. Evidence-based medicine and the NHS: a commentary. *J R Soc Med* 2008;101(7):342-4.

21. Leatherman S, Sutherland K. Designing national quality reforms: a framework for action. *Int J Qual Health Care* 2007;19(6):334-340.

22. Fiscella K, Franks P, Gold MR, Clancy CM. Inequality in quality: addressing socioeconomic, racial, and ethnic disparities in health care. *JAMA* 2000;283(19):2579-84.

23. What role for health care systems in reducing health inequalities? Available at: http://www.epha.org/a/1907. Accessed March 2008.

24. Thomson R, Taber S, Lally J, Kazandjian V. UK Quality Indicator Project (UK QIP) and the UK independent health care sector: a new development. *Int J Qual Health Care* 2004;16 Suppl 1:i51-6.

25. National Surgical Quality Improvement Program. Available at: https://acsnsqip.org/main/about_history.asp. Accessed July 2007.

26. Veillard J, Champagne F, Klazinga N, Kazandjian V, Arah OA, Guisset AL. A performance assessment framework for hospitals: the WHO regional office for Europe PATH project. *Int J Qual Health Care* 2005;17(6):487-96.

27. Organization for Economic Cooperation and Development. Health Care Quality Indicators. 2003. Available at: http://www.oecd.org/document/34/0,3343,en_2649_33929_37088930_1_1_1_37407,00.html. Accessed June 2008.

28. Khuri SF, Daley J, Henderson WG. The Comparative Assessment and Improvement of Quality of Surgical Care in the Department of Veterans Affairs. *Arch Surg* 2002;137(1):20-27.

29. Groene O, Skau JK, Frolich A. An international review of projects on hospital performance assessment. *Int J Qual Health Care* 2008;20(3):162-71.

30. The Leapfrog Group. Available at: http://www.leapfroggroup.org. Accessed June 2007.

31. Medicare Australia. Pratice Incentives Program (PIP). Available at: http://www.medicareaustralia.gov.au/providers/incentives_allowances/pip.shtml. Accessed July 2007.

32. Pink GH, Brown AD, Studer ML, Reiter KL, Leatt P. Pay-for-performance in publicly financed healthcare: some international experience and considerations for Canada. *Healthc Pap* 2006;6(4):8-26.

33. Petersen LA, Woodard LD, Urech T, Daw C, Sookanan S. Does pay-for-performance improve the quality of health care? *Ann Intern Med* 2006;145(4):265-72.

34. Curtin K, Beckman H, Pankow G, Milillo Y, Green RA. Return on investment in pay for performance: a diabetes case study. *J Healthc Manag* 2006;51(6):365-74; discussion 375-6.

35. Roland M. Pay-for-performance: too much of a good thing? A conversation with Martin Roland. Interview by Robert Galvin. *Health Aff (Millwood)* 2006;25(5):w412-9.

36. National Sentinel Stroke Audit Phase 1 Organisational audit 2008. Stroke Programme. 2008. Available at: http://www.rcplondon.ac.uk/clinical-standards/ceeu/Current-work/Pages/Stroke-programme.aspx Accessed January 2009.

37. National service framework for coronary heart disease. Available at: http://www.dh.gov.uk/en/Healthcare/NationalServiceFrameworks/Coronaryheartdisease/index.htm Accessed January 2009.

38. Department of Health. Cancer Reform Strategy. Available at: http://www.dh.gov.uk/en/Healthcare/NationalServiceFrameworks/Cancer/index.htm Accessed January 2009.

39. Ricciardi R, Virnig BA, Ogilvie JW, Jr., Dahlberg PS, Selker HP, Baxter NN. Volume-outcome relationship for coronary artery bypass grafting in an era of decreasing volume. *Arch Surg* 2008;143(4):338-44; discussion 344.

40. Brook RH, Park RE, Chassin MR, Solomon DH, Keesey J, Kosecoff J. Predicting the appropriate use of carotid endarterectomy, upper gastrointestinal endoscopy, and coronary angiography. *N Engl J Med* 1990;323(17):1173-7.

41. Brook RH, Park RE, Chassin MR, Kosecoff J, Keesey J, Solomon DH. Carotid endarterectomy for elderly patients: predicting complications. *Ann Intern Med* 1990;113(10):747-53.

