Essays on Health Outcomes and Physician Practice Variation within a Public Single Hospital: the Case of Malta

CARL CAMILLERI

CITY UNIVERSITY LONDON

DEPARTMENT OF ECONOMICS

A DISSERATION SUBMITTED FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

AT CITY UNIVERSITY LONDON

DECEMBER 2015

Table of Contents

1 Introduction....................17

2 The Appropriateness of a DRG Casemix System in the Maltese Hospital Setting....................26

2.1 Introduction and Motivation....................29

2.2 Literature....................32

2.3 The situation in Malta....................38

2.4 Data....................40

2.5 Methodology....................43

2.6 Results....................50

2.6.1 General description of hospital output....................50

2.6.2 Statistical Analysis of Goodness of Fit....................55

2.7 Discussion....................60

2.8 Conclusion....................62

References....................66

Appendix....................70

3 Volume and Competing Risk Outcomes under Different Consultant Contract Types: the Case of Malta 71

3.1 Introduction and Motivation 74

3.2 Literature 78

3.3 Data 87

3.3.1 Measures for outcome indicators 89

3.3.2 Measures of volume 90

3.3.3 Description of data and preliminary evidence 90

3.3.4 Data descriptive statistics at the surgeon level 91

3.3.5 Data descriptive statistics at the consultant level 92

3.4 Methodology 94

3.4.1 Modelling based on cause-specific hazards 96

3.4.2 Modelling the cumulative incidence function 98

3.5 Results and Discussion 101

3.5.1 Non-parametric estimation 102

3.5.1.1 The Cause Specific Hazard Function 102

3.5.1.2 The Cumulative Incidence Function 104

3.5.2 Semi-parametric estimation: the CSHF and the CIF 107

3.5.2.1 Modelling the CSHF 107

3.5.2.2 Modelling the CIF 114

3.6 Conclusion 122

References 127

Appendix 136

4 Modelling Consultant Practice Heterogeneity arising from Contract Type within the Context of an Acute General Public Hospital ... 144

4.1 Introduction and Motivation ... 146

4.2 Literature ... 150

4.3 Theoretical model ... 163

4.4 Empirical approach ... 165

4.5 Data and variables ... 172

4.6 Results and Discussion ... 180

4.6.1 First stage regression ... 180

4.6.2 Second stage regression – all investigations ... 182

4.6.3 Second stage regression – with dated investigation categories ... 183

4.6.4 Second stage regression – all investigations with interaction term and marginal effects ... 185

4.6.5 Second stage regression - with dated investigation categories – with interactions and marginal effects ... 189

4.7 Conclusion ... 195

References ... 202

Appendix ... 210

5 Conclusions and Recommendations ... 225

List of Tables

Table 2.1: DRGs with the 10 most common occurrences for the 2009-2011 period 51

Table 2.2: ALOS (days) after removing outlier cases 52

Table 2.3: MDCs with the 10 most common occurrences 53

Table 2.4: Analysis of outlier cases by different trimming methods 55

Table 2.5: Distribution of the DRGs with CV>1 56

Table 2.6: R^2 under different trimming methods 57

Table 2.7: R^2 by MDC categories under different trimming scenarios 59

Table 2.8: A description of the MDC categories 70

Table 3.1: Definition and description of the principal variables 88

Table 3.2: Characteristics of PTCA activity 91

Table 3.3: Surgeons undertaking PTCA procedure (2009-2011) 92

Table 3.4: Consultants responsible for the PTCA procedure (2009-2011) 93

Table 3.5: Log rank test for equality of the cause specific hazard 104

Table 3.6: Pepe and Mori (1993) test comparing the cumulative incidence for the two groups: consultant and surgeon volume levels 106

Table 3.7: Cause Specific Hazard at particular event of interest (countc and counts) 108

Table 3.8: Cause Specific Hazard at particular event of interest (vol90c and vol90s) 109

Table 3.9: Cause Specific Hazard at particular event of interest (cvol and svol) 109

Table 3.10: Cause Specific Hazard at particular event of interest – with interaction terms 111

Table 3.11: Cumulative Incidence Function at particular events of interest (countc and counts) 116

Table 3.12: Cumulative Incidence Function at particular events of interest (vol90c and vol90s) 116

Table 3.13: Cumulative Incidence Function at particular events of interest (cvol and svol) 116

Table 3.14: Cumulative Incidence Function at particular events of interest – with interaction terms 117

Table 3.15: Mulitnomial logistic regression (countc and counts) – odds ratio 140

Table 3.16: Mulitnomial logistic regression (vol90c and vol90s) - odds ratio 140

Table 3.17: Mulitnomial logistic regression (cvol and svol) – odds ratio 140

Table 3.18: Mulitnomial logistic regression output (with interactions) 142

Table 3.19: Proportional Hazard test based on the 'estat phtest' command in Stata 143

Table 4.1: Definitions and sample statistics of the main variables 172

Table 4.2: Descriptive statistics for variables of interest by consultant job plan...176

Table 4.3: First stage regression results 181

Table 4.4: Second stage regression results – all investigations 183

Table 4.5: Second stage regression results–investigations (<2 and >2 days)...... 184

Table 4.6: Second stage regression results–all investigations–with interactions..186

Table 4.7: Marginal effects (all investigations) 187

Table 4.8: 2nd stage regression results – investigations with interactions (<2 and >2 days) 190

Table 4.9: Marginal effects – investigations (<2 and >2 days) 191

Table 4.10: Differences in practice variation levels – all investigations 210

Table 4.11: Differences in practice variation levels – investigations <2 days 211

Table 4.12: Differences in practice variation levels – investigations >2 days 212

Table 4.13: Main investigations ordered for patients undergoing PTCA treatment 213

Table 4.14: One-stage multilevel model results with 'inv' as the dependent variable 217

Table 4.15: One-stage multilevel model results with 'invf2d' as the dependent variable 220

Table 4.16: One-stage multilevel model results with 'inva2d' as the dependent variable 223

List of Figures

Figure 2-1: Frequency distribution by LOS (less than 31 days)........................... 53

Figure 2-2: Frequency distribution by MDC groups 54

Figure 3-1: Hazard for event *failure* 102

Figure 3-2: Hazard for event *death* 102

Figure 3-3: Hazard for event *readmission*.. 102

Figure 3-4: By consultant volume levels (CSHF) 103

Figure 3-5: By surgeon volume levels (CSHF)... 103

Figure 3-6: By consultant by job plan contract type (CSHF).............................. 103

Figure 3-7: CIF for event failure 105

Figure 3-8: CIF for event death 105

Figure 3-9: CIF for event readmission 105

Figure 3-10: CIF by consultant volume 106

Figure 3-11: CIF by surgeon volume 106

Figure 3-12: CIF by consultant by job plan contract type 106

Figure 3-13: The hazard function of failing from the procedure for different volume levels for consultants on *public-only* contracts 113

Figure 3-14: The hazard function of failing from the procedure for different volume levels for consultants on *dual practice* contracts 113

Figure 3-15: Schoenfeld residuals 120

Figure 3-16: Schoenfeld residuals with time varying coefficients ...., 120

Figure 3-17: Volume variation for consultants on *public-only* contracts ………..... 121

Figure 3-18: Volume variation for consultants on *dual practice* contracts ……….. 121

Figure 3-19: Cause Specific Hazard Function – for event readmission …………….. 136

Figure 3-20: Cause Specific Hazard Function – for event death ……………………… 136

Figure 3-21: Cumulative Incidence Function – for event readmission …………... 137

Figure 3-22: Cumulative Incidence Function – for event death ……………………… 137

Figure 4-1: Variation in number of total investigations by consultant (left) and by consultant job plan contract type (right) …………………………….. 178

Figure 4-2: Variation in number of investigations for all consultants (left) and by consultant job plan type (right) for the first 2 days of hospital stay ……………………………………………………………………………………… 178

Figure 4-3: Variation in number of investigations for all by consultants (left) and by consultant job plan type (right) after the first 2 days of hospital stay……………………………………………………………………………………..... 179

Figure 4-4: Differences in practice variation levels (*inv*) between *public-only* and *dual practice* contracts for different volume indicators at different volume values ……………………………………………………………………......... 188

Figure 4-5: Differences in practice variation levels (*invf2d*) between *public-only* and *dual practice* contracts for different volume indicators at different volume values …………………………………………………………………….... 193

Figure 4-6: Differences in practice variation levels (*inva2d*) between *public-only* and *dual practice* contracts for different volume indicators at different volume values ……………………………………………………….. 194

Acknowledgments

The writing of this thesis would not have been possible without the help, guidance and support of many colleagues and friends.