42. The Leapfrog Group. Evidence-Based Hospital Referral. 2004. Available at: http://www.leapfroggroup.org Accessed June 2007.

43. Department of Health. Improving Outcomes Guidance. Available at: http://www.dh.gov.uk/en/Policyandguidance/Healthandsocialcaretopics/Cancer/DH_4063775. Accessed August 2007.

44. Elixhauser A, Steiner C, Fraser I. Volume thresholds and hospital characteristics in the United States. *Health Aff (Millwood)* 2003;22(2):167-77.

45. Pronovost PJ, Angus DC, Dorman T, Robinson KA, Dremsizov TT, Young TL. Physician staffing patterns and clinical outcomes in critically ill patients: a systematic review. *Jama* 2002;288(17):2151-62.

46. Treggiari MM, Martin DP, Yanez ND, Caldwell E, Hudson LD, Rubenfeld GD. Effect of intensive care unit organizational model and structure on outcomes in patients with acute lung injury. *Am J Respir Crit Care Med* 2007;176(7):685-90.

47. Malin JL, Schneider EC, Epstein AM, Adams J, Emanuel EJ, Kahn KL. Results of the National Initiative for Cancer Care Quality: how can we improve the quality of cancer care in the United States? *J Clin Oncol* 2006;24(4):626-34.

48. Lilford RJ, Brown CA, Nicholl J. Use of process measures to monitor the quality of clinical practice. *Bmj* 2007;335(7621):648-50.

49. Chassin MR, Kosecoff J, Park RE, Winslow CM, Kahn KL, Merrick NJ, et al. Does inappropriate use explain geographic variations in the use of health care services? A study of three procedures. *Jama* 1987;258(18):2533-7.

50. Brook RH. Practice guidelines and practicing medicine. Are they compatible? *Jama* 1989;262(21):3027-30.

51. Jamtvedt G, Young JM, Kristoffersen DT, O'Brien MA, Oxman AD. Audit and feedback: effects on professional practice and health care outcomes. *Cochrane Database Syst Rev* 2006(2):CD000259.

52. US News and World Report - Best Hospitals 2006. Available at: http://www.usnews.com/usnews/health/best-hospitals/tophosp.htm. Accessed October 2007.

53. Dr Foster Intelligence. Dr Foster Good hospital Guide. Available at: http://www.drfoster.co.uk/ghg. Accessed May 2007.

54. The Healthcare Commission. Available at: http://www.healthcarecommission.org.uk/homepage.cfm. Accessed March 2006.

55. Pitches DW, Mohammed MA, Lilford RJ. What is the empirical evidence that hospitals with higher-risk adjusted mortality rates provide poorer quality care? A systematic review of the literature. *BMC Health Serv Res* 2007;7(1):91.

56. Hofer TP, Hayward RA. Identifying poor-quality hospitals. Can hospital mortality rates detect quality problems for medical diagnoses? *Med Care* 1996;34(8):737-53.

57. Zalkind DL, Eastaugh SR. Mortality rates as an indicator of hospital quality. *Hosp Health Serv Adm* 1997;42(1):3-15.

58. Thomas JW, Hofer TP. Accuracy of risk-adjusted mortality rate as a measure of hospital quality of care. *Med Care* 1999;37(1):83-92.

59. National Institute for Health and Clinical Excellence. Update on Herceptin Appraisal. 2006. Available at: http://www.nice.org.uk/page.aspx?o=328396 Accessed June 2007.

60. Healthcare Output and Productivity: Accounting for Quality Change: Department of Health, 2005.

61. Patient Reported Outcome Measures (PROMs). 2008. Available at: http://www.glasgows.co.uk/proms/. Accessed July 2008.

62. National Programme Budget project. 2002. Available at: http://www.dh.gov.uk/en/Managingyourorganisation/Financeandplanning/Programmebudgeting/index.htm. Accessed September 2007.

63. Office for National Statistics: Public service productivity: Health. *Economic Trends* 2006;628:26-57.

64. Atkinson T. Atkinson Review of Government Output and Productivity for the National Accounts: Final Report. *London: HMSO* 2005.

65. Hayes JL, Hansen P. Is laparoscopic colectomy for cancer cost-effective relative to open colectomy? *ANZ J Surg* 2007;77(9):782-6.

66. Topal B, Peeters G, Verbert A, Penninckx F. Outpatient laparoscopic cholecystectomy: clinical pathway implementation is efficient and cost effective and increases hospital bed capacity. *Surg Endosc* 2007;21(7):1142-6.