Firstly, I would like to thank my supervisors, Prof. Mireia Jofre-Bonet and Dr Victoria Serra-Sastre. They have always been there when I needed their help and they have supported and encouraged me all the way by offering continual guidance and advice. I am genuinely grateful for the opportunity they have given me to learn from them and for the time they have dedicated over the course of this study.

This research project started with the help of two other professors whom I would like to mention: Prof. David Parkin and Prof. Nancy Devlin. From the first few months when I started this work (back in 2008), both have offered many words of advice and encouragement on how to succeed in this process.

I would also like to thank the University of Malta for providing a bursary to finance my studies and offering me the flexibility to conduct my research work while maintaining my lecturing post at the Department of Economics.

I am also indebted to all my colleagues and friends at the Department of Economics at the University of Malta. You have all been there to hear and share in my anxieties and frustrations. A special mention goes to Dr Saviour Gauci (ex-Dean of the Faculty) who provided support and encouragement in the early years of this work and to Prof. Joseph Falzon (the current Dean of the Faculty), who has ensured that this work has proceeded without unnecessary hiccup. A very special thanks goes to the current head of the Department of Economics, Dr Philip Von Brockdorff. His unceasing support during the last few months of this work and his dedication to the development of the Department of Economics and its staff members have been an example to all.

A special thanks goes to the staff at the Ministry of Health (Malta), especially at the Mater Dei Hospital. I would like to mention the support and help I received from Mr Alan Dimech who has been a constant help in providing the data for this work. There is no doubt that without his help and guidance this project would not have come to a conclusion and most probably would not even have started! A special mention also goes to Dr Hugo Agius Muscat and Dr Neville Calleja who were able to explain some very complex hospital procedures in a way that I could understand.

I would like to thank all my family and friends who have been of constant support throughout this long journey. You have all provided much of your precious time and energy to help me and my family over these years. Your contribution to this work has been immense and cannot be measured. I am indebted to you all.

A special word goes to my wife, Vanessa.

Vanessa, you have been there when it all started, you have been through all the ups and downs and have had to live with the uncertainty of where we will be when this research came to an end. Your patience, understanding and dedication to our family over the last few years have been simply amazing.

Jan and Jacques (my two boys) have also participated and contributed in this research process. Whilst appreciating that I might have been somewhat of a boring "daddy" to live with over the last few years, both have coped well with my relentless PhD work and travel commitments and both have supported me in their own particular and little ways.

This thesis is dedicated to the most important people in my life – Vanessa and my boys!

Declaration

I grant powers of discretion to the University Librarian to allow this thesis to be copied in whole or in part without any further reference to me. This permission covers only single copies made for study purposes, subject to normal conditions of acknowledgment.

Abstract

This thesis is about the measurement of health care output and the relationship between health care outcomes, physician practice patterns and individual physician characteristics within a very specific and particular health care sector, the health care sector on the Islands of Malta.

Chapter 2 focuses on the appropriateness of introducing a Diagnosis Related Group (DRG) casemix classification system on Maltese data. A number of tests are applied to gauge the ability of Grouper software to capture the heterogeneity between the obtained DRG groups and the degree of homogeneity gained in explaining resource use from the grouping of cases by DRG categories. This serves to provide a measure of health care 'output' whilst providing a tool to help describe and manage resource use. Chapter 3 of this thesis explores differences in the expected relationship between volume and competing risk outcomes and whether this relationship varies in view of different consultant job contract conditions. Finally, Chapter 4 of this thesis studies the behaviour of individual consultants working in the context of the specific incentives and work practices of the Maltese health care system. The role of the specific consultant job contract type is investigated to explain heterogeneity arising among treatment practice patterns over two specific periods related to the patients' stay at the hospital: the first two days of hospital stay and their remaining stay.

Abbreviations

ALOS	Average Length of Stay
AP-DRG	All Patient DRG
AR-DRG	Australian Refined DRG
CABG	Coronary Artery Bypass Graft
CIF	Cumulative Incidence Function
CMS	Centers for Medicare and Medicaid Services
CPU	Clinical Performance Unit
CSHF	Cause Specific Hazard Function
CV	Coefficient of Variation
DRG	Diagnosis Related Group
EDV	Estimated Dependent Variable
FOC	First Order Conditions
GEM	General Equivalence Mapping
HRG	Healthcare Resource Group
ICD-10	International Classification of Diseases – 10th revision
ICD-9-CM	International Classification of Diseases – 9th revision – Clinical Modification
KM	Kaplan Meier
LOS	Length of Stay
MDC	Major Diagnostic Categories
MDH	Mater Dei Hospital
MS-DRG	Medical Severity Diagnosis Related Group
NHS	National Health Service

PCI	Percutaneous Coronary Intervention
PTCA	Percutaneous Transluminal Coronary Angioplasty
WHO	World Health Organisation

Chapter

1

Introduction

This thesis is a compilation of three research papers that investigate differences in hospital outcomes and physician practice variation patterns observed within a very specific and particular health care sector. The focus is the activity undertaken within the only publically funded general acute hospital on the Islands of Malta. Whilst acknowledging the importance of differences due to patient, physician and environmental characteristics that can help to explain variations in outcome, particular focus is given to variations arising from individual consultant and surgeon practice related characteristics, especially job volume levels and consultant job contract conditions. Indeed, these particular characteristics of the health care sector are expected to influence and define the relationships between health care providers, the hospital and the patients.

The thesis focuses on the measurement of output within the Maltese health care sector with an emphasis on the factors which are expected to affect hospital outcomes and practice variation among consultants within this hospital setting. The analysis aims to assess the feasibility of introducing a casemix grouper software as a tool in obtaining a measure of current hospital output. The ability to measure outcomes can lead to better accountability in the use of public funds which in turn would enable policy makers to achieve greater efficiency in the overall operation of the hospital whilst ensuring the sustainability of the health care system.

The performance of the health care system can be improved only if its output is adequately measured. This thesis also aims to understand the main factors which affect performance within the health care sector. In view of the contribution of health professionals towards the overall running of the hospital, this study seeks to identify the influence of particular consulant characteristics on patient outcomes and consultant practice patterns. This is important so as to identify the existence, nature and possible channels of the contribution of health professionals towards outcome effectiveness. Variations in practice patterns are expected to have important consequences on the quality and cost of health care treatment. The job contract type of consultants working with the public hospital is expected to be a determining factor in explaining the behaviour of the consultants. The study aims to identify and measure the impact of job contract type on both patient outcomes and on practice behaviour patterns of consultants. This understanding helps in the setting of policy, based on empirical findings, to address the needed reforms which have been at the top of the health care agenda over recent years.

In the absence of a clearly defined casemix classification system for the Maltese health care sector, this thesis opens with an evaluation of the feasibility of introducing a casemix system based on Diagnosis Related Groups (DRGs). DRGs have been advocated in the literature and used in a number of countries as a primary tool to assess the efficiency and effectiveness of hospitals in providing acute patient care. The setting up of DRG groups could provide an effective measure of health care 'output' for the hospital.