67. Fisher JA, Fikry C, Troxel AB. Cutting cost and increasing access to colorectal cancer screening: another approach to following the guidelines. *Cancer Epidemiol Biomarkers Prev* 2006;15(1):108-13.

68. Ganz PA. Impact of quality of life outcomes on clinical practice. *Oncology (Williston Park)* 1995;9(11 Suppl):61-5.

69. Patient-Reported Outcome and Quality of life Instruments Database (ProQolid). 2005. Available at: www.proqolid.org. Accessed July 2008.

70. Lipscomb J, Gotay CC, Snyder CF. Patient-reported outcomes in cancer: a review of recent research and policy initiatives. *CA Cancer J Clin* 2007;57(5):278-300.

71. FACIT.org. Functional Assessment of Chronic Illness Therapy. Available at: http://www.facit.org/about/welcome.aspx. Accessed July 2008.

72. Park SM, Park MH, Won JH, Lee KO, Choe WS, Heo DS, et al. EuroQol and survival prediction in terminal cancer patients: a multicenter prospective study in hospice-palliative care units. *Support Care Cancer* 2006;14(4):329-33.

73. Busija L, Osborne RH, Nilsdotter A, Buchbinder R, Roos EM. Magnitude and meaningfulness of change in SF-36 scores in four types of orthopedic surgery. *Health Qual Life Outcomes* 2008;6(1):55.

74. Osoba D. Translating the science of patient-reported outcomes assessment into clinical practice. *J Natl Cancer Inst Monogr* 2007;37:5-11.

75. Ware JE, Jr., Snyder MK, Wright WR, Davies AR. Defining and measuring patient satisfaction with medical care. *Eval Program Plann* 1983;6(3-4):247-63.

76. Carr-Hill RA. The measurement of patient satisfaction. *J Public Health Med* 1992;14(3):236-49.

77. Shahian DM, Edwards FH, Ferraris VA, Haan CK, Rich JB, Normand SL, et al. Quality measurement in adult cardiac surgery: part 1--Conceptual framework and measure selection. *Ann Thorac Surg* 2007;83(4 Suppl):S3-12.

78. O'Brien SM, Shahian DM, DeLong ER, Normand SL, Edwards FH, Ferraris VA, et al. Quality measurement in adult cardiac surgery: part 2--Statistical considerations in composite measure scoring and provider rating. *Ann Thorac Surg* 2007;83(4 Suppl):S13-26.

79. Birkmeyer JD, Birkmeyer CM, Wennberg DE, Young MP. Leapfrog patient safety standards: the potential benefits of universal adoption. Washington,: Leapfrog Group, 2000.

80. Leatherman S, Sutherland K. Designing national quality reforms: a framework for action. *Int J Qual Health Care* 2007.

81. Glickman SW, Baggett KA, Krubert CG, Peterson ED, Schulman KA. Promoting quality: the health-care organization from a management perspective. *Int J Qual Health Care* 2007.

82. Luft HS, Bunker JP, Enthoven AC. Should operations be regionalized? The empirical relation between surgical volume and mortality. *N Engl J Med* 1979;301(25):1364-1369.

83. Care Quality Commission. Heart Surgery in the United Kingdom. Available at: http://heartsurgery.cqc.org.uk/about-this-site.aspx. Accessed December 2009.

84. Center for Medical Consumers. Available at: www.medicalconsumers.org. Accessed October 2007.

85. Halm E, Lee C, Chassin M. How Is Volume Related to Quality in Health Care? A Systematic Review of the Research Literature. Appendix C, 27-102. In: Hewitt M (ed.), Interpreting the Volume-Outcome Relationship in the Context of Health Care Quality: Workshop Summary; 2000. Institute of Medicine, Washington DC. National Academic Press.

86. Hong D, Tandan VR, Goldsmith CH, Simunovic M. Users' guide to the surgical literature: how to use an article reporting population-based volume-outcome relationships in surgery. *Can J Surg* 2002;45(2):109-15.

87. Aylin P, Bottle A, Majeed A. Use of administrative data or clinical databases as predictors of risk of death in hospital: comparison of models. *Bmj* 2007;334(7602):1044.

88. Christian CK, Gustafson ML, Betensky RA, Daley J, Zinner MJ. The volume-outcome relationship: don't believe everything you see. *World J Surg* 2005;29(10):1241-4.

89. Spiegelhalter DJ. Funnel plots for comparing institutional performance. *Stat Med* 2005;24(8):1185-202.

90. Hewitt M (ed.). Interpreting the Volume-Outcome Relationship in the Context of Health Care Quality: Workshop Summary; 2000. Institute of Medicine, Washington DC. National Academic Press.

91. Killeen SD, O'Sullivan MJ, Coffey JC, Kirwan WO, Redmond HP. Provider volume and outcomes for oncological procedures. *Br J Surg* 2005;92(4):389-402.

92. Panageas KS, Schrag D, Riedel E, Bach PB, Begg CB. The effect of clustering of outcomes on the association of procedure volume and surgical outcomes. *Ann Intern Med* 2003;139(8):658-65.

93. Begg CB, Riedel ER, Bach PB, Kattan MW, Schrag D, Warren JL, et al. Variations in morbidity after radical prostatectomy. *N Engl J Med* 2002;346(15):1138-44.

94. Bianco FJ, Jr., Riedel ER, Begg CB, Kattan MW, Scardino PT. Variations among high volume surgeons in the rate of complications after radical prostatectomy: further evidence that technique matters. *J Urol* 2005;173(6):2099-103.

95. Roohan PJ, Bickell NA, Baptiste MS, Therriault GD, Ferrara EP, Siu AL. Hospital volume differences and five-year survival from breast cancer. *Am J Public Health* 1998;88(3):454-7.

96. Halm EA, Lee C, Chassin MR. Is volume related to outcome in health care? A systematic review and methodologic critique of the literature. *Ann Intern Med* 2002;137(6):511-20.

97. Goodney PP, Stukel TA, Lucas FL, Finlayson EV, Birkmeyer JD. Hospital volume, length of stay, and readmission rates in high-risk surgery. *Ann Surg* 2003;238(2):161-7.

98. Mason A, Goldacre MJ, Bettley G, Vale J, Joyce A. Using routine data to define clinical case-mix and compare hospital outcomes in urology. *BJU Int* 2006;97(6):1145-7.

99. Birkmeyer JD, Stukel TA, Siewers AE, Goodney PP, Wennberg DE, Lucas FL. Surgeon volume and operative mortality in the United States. *N Engl J Med* 2003;349(22):2117-27.

100. Schrag D, Panageas KS, Riedel E, Cramer LD, Guillem JG, Bach PB, et al. Hospital and surgeon procedure volume as predictors of outcome following rectal cancer resection. *Ann Surg* 2002;236(5):583-92.

101. Hillner BE, Smith TJ, Desch CE. Hospital and physician volume or specialization and outcomes in cancer treatment: importance in quality of cancer care. *J Clin Oncol* 2000;18(11):2327-40.

102. Harmon JW, Tang DG, Gordon TA, Bowman HM, Choti MA, Kaufman HS, et al. Hospital volume can serve as a surrogate for surgeon volume for achieving excellent outcomes in colorectal resection. *Ann Surg* 1999;230(3):404-11; discussion 411-3.

103. Schrag D, Panageas KS, Riedel E, Hsieh L, Bach PB, Guillem JG, et al. Surgeon volume compared to hospital volume as a predictor of outcome following primary colon cancer resection. *J Surg Oncol* 2003;83(2):68-78; discussion 78-9.

104. Thiemann DR, Coresh J, Oetgen WJ, Powe NR. The association between hospital volume and survival after acute myocardial infarction in elderly patients. *N Engl J Med* 1999;340(21):1640-8.

105. Konety BR, Dhawan V, Allareddy V, Joslyn SA. Impact of hospital and surgeon volume on in-hospital mortality from radical cystectomy: data from the health care utilization project. *J Urol* 2005;173(5):1695-700.

106. McCabe JE, Jibawi A, Javle P. Defining the minimum hospital case-load to achieve optimum outcomes in radical cystectomy. *BJU Int* 2005;96(6):806-10.

107. Hollenbeck BK, Wei Y, Birkmeyer JD. Volume, process of care, and operative mortality for cystectomy for bladder cancer. *Urology* 2007;69(5):871-5.