DRG categories provide a measure for identifying hospital patient cases which involve the use of excess resources in comparison to the average patient case. There is a continual strive to understand these differences in resource use, identify reasons for such differences and address such differences. The measurement of 'output' within this setting would thus provide a consistent and reliable indicator for the analysis of differences in practice arising from factors which are not taken into account by the DRG grouping method. Such factors are likely to include, amongst others: the behavioural characteristics of the consultant and the surgeon when dealing with the patient. This can have important consequences for the implementation of policy measures within the health care sector.

The data collated and prepared for use in this thesis has been provided by the Clinical Performance Unit within Mater Dei Hospital,[1] and cover a three year period (2009-2011). The dataset provides information on individual patients admitted to hospital and the consultants and surgeons practicing within the hospital. The data also identify the diagnosis (a six level digit code based on ICD-10)[2] and procedures (a three level code based on ICD-9-CM)[3] related to each patient admitted to hospital, the date of the patient's admission and discharge, their age and sex, the final destination of the patient following discharge and all of the investigations and tests that were ordered and carried out during their hospital stay, including the date when the investigations were performed.

Furthermore, the dataset includes a number of characteristics which describe the surgeons who carried out the procedures and the consultants who had overall responsibility for each of the patients treated at the hospital. A data preparation exercise was undertaken to eliminate cases with invalid data and a number of technical edits and algorithms on the coding structure of the data were performed to ensure it was able to be used with the DRG Grouper software. A set of diagnostic and procedural variables was also constructed. This included counts of diagnoses per patient, counts of procedures and also a range of differing counts of consultant

---

[1] Mater Dei Hospital is the only public general acute hospital providing services on the Islands of Malta.

[2] International Classification of Diseases – 10th revision (WHO).

[3] International Classification of Diseases – 9th revision – Clinical Modification (WHO).

and surgeon volume measures. The availability of individual level data provided a more robust dataset for estimation and analysis purposes.

Consultants are major players in the overall running of a public hospital and their level of activity and behaviour decisions are expected to significantly influence the overall outcomes for patients. Consultants are employed within the public hospital on one of two possible contract options: they can work exclusively with the public hospital or they can decide to opt for a contract which allows them to work in private practice alongside their public sector commitments. It is expected that within the context and characteristics of the Maltese health care system, the prevailing consultant job contract conditions will explain some of the variation observed in practice patterns evidenced from the data.

As suggested from a review of the literature, a first attempt by an organisation at implementing a DRG casemix system involves the use and application of a tried and tested casemix system imported from another country. Chapter 2 of this thesis looks at the specific application of the MS-DRG[4] (Version 27) Grouper system on the Maltese data and applies the Coefficient of Multiple Determination (R^2) and the Coefficient of Variation (CV) to assess the heterogeneity between the obtained DRG classes and the degree of homogeneity gained in explaining resource use from the grouping of cases by DRG categories respectively. The Length of Stay (LOS) at the hospital for each individual patient is used as a proxy for resource intensity use. The results obtained from the study compare well with those obtained in other countries

[4] Medical Severity Diagnosis Related Groups (MS-DRG) provided by the Centers for Medicare and Medicaid Services, Department of Health and Human Services, USA.

and this supports the view that the management of the health care system in Malta would benefit from the implementation of a DRG casemix system.

Following on from the findings presented in Chapter 2, the analysis presented in Chapter 3 and Chapter 4 of this thesis focuses on a particular procedure, Percutaneous Transluminal Coronary Angioplasty (PTCA), performed within this particular hospital setting. Chapter 3 looks at differences in the expected relationship between volume and competing risk outcomes and whether this relationship varies in view of different consultant job contract conditions. An extensive literature review summarises the wide-ranging work on the relationship between volume and outcome within varying hospital settings and between varying individual physicians. Whilst acknowledging this relationship, the study specifically considers the strength of this relationship in the context of a single provider public hospital. Furthermore, it was deemed important to look deeper into the resulting outcomes following a hospital intervention and analyse the factors which affect such outcomes.

The main event of interest is *failure* within the 60 day period following the discharge from hospital after the undertaking of a PTCA procedure. Patients can be readmitted to hospital following discharge within the 60 day period or may otherwise die either during the hospitalisation period or within the 60 day period following discharge. The occurrence of the events *readmission* or *death*, are both treated as failure events. An event is considered to be a *success* if no event takes place within the 60 day time period following discharge. In line with a number of suggestions in the literature, both non-parametric and semi-parametric survival analysis methods are applied to study the rates for the occurrence of the events of interest. A multinomial logistic model is also estimated to gauge the robustness of the results obtained from the survival analysis methods.

The results obtained support the view that differences in outcomes for the event of interest are dependent on the patient volume levels of consultants and surgeons and also on the particular job plan contract type of the consultants. Even within a single hospital setting, one would expect differences in the behaviour of consultants and surgeons when treating patients to occur. Variations in medical practice arise when patients with similar characteristics make use of different levels of hospital resources. Such differences at the individual level are expected to have an impact on the overall performance, both in terms of quality and cost, of the health care system. A reduction in variation would be expected to lead to an improvement in the overall efficiency of the health care system and would have an impact on patient welfare and overall health outcomes.

Whilst many studies focus on the variation between hospitals, Chapter 4 of this thesis studies the behaviour of individual consultants working in the context of specific incentives and work practices of the Maltese health care system. Within this context, the role of the consultant job contract type is expected to be a prime determinant of heterogeneity arising among treatment patterns. Practice variation is measured by the number of investigations ordered and carried out on the patient undergoing a Percutaneous Transluminal Coronary Angioplasty (PTCA). In the study, data were used to distinguish between 'necessary' investigations in the treatment of PTCA undertaken in the first two days of a patient's hospital stay, and those undertaken afterwards and thus considered 'less important'. This paper contributes to the literature by analysing how practice variation differs by consultant job contract type in relation to the different time periods during which the patient is in hospital.

The consultant responsible for the admitted patient has direct control over the number of investigations carried out at the hospital. Variation in such decisions could reflect consultant's strategic behaviour due to his type of contract and/or consultant treatment preferences. In this study, a theoretic utility function is developed to analyse the utility associated with the behaviour of doctors working within the context of a single hospital who choose to either work exclusively for the government or also practice in the private sector.

In the literature, the variation in utility has been linked to physician and patient characteristics in addition to the organisational practices within the health care system. Both the uncertainty of a choice of treatment and its resulting effect have been identified as prime determinants of variation. Of importance to the study is the fact that due to the small size of the health care sector, patients are expected to be knowledgeable of physicians who carry out this particular procedure, whilst practicing physicians are concerned about their overall reputation.

A two-stage multilevel modelling approach is used in Chapter 4 of this thesis to study the relationship between the specific consultant contract conditions and the number of investigations ordered by the consultant. This level of analysis ensures that causes of variation, which might be omitted if the analysis is carried out at an aggregated level, are included and taken into account. A one-stage multilevel model, including group level effects, is also used to gauge the robustness of the obtained results.

The results show that both the job contract type and the volume levels of the consultant significantly affect practice variation. Furthermore, it was found that the responsiveness of practice variation to such characteristics varies in relation to the period of analysis, i.e. during the first two days of hospital stay compared to the rest of the days spent at hospital.

This thesis concludes with a summary of the main findings and implications of the analysis undertaken and presented in the respective chapters. In addition, some possible extensions to the research are discussed. The need to be able to benchmark hospital activity against an ‘acceptable’ practice which is viable, is crucial information for the better management of the available limited resources and in monitoring health care providers’ behaviour effectively. The understanding of the relationship between health care outcomes, patient, consultant and surgeon characteristics is crucial for the hospital authorities in the setting up of effective policy throughout the health care system.

Chapter

2

The Appropriateness of a DRG Casemix System in the Maltese Hospital setting

Abstract

The healthcare system in Malta is financed through public funds with free healthcare provided at the point of service. The aim of this paper is to examine the feasibility of introducing a Diagnosis Related Groups (DRGs) casemix system for Malta, not necessarily for payment and funding purposes, but as a tool to help describe, manage and measure resource use.