108. Elting LS, Pettaway C, Bekele BN, Grossman HB, Cooksley C, Avritscher EB, et al. Correlation between annual volume of cystectomy, professional staffing, and outcomes: a statewide, population-based study. *Cancer* 2005;104(5):975-84.

109. Nelson CP, Dunn RL, Wei JT, Gearhart JP. Surgical repair of bladder exstrophy in the modern era: contemporary practice patterns and the role of hospital case volume. *J Urol* 2005;174(3):1099-102.

110. Ellison LM, Trock BJ, Poe NR, Partin AW. The effect of hospital volume on cancer control after radical prostatectomy. *J Urol* 2005;173(6):2094-8.

111. Eastham JA, Kattan MW, Riedel E, Begg CB, Wheeler TM, Gerigk C, et al. Variations among individual surgeons in the rate of positive surgical margins in radical prostatectomy specimens. *J Urol* 2003;170(6 Pt 1):2292-5.

112. Yao SL, Lu-Yao G. Population-based study of relationships between hospital volume of prostatectomies, patient outcomes, and length of hospital stay. *J Natl Cancer Inst* 1999;91(22):1950-6.

113. Hu JC, Gold KF, Pashos CL, Mehta SS, Litwin MS. Role of surgeon volume in radical prostatectomy outcomes. *J Clin Oncol* 2003;21(3):401-5.

114. Davenport K, Timoney AG, Keeley FX, Joyce A, Downey P. A 3-year review of The British Association of Urological Surgeons Section of Endourology Laparoscopic Nephrectomy Audit. *BJU Int* 2006;97(2):333-7.

115. Taub DA, Miller DC, Cowan JA, Dimick JB, Montie JE, Wei JT. Impact of surgical volume on mortality and length of stay after nephrectomy. *Urology* 2004;63(5):862-7.

116. Konety BR, Allareddy V, Modak S, Smith B. Mortality after major surgery for urologic cancers in specialized urology hospitals: are they any better? *J Clin Oncol* 2006;24(13):2006-12.

117. Thorpe AC, Cleary R, Coles J, Vernon S, Reynolds J, Neal DE. Deaths and complications following prostatectomy in 1400 men in the northern region of England. Northern Regional Prostate Audit Group. *Br J Urol* 1994;74(5):559-65.

118. Konety BR, Allareddy V, Herr H. Complications after radical cystectomy: analysis of population-based data. *Urology* 2006;68(1):58-64.

119. Chun FK, Briganti A, Antebi E, Graefen M, Currlin E, Steuber T, et al. Surgical volume is related to the rate of positive surgical margins at radical prostatectomy in European patients. *BJU Int* 2006;98(6):1204-9.

120. Hollenbeck BK, Daignault S, Dunn RL, Gilbert S, Weizer AZ, Miller DC. Getting under the hood of the volume-outcome relationship for radical cystectomy. *J Urol* 2007;177(6):2095-9; discussion 2099.

121. Judge A, Evans S, Gunnell DJ, Albertsen PC, Verne J, Martin RM. Patient outcomes and length of hospital stay after radical prostatectomy for prostate cancer: analysis of hospital episodes statistics for England. *BJU Int* 2007;100(5):1040-9.

122. Callcut RA, Breslin TM. Shaping the future of surgery: the role of private regulation in determining quality standards. *Ann Surg* 2006;243(3):304-12.

123. Heart Surgery in Great Britain. 2006. Available at: http://heartsurgery.healthcarecommission.org.uk/index.aspx. Accessed December 2006.

124. Tekkis PP, McCulloch P, Steger AC, Benjamin IS, Poloniecki JD. Mortality control charts for comparing performance of surgical units: validation study using hospital mortality data. *Bmj* 2003;326(7393):786-8.

125. Marshall T, Mohammed MA, Rouse A. A randomized controlled trial of league tables and control charts as aids to health service decision-making. *Int J Qual Health Care* 2004;16(4):309-15.

126. Benneyan JC, Lloyd RC, Plsek PE. Statistical process control as a tool for research and healthcare improvement. *Qual. Saf. Health Care* 2003;12(6):458-464.

127. The Information Centre. Available at: http://www.hesonline.nhs.uk. Accessed April 2007.

128. Fourth National Vascular Dataset Report. The Vascular Society of Great Britain and Ireland. 2004. Available at: http://www.vascularsociety.org.uk/committees/audit.asp. Accessed May 2007.