Like many other countries, Malta faces a growing concern of how to cope with the ever-increasing demands for health services within the context of limited resources available for health care needs. To date there has been no previous application of a casemix system for Malta. Based on the experience of other countries, a first step towards the use of a patient classification system normally involves the importation of a pre-existing casemix system already used and tested elsewhere. This study evaluates the practicability of the MS-DRG (Version 27.0) Grouper casemix software developed in the US within the context of the existing acute hospital care sector in Malta.

The classification of 151,615 individual patient cases admitted to hospital between 2009 and 2011 resulted in 636 different DRG categories being defined. Around half of these DRGs accounted for 99% of the total activity at the hospital. 296 DRG categories had less than 15 cases over the period, which highlights the need to carefully deal with relatively small sized DRG groups. The patient Length of Stay (LOS) is the indicator used as a proxy for resource use and the extent to which the DRG system is able to explain resource use is measured by two coefficients: the Coefficient of Multiple Determination (R^2) and the Coefficient of Variation (CV). In this study an initial R^2 value of 0.19 was obtained and this improved to

approximately 0.25 when a number of trimming algorithm methods were applied to the data. A good proportion of the resulting DRGs had a CV which was less than 1, indicating a considerable low degree of variability within the obtained DRG groups.

Based on the analysis undertaken, there is good evidence to support the introduction and use of a DRG based system in Malta. The derived DRGs will help to define the service mix of the hospital and can be further used by the management of the hospital to assess efficiency and effectiveness within the Maltese hospital setting.

2.1 Introduction and Motivation

The ability to measure the outcome of health care is critical to improving the effectiveness, efficiency and accountability of any health care system. Hospitals are deeply rooted in the political and administrative organisation of their country and typically account for the majority of spending by Government within the health care sector of a country. As stated by Lord Darzi (2008) 'we can only be sure to improve what we actually measure'. Casemix systems, particularly Diagnosis Related Groups (DRGs), have been implemented in a number of countries as part of a reform process to improve efficiency within the health care sector. DRGs help policy makers obtain an estimate of the activity undertaken within the hospital and thus assist in the understanding and measuring of the output of the hospital entity.

Measuring health care output through the use of indicators such as the number of beds, deaths or discharges, although all valid, do not provide the necessary measure of output from the health care system. The primary focus of this paper is the measurement of output in the Maltese health care sector through the use of a DRG casemix system.

This study seeks to assist policy makers in their endeavour to evaluate and adopt the appropriate policy guidelines to ensure the sustainability of the provision of free health care services. By applying DRGs in this context, the study will contribute to the debate of the relevance of such systems when applied to small countries with very particular hospital characteristics. Furthermore, a DRG casemix system would provide the necessary framework to studying the performance of physicians working within the hospital system.

To date, there has not been a study on the application of a DRG casemix classification system to health care in Malta. Scheller-Kreinsen et al., (2009) noted that most countries which introduced DRG systems as part of their reform initiatives have often imported a pre-existing casemix system from another country even though it may not fully reflect their own health care practice patterns. It is only later that countries decide to refine the implemented casemix system to better reflect their own health care system. This paper explores the practicability of applying the Casemix Grouper[5] software MS-DRG[6] (Version 27.0) used in the US in the context of the Maltese health care sector data. The derived DRG categories will be evaluated in terms of the homogeneity and power of the generated DRG groups to explain resource use. The DRG casemix system is now in use in a number of countries and has been applied to a variety of health care systems.

The multi-product nature of hospital output is a major factor to be dealt with when defining hospital activity. Classes of patients with similar clinical attributes and similar processes of care provide the necessary framework to aggregate patients into case types or *products* which entail the use of similar resources. DRGs provide a management tool which views the delivery of health care as a service, a production process in which outputs (health care episodes) are delivered to consumers (patients).

The results presented in this paper show that there is a good basis for recommending the introduction of a DRG based system to describe the hospital

[5] A DRG Grouper is an algorithm which takes clinical and demographic data as input and gives a corresponding Diagnosis Related Group as output.
[6] Medical Severity Diagnosis Related Groups (MS-DRG) is a grouper software provided by the Centers for Medicare and Medicaid Services, Department of Health and Human Services, USA.

output activities in the Maltese health care system. This will help to partition the episodes of care in an informative and meaningful way for the better management of resources within the health care sector. The formulation of DRGs would encourage administrators to view the use and costs of hospital services along product lines based on DRGs and in so doing, provide information on whether resources used for particular episodes of care are in line with what is expected for an average case within a particular DRG group.

The paper is organized as follows: the following section, Section 2.2, provides an overview of the literature related to the process of introducing a DRG casemix system together with the expected consequences on hospital activity resulting from its implementation. Section 2.3 gives a brief description of the main characteristics of the health care sector in Malta together with a detailed analysis of the available data and the necessary adaptations required to ensure that the grouper software is suitably applied. The methodology used to assess the reliability of the obtained DRGs is described next, followed by a discussion of the results obtained when measuring hospital output within the Maltese acute health care sector using DRGs. Finally a number of conclusions are presented.

2.2 Literature

Casemix is defined by Anderson (1984) as a system which "groups individual cases into a smaller number of case types so that the cases within a given group are homogeneous and the case types themselves are heterogeneous from other groups in some meaningful way". Patients within a clinical group are expected to have reasonably consistent resource consumption levels and such levels are expected to differ from other groups. DRGs are a tool which describes the number and type of treated patients in a hospital setting. According to Zhiping et al., (2004b) a casemix system has to have three important characteristics: clinical meaningfulness, resource use homogeneity and a manageable number of classes. The DRG system is the most well known casemix system, designed to group together acute inpatients, who are similar clinically and who have a similar pattern of resource use.

The first DRG classification system was designed in the United States following work at Yale University by Fetter, Thompson and Averill in 1969. The intended primary use was to achieve quality improvement and to review utilisation. However this changed in the funding of hospitals in the 1980s. The constant rising cost of health care in the US was the primary motive for developing a prospective payment system. Scheller-Kreinsen et al., (2009) highlight three main reasons for the introduction of DRGs: increased transparency, greater efficiency and supporting in-hospital management.

Most EU nations recorded a period of recession throughout the 1980s and this was the time when most countries initiated a process of reform within their health care systems. Given the fact that the acute hospital sector generally constitutes a major component of health care expenditure, most European governments sought to use the introduction of a DRG system to primarily target improvements in efficiency, in resource allocation and cost containment within their respective health care system. Governments also sought other objectives, not primarily economic in nature, such as reductions in waiting lists and improvements in the quality of care. The 1990s saw the implementation and application of DRG systems within most European countries such as the UK, Italy, France, Ireland and Portugal to mention just a few. As highlighted by Wiley (1999) this was at a time of increasing pressure on health care expenditure within a number of countries leading to the implementation of a number of reform options within the respective health care systems.

The motives underlying the introduction and the development of DRG systems, together with the particular design features of the system, vary greatly across countries (HOPE, 2006). The provision of health care systems, their historical background and cultural environment within which they operate, have impacted on the process of implementing a DRG system. Within the setting of global budgets (as was the case seen in most European countries) the introduction of DRGs was mostly directed towards fairness and efficiency rather than cost containment. Countries adopting a DRG casemix system now had a way of obtaining information on individual services offered and also on the costs per case incurred (provided that these were available).

In countries where the inpatient sector is dominated by the public sector and where most workers are on fixed salaries, the incentives of introducing DRGs tend to differ. Scheller-Kreinsen et al., (2009) concludes though that one would still expect DRGs to increase incentives for effective and efficient service delivery in such a setting. Introducing DRGs would also help to establish a financial mechanism which would ensure that production costs, in terms of resource use were capped to a certain level. Once hospital activity levels and casemix have been determined, health authorities can allocate annual hospital budgets, both prospectively and retrospectively. Furthermore, the documentation needs associated with the setting up of DRGs would provide managers with the information needed to monitor and control the work of clinicians. DRGs provide the information to help identify cases which could be judged as being relatively 'in line with the norm' in comparison with cases for which treatment is less efficient in terms of resource use.

The widespread use of DRGs across countries is in itself a sign of the success of such systems in helping health services achieve pre-set targets of better management and greater efficiency levels. However, the success of adopting DRGs depends on the interlinkages between the various entities which make up a health care system, in particular, the role of clinicians and the political commitment of governments to implement the necessary changes within the health care system.

The introduction of a DRG system led to a number of benefits within the various health care systems in which they where introduced, in particular, as presented in the literature these include: reducing waiting times, increasing activity, stimulating provider competition and facilitating patient choice of hospital, controlling costs, improving transparency in hospital facilities and harmonising payment systems (Dismuke and Sena, 1999; Louis et al., 1999; Mikkola et al., 2002). Furthermore, hospitals tend to be better at coding their activity in view of the fact that this would serve to ensure that funding is in line with the activity carried out within the hospital. The better coding of activity undertaken within the hospital ensures that more accurate information and more transparency is achieved in terms of undertaken activity. The DRG system provides more data on hospital production and costs in comparison to other management systems. The data used for the DRG system would serve as an instrument for the allocation of public hospital budgets funded by the national government.

Following the introduction of a new financing system based on DRGs, Aparo et al., (2009) found significant lower costs per discharge and a reduction of more than 50% in patient Length of Stay (LOS) having thus an impact on resource allocation within the hospital. This study however fails to consider the possible impact on the quality of care. DRG's provide the necessary data which allows for the setting up of an appropriate benchmark in terms of resource use for each of the assigned DRG categories within the hospital setting. The hospital would thus be in a position to plan prospectively in terms of resource usage.

There is broad agreement in academic and policy circles that the introduction of DRGs also affects provider behaviour (Scheller-Kreinsen et al., 2009). The introduction of DRGs has led to an increased level of activity in the short term, which results from the fact that the introduction of DRGs provides an incentive for shorter hospital LOS periods (Dismuke and Sena, 1999; Louis et al., 1999; Mikkola et al., 2002). In a study of the impact on outcomes following the introduction of changes in the hospital payment system, Moreno-Serra and Wagstaff (2010) conclude that inpatient admissions do not seem to be affected by a shift from budgets to patient based payment methods whilst Average Length of Stay (ALOS) seems to be reduced by about 4%.

DRGs offer hospital management not only the reason, but also the management and measurement tools needed to have better control over resource use. In particular, Berki (1985) finds that one could collect information on hospital resources used by individual physicians and those identified as using *wasteful* resources could be persuaded to change their practices. Therefore, the success of introducing a DRG system should not only be gauged by changes in efficiency but also by the effect of incentives on provider behaviour that result from the introduction and implementation of the DRG system.

Busse et al., (2006) balances the gains from DRG use of generating valuable information on costs, casemix and better cost control per diagnosis, against the problems resulting from cream skimming, up-coding or DRG creep, cost shifting and quality skimping. Indeed physicians working within a health system, based on case mix classification, would prefer to treat low risk patients over other patients with a higher risk factor assuming that both are classified under the same DRG category. Furthermore there is also the possibility of shifting patients, which are relatively more costly to treat to other centres/hospital.

There is infact an added incentive for the hospital concerned to up-code cases treated at the hospital. DRG creep would help the hospital to attract more funds or would serve to justify the use of more resources within the hospital. The fact that the use of resources is specifically set by the assigned DRG category reduces the incentive of practitioners to seek improvements in quality (referred to as quality skimping) over and above that required by the DRG assigned category. This resultant unintended behaviour, which can arise after DRG implementation, is identified as a cause of concern.

Kahn et al., (1990) find that after the adoption of the DRG based payment system patients were overall sicker at admission, the LOS of patients dropped, patients became less stable at discharge whilst in-hospital mortality decreased. Similar results were obtained by Louis et al., (1999) in a study using Italian data whereby the implementation of the DRG system resulted in a decrease in ordinary hospital admissions, a decrease in ALOS, an increase in day admissions and greater severity of illness amongst hospitalised patients. No change is noted in mortality and readmission rates. Finger (2000) however, finds that the authorities in the US became concerned about the number of readmissions registered in hospital cases following the introduction of a DRG system. The introduction of DRGs led to cost shifting whereby costs were shifted to other parts of the health care system away from the acute hospital (Jonsson, 1996; Morrissey et al., 1988) and/or cream skimming practised by hospitals as evidenced by Bibbee and Padrini (2006).

Experience shows that many countries introducing a DRG system share common difficulties, in particular the ever increasing demand to meet the expectations of their population in terms of health care needs and the pressure to curtail expenditure. The response of countries varies in relation to the historical, cultural and political interlinkages within their particular societies. What seems clear though is that most countries have turned towards DRGs and casemix tools to control costs, improve quality and achieve improved efficiency levels. The DRG tool has been seen to be flexible enough to be adapted to support different country needs and health care system characteristics, whether tax-based, insurance based or even financed through budgeting practices or contracting.

2.3 The situation in Malta

Malta has an integrated health care service, organized at a national level with one main acute general hospital which services the entire population. Health services in Malta are provided mainly by the state with an increasing contribution from the private sector being registered over the most recent years. Health care expenditure is primarily of a public nature, which expenditure is though complemented by private financing through out of pocket payments and private insurance. The total expenditure on health as a percentage of GDP is of around 8% and two-thirds of this is state financed. The statutory health care system is funded by general tax revenue, with all forms of taxation feeding into the Consolidated Fund from which the annual budget allocation for health is drawn. The Ministry for Finance determines the size of the health care budget yearly. There is no reimbursement system operating in the Maltese health care system as all public health care services are provided on a free basis at the point of use. There are no co-payments for any part of the health care system.

The public health care centres and hospitals provide free access to preventive, investigative, curative and rehabilitation services, irrespective of income or ability to pay. Most hospital care is provided within the Mater Dei Hospital (MDH), which is the only acute general hospital in Malta. It is estimated that over 80% of the population of the Maltese Islands rely on the health care services provided by the state. MDH also offers services in all other specialties which include amongst others: radiology, ophthalmology, urology and pathology. Given the geographical characteristics of Malta and particularly its relatively small size, the hospital caters for all the needs of the population, even though some cases which require particular treatment or which are relatively uncommon are usually referred to specialized centres abroad.

Although universal coverage for government health services exists, people tend to use the private sector for what they perceive to be more personal attention and care together with the desire to be seen by the consultant of their choice. Until fairly recently, direct out-of-pocket payment used to be the method of payment for all private health care services ranging from consultations to interventions. The setting up of a number of private hospitals and clinics over the recent years has led to considerable changes within the private health care insurance market.

Out of pocket payments are still though the predominant method of payment for private general treatment given that there are no incentives for people to take out private health insurance at present. Private health insurance is mainly taken out for the eventuality of elective surgery, hospital care and medical treatment overseas. There exists a perception that "complex" interventions are only possible, or are safer if carried out at the public hospital. Private insurance companies also offer a cash rebate per diem for insured persons who opt to make use of the state hospital to treat particular and specific conditions. As a result, the state bears most of the brunt of the financial burden of health care.

Consultants working at MDH are engaged as government employees, paid in accordance with a salary grade scheme which applies to all other government employees. Consultants choose between two work plan options, *public-only,* whereby consultants cannot undertake any private practice and *dual practice* whereby consultants can practice in both the private and the public sector.

Within this context the Government of Malta is committed to keeping health care free at the point of delivery and it is the Government's intention to improve resources and ensure that they are managed efficiently (Ministry of Health - Government of Malta, 2008). A major challenge of the current health care system in Malta is to ensure the sustainability of the system whilst developing the necessary mechanisms to adequately measure quality and outcomes. It is in this context that this paper evaluates the construction and relevance of DRG practicability for Malta's health care system.