129. Kunadian B, Dunning J, Roberts AP, Morley R, Twomey D, Hall JA, et al. Cumulative funnel plots for the early detection of interoperator variation: retrospective database analysis of observed versus predicted results of percutaneous coronary intervention. *BMJ* 2008;26(336(7650)):931-4.

130. Healthcare Commission. Performance Ratings. 2005. Available at: http://ratings2005.healthcarecommission.org.uk/. Accessed July 2007.

131. Mason A, Goldacre M, Meddings D, Woolfson J. Use of case fatality and readmission measures to compare hospital performance in gynaecology. *Bjog* 2006;113(6):695-9.

132. Poloniecki J, Valencia O, Littlejohns P. Cumulative risk adjusted mortality chart for detecting changes in death rate: observational study of heart surgery. *Bmj* 1998;316(7146):1697-700.

133. Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics* 2000;1(4):441-52.

134. Grigg OA, Farewell VT, Spiegelhalter DJ. Use of risk-adjusted CUSUM and RSPRT charts for monitoring in medical contexts. *Stat Methods Med Res* 2003;12(2):147-70.

135. Department of Health. Hospital Activity Statistics - Diagnostics Statistics. Available at: http://www.performance.doh.gov.uk/diagnostics/index.htm. Accessed July 2007.

136. Spiegelhalter DJ. Handling over-dispersion of performance indicators. *Qual Saf Health Care* 2005;14(5):347-51.

137. Houghton A. Variation in outcome of surgical procedures. *Br J Surg* 1994;81(5):653-60.

138. Birkmeyer JD, Finlayson EV, Birkmeyer CM. Volume standards for high-risk surgical procedures: potential benefits of the Leapfrog initiative. *Surgery* 2001;130(3):415-22.

139. Birkmeyer JD, Dimick JB. Potential benefits of the new Leapfrog standards: effect of process and outcomes measures. *Surgery* 2004;135(6):569-75.

140. The Bristol Royal Infirmary Inquiry, Crown Copyright. The Inquiry into the management of care of children receiving complex heart surgery at the Bristol Royal Infirmary. 2001. Available at: http://www.bristol-inquiry.org.uk/final_report/index.htm. Accessed September 2007.

141. Khuri SF, Daley J, Henderson W, Hur K, Hossain M, Soybel D, et al. Relation of surgical volume to outcome in eight common operations: results from the VA National Surgical Quality Improvement Program. *Ann Surg* 1999;230(3):414-29; discussion 429-32.

142. Egger M, Smith GD. Bias in location and selection of studies. *BMJ* 1998;316(7124):61-6.

143. Dickersin K, Min YI, Meinert CL. Factors influencing publication of research results. Follow-up of applications submitted to two institutional review boards. *JAMA* 1992;267(3):374-378.

144. Easterbrook PJ, Gopalan R, Berlin JA, Matthews DR. Publication bias in clinical research. *The Lancet* 1991;337(8746):867-872.

145. Sutton AJ, Duval SJ, Tweedie RL, Abrams KR, Jones DR. Empirical assessment of effect of publication bias on meta-analyses. *Bmj* 2000;320(7249):1574-7.

146. Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315(7109):629-34.

147. Tang JL, Liu JL. Misleading funnel plot for detection of bias in meta-analysis. *J Clin Epidemiol* 2000;53(5):477-84.

148. Sterne JA, Egger M. Funnel plots for detecting bias in meta-analysis: guidelines on choice of axis. *J Clin Epidemiol* 2001;54(10):1046-55.

149. Lau J, Ioannidis JP, Terrin N, Schmid CH, Olkin I. The case of the misleading funnel plot. *Bmj* 2006;333(7568):597-600.

150. Association of Public Health Observatories. Indicators of Public Health in the English Regions. 2004. Available at: http://www.apho.org.uk/resource/view.aspx?QN=IND_SERIES. Accessed January 2008.

151. Department of Health. The NHS Plan: a plan for investment – a plan for reform. 2000. Available at: http://www.dh.gov.uk/assetRoot/04/05/57/83/04055783.pdf. Accessed April 2007.

152. Department of Health. National standards, local action: health and social care standards and planning framework 2005/06-2007/08. 2004. Available at: www.dh.gov.uk/assetRoot/04/08/60/58/04086058.pdf. Accessed May 2007.