2.4 Data

This study uses patient level data provided by the hospital management unit[7] of MDH for the years 2009, 2010 and 2011. Given the high degree of data fragmentation, a matching and database linking process was carried out to ensure that a workable dataset was available. This involved the use of a mapping algorithm to map the Maltese data to the requirements of the Grouper software. Three different datasets[8] at the patient level were integrated together through the use of an encrypted patient ID code. Data up to a six level digit diagnosis code, based on the ICD-10[9] classification system of the World Health Organisation (WHO), was made available for this study whilst data for procedure codes up to three coding

[7] The Clinical Performance Unit (CPU).
[8] The Surgical and Operations Register, The Admissions and Transfers Discharge Database and the Hospital Activity Analysis Database. All databases are maintained by the CPU at MDH.
[9] International Classification of Diseases – 10th revision (WHO).

levels was based on the ICD-9-CM[10] classification system. A number of technical edits and assumptions, in consultation with the hospital management authorities, were applied to the dataset in order to obtain data in the requested classification system (ICD-9-CM) for the diagnosis codes required by the MS-DRG Grouper[11] used in this study. In particular, this involved eliminating cases with invalid data such as invalid age or sex, dealing with cases with a missing primary diagnosis, re-specifying the admitting diagnosis codes as well as fine-tuning the discharge status codes for a number of cases. The DRG software was applied to a final dataset of 151,615 patient cases.

Data drawn from the following categories at patient level were the main inputs for the DRG classification Grouper: admission and discharge date, age, sex, diagnosis and procedure codes together with discharge status. Other data related to different consultant and surgeon characteristics and to investigations carried out during a hospital stay was provided. The dataset used in this study relates to episodes of inpatient care for individuals who have been given a bed at MDH. This includes day cases for which the LOS would be recorded as zero.

LOS: The LOS variable is an important indicator of resource utilization. It is also easily extractable, well standardized and generally reliable. LOS is a physical measure of hospital resource use and provides a suitable measure for intermediate

[10] International Classification of Diseases – 9th revision – Clinical Modification (WHO).

[11] Data for diagnosis codes was initially converted from the ICD-10 coding structure to the ICD-10-CM structure. This involved a detailed and meticulous process of allocating cases from the more general ICD-10 structure to the more specific ICD-10-CM. A number of assumptions, based on the recommendations of the clinincal performance unit within the hospital, had to be implementated given the lack of required detail in the available ICD-10 database. A backward mapping algorithm (General Equivalence Mapping (GEM)) provided by the Centers for Medicare and Medicaid Services (CMS) of the US government was then applied to the data to obtain data in the ICD-9-CM coding structure required by MS-DRG version 27.

products used in the treatment of the inpatient. In the original design of the DRG system by Fetter and colleagues in 1969, the LOS variable was found to correlate well with resource consumption (Fetter et al., 1980). Whilst recognising the shortcomings of using LOS as a proxy for resource intensity in the production of hospital products, Rhodes et al., (1997), highlight that it is nevertheless, in the absence of a practical alternative, the measure most often used. Cost-accounting data for measuring resource consumption in the treatment of patients in Malta is not available[12] and therefore in this study, the LOS variable will be the proxy variable for resource use.

Outliers: Of particular interest and concern to the analysis of DRG groups is the treatment of outlier cases and their expected impact on the homogeneity of the assigned DRG groups. The treatment of outlier cases in Malta's context is of prime relevance given the relative small size of the health care sector and the likelihood that a considerable amount of DRG categories would only comprise relatively few observations. As highlighted by Cots et al., (2003) outliers ought to be valued differently from inliers as their presence would lead to a mean value of resource use which is not representative of the DRG group. There is thus a case for removing outlier observations which cause much of this disturbance.

A common approach applied in the literature is to remove at the outset values outside a pre-set limit given that they have an unrepresentative impact on the overall average within that particular group. The real reason for excluding such 'extreme' cases (apart from the statistical problems which they cause) is that such cases represent unusual occurrences not necessarily predictable from diagnosis or

[12] A cost accounting exercise is currently being undertaken at the Ministry of Health.

procedural information such as, lack of beds for transfer and lack of available services related to home assistance and care.

The identification and removal of outlier observations in statistical analysis helps to provide a sounder basis for examining the characteristics of a population under review (Reid et al., 1997). Outliers (following the removal of the 'extreme' cases) can be defined as observations that appear to be different from other similar observations in the dataset or DRG group. The inlier status of a patient is thus defined on the basis of the assigned DRG and on the population of episodes selected.

Two types of outliers are identified: cases of patients with LOS which is much longer or shorter than the average (a stay outlier) or cases of patients with costs which are more or less than the average for that group (a cost outlier). Only stay outliers will be considered in this study. Lichtig (1986) and Reid et al., (1997) identify a number of reasons related to the existence of outliers: data errors, unusual combinations of clinical conditions, hospital-acquired complications and misadventures. Some DRG categories might be more susceptible to problems of generating outliers, whilst treatment mishaps in particular categories can also be highlighted as possible causes of outlier cases.

2.5 Methodology

In this section the main features of the Grouper used to derive the DRG categories will be reviewed, highlighting the characteristics which will determine the assigned DRG groups. An overview of the trimming methods applied in this paper to identify outlier cases are included in this section. This is followed by a description of the Coefficient of Variation, (*CV),* and the Coefficient of Multiple Determination, *(*R^2), used to assess the performance of the casemix classification Grouper.

MS-DRG: The MS-DRG Grouper delivers the operational means of determining the types of patients treated and relates this to the resources consumed within the hospital. The Grouper partitions the data, prepared to specification, in such a way, that information is provided to the user on the clinical inputs required and on the resources expected to be consumed by each episode in question. Although all patients are unique, groups of patients have demographic, diagnostic, and therapeutic attributes in common that determine the intensity level of their resource usage.

The guiding principle of DRG casemix systems is to divide patients into homogeneous groups in terms of age, sex, diagnosis, procedures and discharge status, groups which are clinically meaningful and relatively homogeneous in resource use. Such characteristics are important in explaining resource use and are thus incorporated into the preset algorithm used to determine the particular DRG. MS-DRG executes an algorithm based on a number of characteristics reviewed in a particular order: presence of operating room procedure, principle diagnosis, age, sex, complication or comorbidity, various secondary diagnoses, and discharge status.

The MS-DRG version chosen for this study uses the ICD-9-CM international coding classification for both diagnosis and procedures. DRGs are formed by first allocating all possible diagnoses into 25 mutually exclusive Major Diagnostic Categories

(MDCs)[13] reflecting the part of the body organ system which is being affected. Cases with at least one operating room procedure are referred to as *surgical* and those with no procedures are treated as *medical* cases; these are characterised by the principle diagnosis for which patients are admitted. Once the medical and surgical classes for an MDC are formed, each class of patients is evaluated to determine whether complications and comorbidities, the patient's age or discharge status, consistently affect the consumption of hospital resources. Grouping continues by assigning cases from the appropriate MDC categories to the particular DRG groups within each MDC based on those patient characteristics which are expected to impact on resource consumption patterns. The DRG logic algorithm within MS-DRG defines around 1,000 different DRG groups.[14]

Trimming: A number of techniques have been applied in the literature for defining outliers, many of which have a purely statistical underpinning (Bender and McGuire, 1995; Coombes et al., 1995; Reid et al., 1997). Lichtig (1986) uses multiples of the standard deviation of the original or logarithms of the data to define trimming points. Reid et al., (1997) and Zhiping et al., (2004a) stress the use of an algorithm based on the interquartile range (IQR) whereby the low trim point is defined as $Q1-1.5*(Q3-Q1)$ where $Q1$ is the first quartile, $Q3$ is the third quartile and $(Q3-Q1)$ is the interquartile range. The high trim point is defined at $1.5*(Q3-Q1)+Q3$. Palmer et al., (2001) also use the interquartile range method to define an upper trim point beyond which all cases are defined as outliers.

[13] A list of all the 25 Major Diagnostic Categories is provided in the Appendix to this chapter (Table2.8).
[14] The definitions manual with the full logic tables is provided by 3M Health Information Systems 2010, USA.

The trimming methods described above are applied to the dataset, following the removal of cases considered by the management of the hospital as 'extreme' cases. Casemix systems have not been designed to deal with such cases and thus the removal of such cases has to be done arbitrarily. It is within this context and in consultation with the management of the hospital that it was decided to remove from the data set all 'extreme' cases with a LOS in excess of 60 days.

The exclusion of extreme cases is imperative for the elimination of outlier values that otherwise would have had a significant and unrepresentative impact on the average for the respective DRG groups. Following the arbitrary elimination of the extreme cases a further process of distinction between inlier and outlier cases within the respective DRGs is carried out. The trimming of outliers would remove from the DRG group, cases which do not appear to belong to the underlying distribution postulated by the LOS for the majority of the cases within the DRG.

A number of trimming methods[15] will be evaluated in this study. These methods are applied to each of the particular DRG groups generated by the Grouper software. Two particular trimming methods stand out in the literature. One is based on the distribution of the elements that make up the DRG group, attempting to make the arithmetic mean of the group more robust (referred to in this study as the GM2 and GM3 methods), A second group of trimming methods is based on the interquartile range whereby a multiple of the range between the 25th percentile and the 75th

[15] A number of trimming methods are tested:
GM2 = geometric mean + 2*standard deviation
GM3 = geometric mean + 3*standard deviation
IR1.5 = 75th Percentile + 1.5*interquartile range
IR2.0 = 75th Percentile + 2*interquartile range

percentile is added to the 75th percentile (referred to in this study as the IR1.5 and the IR2.0 methods).

The preferred method in this study, the IR1.5, also used by the British National Casemix Office and the Australia Department of Health and Family Services is based on the interquartile method (percentile methods). This non parametric method makes no strong assumptions on the distributional form of the dependent variable and the trim points are not readily distorted by extreme values (Queensland Health, 2004). Furthermore, this method is relatively easy to understand apart from the fact that the method has a long history in the use of preliminary editing of data in statistical analysis (Palmer, Reid 2001). The IQR method has been used in a number of evaluation studies for countries adopting DRG systems; Freeman (1991) used IQR for European data (UK, Norway, Spain), whilst Casas and Tomas (1993) used IQR for data on Ireland, Portugal, Switzerland and Spain. The choice of this trimming method is also based on work by Soderlund (1996) for the National Health Service (UK) wherby the trimming method is chosen on the basis of the fact that such method should detect the greatest number of outlier cases whereby resulting in a more robust mean value for the DRG.

Furthermore, given the characteristics of the dataset, this study will thus not apply a lower trim point to the data. This strategy was also applied in a number of cases in the literature (Palmer et al., 2001). Trimming at the lower end would result in deleting LOS data on patients which are very important for the overall understanding of LOS patterns in the hospital at large.

The performance and adequacy of the MS-DRG Grouper to describe Maltese health care data, in terms of grouping episodes of care is evaluated through the use of the

CV and R^2 statistics. These are the two main statistical tools which are used for the evaluation of heterogeneity in LOS between DRG classes and homogeneity in LOS within DRG classes.

CV: This statistic is used to measure the within group variability or homogeneity in terms of LOS within established DRG categories and is obtained by dividing the standard deviation of LOS by the mean of LOS for each of the DRG groups. A value of zero is an indication that the group has no variance from the mean whilst a *CV* greater than 1 is an indication of heterogeneity within the group. The homogeneity of the DRG groups adds to the robustness of the DRG design. Fischer (2000) and Palmer et al., (2001) conventionally take a *CV* of less than 1 as an indication of DRGs with an acceptable degree of variation. Other studies by (Aisbett et al., 2007; Ghaffari et al., 2008; Reid et al., 1991; Zhiping et al., 2004a) use the same criteria to assess the performance of the DRG classification structure.

R^2: This statistic is a measure of the gain in explanatory power of the grouping classification in terms of resource usage (LOS in this case), which provides a measure of the extent to which the DRG system explains variation in resource use based on the characteristics of the individual patient. Averill et al., (1998) describe R^2 as the proportion of the variance of the whole population around the population mean (total variance) that is due to the variance of group means around the population mean (between group variance). The extent to which variation of LOS occurs between casemix groups, rather than within them, determines the strength of the grouping system. If n observations are divided into n different groups, and if all the variation is between groups and none within the group, then an R^2 value of 1 is obtained. The higher this ratio, the more of the total variance is said to be explained by the variance between groups as opposed to variance within groups and this is thus an indication that the groups are relatively heterogeneous between one another, as is required.

As is common practice in casemix research, a regression[16] model is used to explain the variation in resource use through the use of a LOS variable. The assigned DRG index category for the individual patient case is the independent variable used in the estimation process. This variable incorporates all the visit related information (including the patient characteristics) used by the grouper software to assign the particular DRG category. Most LOS distributions are asymmetric, usually with a long right tail and some very large observations (Zhiping, 2004). As clearly highlighted in most literature, the presence of even a few high observations (outliers) of the LOS dependent variable will greatly affect the R^2 value due to the skewed nature of the distribution for the LOS variable. The OLS method requires that the distribution assumption of normality in the error term holds and thus the emphasis of the analysis shifts to the trimmed data, after the treatment of such outliers. The trimming and the removal of extreme cases significantly contributes towards the attainment of distributional properties which are more in line with the requirements of the OLS method (Reid et al., 1991).

A high R^2 is an indication that the DRG classification explains a significant proportion of the variation in LOS, and thus the DRG could be used as a suitable basis for explaining variations in resource use for different hospital outputs. The explanatory power of the DRGs, represented by the R^2 in this study is compared to that obtained

[16] Based on the Ordinary Least Squares (OLS) method.

in other studies which seek to evaluate the appropriateness of introducing a casemix system within a health care setting.

R^2 values are also estimated for different MDC categories to show that variations are registered not only across DRG classifications but also for different MDCs. In line with the distributional property requirements of the applied estimation method, the trimming process applied within the MDC categories is expected to yield significant improvement in terms of such properties within the respective MDC groups. Averill et al., (1998) highlight that there is a systemic variation in R^2 across MDCs reflecting the fact that LOS is more predictable in some MDCs such as the circulatory system and less predictable in others such as mental health. The R^2 results obtained from this study, for the different MDC categories, are compared to similar results in the literature to gauge the adequacy of the DRG casemix system in such a context.

2.6 Results

This section commences with a general description of the hospital output results obtained from the DRG Grouper applied to the acute health care sector in Malta. Basic descriptive statistics for the most commonly assigned DRG cases are also provided. There then follows a presentation of the results obtained for both the CV statistic and the R^2 statistic under the different trimming options and for the various assigned MDC categories.

2.6.1 General description of hospital output

The classification of the 151,615 cases using MS-DRG resulted in 636 different DRGs, approximately 55% of which accounted for 99% of the total activity generated in the hospital. Further analyses showed that around 2% of the DRGs represented approximately 31% of the activity at the hospital. There were 296 DRGs with fewer than 15 cases each over the three year period (1,481 patient cases in total). Out of these, only 31 DRGs with fewer than 15 cases each had an ALOS shorter than 4 days.[17] This indicates that most of these cases are *complex* cases that absorb a significant amount of hospital resources to be treated. DRGs with few episodes are harder to interpret, as utilisation measures obtained for them are subject to relative sample variation not reduced by the law of large numbers (Aisbett et al., 2007).[18] This also applies to particular DRG groups with very low volumes, as can be found in health care systems with significantly large datasets as well as those like Malta.