153. Department of Health. The NHS Cancer Plan. 2000. Available at: http://www.dh.gov.uk/assetRoot/04/01/45/13/04014513.pdf. Accessed February 2007.

154. National Patient Safety Agency. Clean Your Hands Campaign. Available at: http://www.npsa.nhs.uk/cleanyourhands. Accessed 1st May 2007.

155. Healthcare Commission. MRSA Bacteraemia. 2007/2008. Available at: http://www.healthcarecommission.org.uk/nationaltargets2007-2008/newnationaltargets/acuteandspecialisttrusts/indicators/mrsabacteraemia.cfm. Accessed December 2008.

156. Care Quality Commission. MRSA Bacteraemias. 2009/2010. Available at: http://www.cqc.org.uk/guidanceforprofessionals/healthcare/nhsstaff/periodicreview2009/10/qualityofservices/existing/incidenceofmrsabacteraemiaacsp.cfm. Accessed December 2009.

157. Spiegelhalter DJ. Problems in assessing rates of infection with methicillin resistant Staphylococcus aureus. *Bmj* 2005;331(7523):1013-5.

158. Association of Public Health Observatories. Available at: http://www.apho.org.uk/. Accessed March 2008.

159. South West Public Health Observatory. 3-year Funnel plots for PCTs in South West and Hampshire and Isle of Wight 2001-2003. Available at: http://www.swpho.nhs.uk/resource/item.aspx?RID=9400. Accessed May 2007.

160. Halm E, Lee C, MR C. Interpreting How Is Volume Related to Quality in Health Care? A Systematic Review of the Research Literature. Appendix C, 27-102. In: Hewitt M (ed.), Interpreting the Volume-Outcome Relationship in the Context of Health Care Quality: Workshop Summary; 2000. Institute of Medicine, Washington DC. National Academic Press

161. National Institute for Health and Clinical Excellence. Guidance on Cancer Services, Improving Outcomes in Urological Cancers, The Manual. 2002. Available at: http://www.nice.org.uk/nicemedia/pdf/Urological_Manual.pdf. Accessed March 2005.

162. Ellison LM, Heaney JA, Birkmeyer JD. The effect of hospital volume on mortality and resource use after radical prostatectomy. *J Urol* 2000;163(3):867-9.

163. Karakiewicz PI, Bazinet M, Aprikian AG, Tanguay S, Elhilali MM. Thirty-day mortality rates and cumulative survival after radical retropubic prostatectomy. *Urology* 1998;52(6):1041-6.

164. Hollenbeck BK, Dunn RL, Miller DC, Daignault S, Taub DA, Wei JT. Volume-based referral for cancer surgery: informing the debate. *J Clin Oncol* 2007;25(1):91-6.

165. Birkmeyer JD, Siewers AE, Finlayson EV, Stukel TA, Lucas FL, Batista I, et al. Hospital volume and surgical mortality in the United States. *N Engl J Med* 2002;346(15):1128-37.

166. Deeks JJ H, JPT, Altman DG. Analysing and presenting results. In: Higgins JPT, Green S, editors. Cochrane Handbook for Systematic Reviews of Interventions 4.2.6 [updated September 2006]; Section 8. 2006.

167. Van Poppel H, Collette L, Kirkali Z, Brausi M, Hoekstra W, Newling DW, et al. Quality control of radical prostatectomy: a feasibility study. *Eur J Cancer* 2001;37(7):884-91.

168. Nuttall M, van der Meulen J, Phillips N, Sharpin C, Gillatt D, McIntosh G, et al. A systematic review and critique of the literature relating hospital or surgeon volume to health outcomes for 3 urological cancer procedures. *J Urol* 2004;172(6 Pt 1):2145-52.

169. Joudi FN, Konety BR. The impact of provider volume on outcomes from urological cancer therapy. *J Urol* 2005;174(2):432-8.

170. Sundararajan V, Henderson T, Perry C, Muggivan A, Quan H, Ghali WA. New ICD-10 version of the Charlson comorbidity index predicted in-hospital mortality. *J Clin Epidemiol* 2004;57(12):1288-94.

171. Office for National Statistics Gridlink. Available at: http://www.statistics.gov.uk/geography/gridlink.asp. Accessed December 2007.

172. Black N, Barker M, Payne M. Cross sectional survey of multicentre clinical databases in the United Kingdom. *BMJ* 2004;328(7454):1478.

173. Morris E, Quirke P, Thomas JD, Fairley L, Cottier B, Forman D. Unacceptable variation in abdominoperineal excision rates for rectal cancer - time to intervene? *Gut* 2008.

174. PbR data assurance framework 2008/09. Key messages from Year 2 of the national clinical coding audit programme. 2009. Available at: http://www.audit-commission.gov.uk/SiteCollectionDocuments/AuditCommissionReports/NationalStudies/20090827pbrdataassuranceframework0809rep.pdf Accessed August 2009.

175. Information and data quality in the NHS. Key messages from three years of independent review. 2004. Available at: www.auditcommission.gov.uk/reports/NATIONAL-REPORT.asp?CategoryID=&ProdID=4D598AF6-3894-401d-AA48-1076125DA38D&SectionID=toc#. Accessed April 2008.

176. Hollenbeck BK, Miller DC, Taub DA, Dunn RL, Khuri SF, Henderson WG, et al. The effects of adjusting for case mix on mortality and length of stay following radical cystectomy. *J Urol* 2006;176(4 Pt 1):1363-8.

177. Sheldon TA. The volume-quality relationship: insufficient evidence for use as a quality indicator. *Qual Saf Health Care* 2004;13(5):325-6.

178. PbR Data Assurance Framework 2007/08. Findings from the first year of the national clinical coding audit programme. 2008. Available at: http://www.audit-commission.gov.uk/SiteCollectionDocuments/AuditCommissionReports/NationalStudies/PbRreport.pdf Accessed February 2009.

179. Barbieri CE, Lee B, Cookson MS, Bingham J, Clark PE, Smith JA, Jr., et al. Association of procedure volume with radical cystectomy outcomes in a nationwide database. *J Urol* 2007;178(4 Pt 1):1418-21; discussion 1421-2.

180. Eastern Region Public Health Observatory. Available at: http://www.erpho.org.uk/. Accessed May 2006.

181. Klein EA, Bianco FJ, Serio AM, Eastham JA, Kattan MW, Pontes JE, et al. Surgeon Experience is Strongly Associated With Biochemical Recurrence After Radical Prostatectomy for All Preoperative Risk Categories. *J Urol* 2008;179(6):2212-6.

182. Rosser CJ, Kamat AM, Pendleton J, Robinson TL, Pisters LL, Swanson DA, et al. Impact of fellowship training on pathologic outcomes and complication rates of radical prostatectomy. *Cancer* 2006;107(1):54-9.

183. Schulman J, Spiegelhalter DJ, Parry G. How to interpret your dot: decoding the message of clinical performance indicators. *J Perinatol* 2008;28(9):588-96.

184. Eastern Region Public Health Observatory. Catchment areas and populations. 2003. Available at: http://www.erpho.org.uk/Download/Public/9480/1/INPHORM%202%20Catchment%20areas%20and%20populations[1].pdf. Accessed February 2007.

185. National Institute for Health and Clinical Excellence. Improving Outcomes in Urological Cancers - Research Evidence. 2002. Available at: http://www.nice.org.uk/nicemedia/live/10889/28772/28772.pdf Accessed October 2009.

186. Gilbert SM, Dunn RL, Miller DC, Daignault S, Ye Z, Hollenbeck BK. Mortality after urologic cancer surgery: impact of non-index case volume. *Urology* 2008;71(5):906-10.

187. Fradet V, Allareddy V, Aaronson DS, Konety BR. Is the complication rate of radical cystectomy predictive of the complication rate of other urological procedures? *J Urol* 2009;181(3):1054-9; discussion 1059-60.

188. Nuttall MC, van der Meulen J, McIntosh G, Gillatt D, Emberton M. Threshold volumes for urological cancer surgery: a survey of UK urologists. *BJU Int* 2004;94(7):1010-3.

189. Sowden AJ, Sheldon TA. Does volume really affect outcome? Lessons from the evidence. *J Health Serv Res Policy* 1998;3(3):187-90.

190. Hibbard JH, Stockard J, Tusler M. Does publicizing hospital performance stimulate quality improvement efforts? *Health Aff (Millwood)* 2003;22(2):84-94.

191. Dr Foster Intelligence. Hospital Guide 2008. Available at: http://www.drfosterintelligence.co.uk/newsPublications/localDocuments/hospitalGuide2008.pdf Accessed February 2009.