Table 2.1 shows information on the DRGs with the most common occurrences for the years 2009-2011. The ALOS for each of the DRG categories and the size of each of the DRGs in percentage terms is given. The ALOS varied between DRGs and the sensitivity of the ALOS to outliers within the DRGs categories was analysed further by applying a number of trimming options to the data.

Table 2.1: DRGs with the 10 most common occurrences for the 2009-2011 period

DRG	Description	Occurrences	ALOS	% of cases
392	Esophagitis, Gastroent. & Misc. Digest Disorders w/o mcc	9727	2.5	6.5
313	Chest Pain	7508	1.9	5.0
117	Intraocular procedures w/o cc/mcc	6776	0.5	4.5
775	Vaginal delivery w/o complicating diagnoses	5437	2.7	3.6
951	Other factors influencing health status	4125	0.6	2.7
950	After care w/o cc/mcc	3531	0.3	2.3
581	Other skin, subcut. tissue & breast proc. w/o cc/mcc	3179	1.1	2.1
293	Heart Failure and shock w/o cc/mcc	2728	6.4	1.8
766	Cesarean section w/o cc/mcc	2445	6.6	1.6
607	Minor skin disorders w/o mcc	2312	0.7	1.5

Source: Analysis of MS-DRG Grouper output.

[17] 3.97 days is the ALOS when all patients in this study are taken into account.
[18] DRG categories with less than 15 cases over the 3 years of data are not considered in the analysis.

Table 2.2 shows the variation in ALOS when cases with different LOS periods were removed from the analysis. When cases considered as *extreme* (more than 60 day LOS) were excluded from the analysis, the ALOS fell to 3.63 days. In total there were 519 cases which reported a LOS in excess of 60 days. There are only 202 cases with LOS in excess of 90 days. Approximately 44% of the cases accounted for short stay patients (0 or 1 day) and the ALOS, when excluding short stay patients, was observed to go up to 6.77 days. Around 60% of the cases treated in the hospital had an ALOS of less than or equal to 2 days. Most distributions of LOS were asymmetric, with a long right tail and some very large frequency observations.

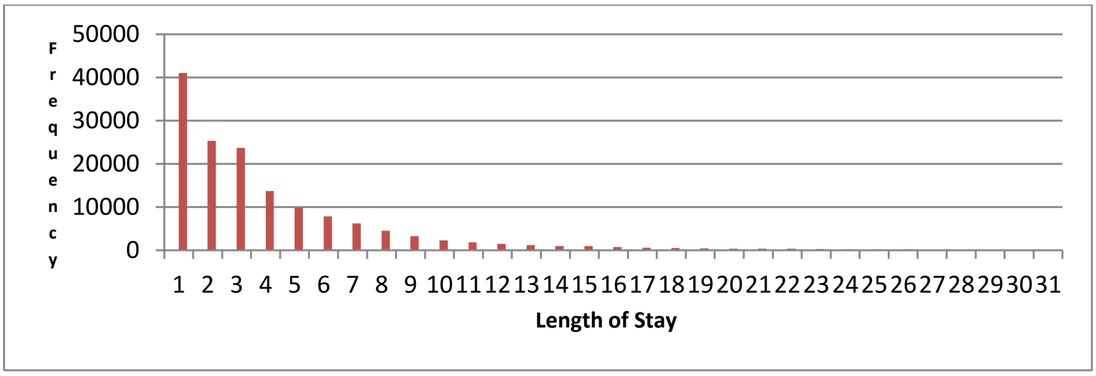
Table 2.2: ALOS (days) after removing outlier cases

	All cases	<120 days	<90days	<60days	<30days
Average length of stay	3.97	3.84	3.78	3.63	3.27

Source: Analysis of hospital episode data.

Figure 2-1 shows the frequency distributions for all cases with LOS which are less than or equal to 30 days. All other cases with LOS in excess of 30 days had an occurrence of less than 1.5% in the overall dataset. Patients with very large LOS were treated as extreme cases and were removed for analytical purposes, while others with atypical long or short stays, which are usually treated as outliers, were also trimmed.

Figure 2-1: Frequency distribution by LOS (less than 31 days)



Source: Analysis of hospital episode data.

As part of the preliminary output, the MS-DRG Grouper initially assigned individual cases to one of 25 MDCs reflecting the part of the organ system which was being affected in the particular case. The study found that MDC 5 (Diseases and Disorders of the Circulatory System) was the MDC with the highest volume at 23,325 cases, accounting for 15.4 per cent of the total cases in the three year period.

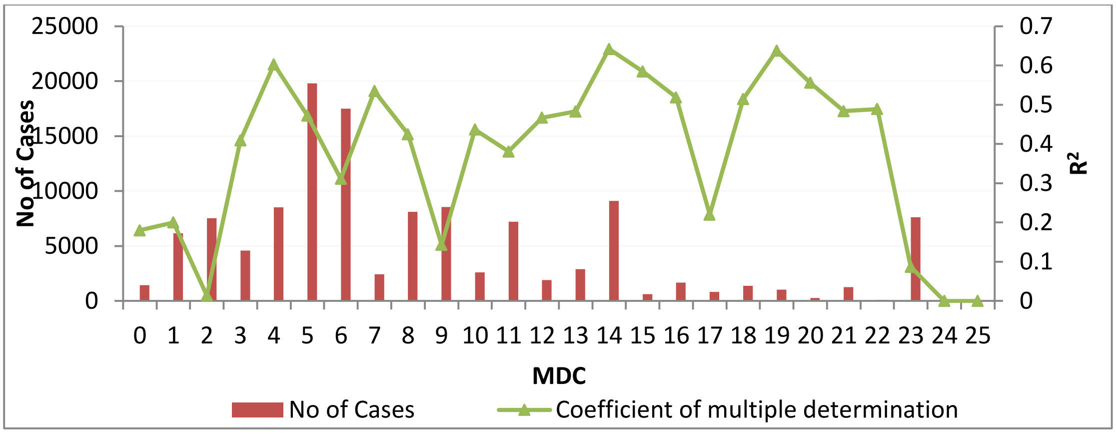
Table 2.3: MDCs with the 10 most common occurrences

MDC	Description	Occurrences	ALOS	% of cases
5	Circulatory System, Diseases & Disorders	23325	4.8	15.4
6	Digestive System, Diseases & Disorders	20685	3.1	13.6
14	Pregnancy, Childbirth, & The Puerperium	10942	3.3	7.2
9	Skin, Subcutaneous Tissue & Breast, Diseases & Disorders	10723	2.5	7.1
8	Musculoskeletal System & Connective Tissue, Diseases & Disorders	10671	5.4	7.0
2	Eye, Diseases & Disorders	10151	0.5	6.7
23	Factors On Health Status & Other Contacts With Health Services	9883	1.2	6.5
4	Respiratory System, Diseases & Disorders	9877	6.1	6.5
11	Kidney And Urinary Tract, Diseases & Disorders	9124	3.9	6.0
1	Nervous System, Diseases & Disorders	7374	5.3	4.9

Source: Analysis of MS-DRG Grouper output.

Table 2.3 and Figure 2-2 show that a significant proportion of hospital activity is grouped into a small number of MDC categories, in fact 81% of all the cases fell within 10 MDCs. The variation in ALOS between different MDCs reflects the particular characteristics of each MDC group. Groups with very low ALOS reflected treatment undertaken on a day care basis or treatment requiring only an overnight stay in hospital.

Figure 2-2: Frequency distribution by MDC groups



Source: Analysis of MS-DRG Grouper output.

The presence of outliers in the dataset is due to a number of factors, such as the age of the individual, case severity, the availability of discharge to other institutions, and the social circumstances of the patients admitted to hospital. Table 2.4 shows the number of outliers identified using the trimming methods applied in the analysis. It also includes the ALOS as a result of the trimming method adopted and the percentage of outlier cases which were trimmed. A universal trim was applied to the data and cases with a LOS in excess of 60 days were removed from the analysis. This resulted in the removal of 519 cases. It is recommended that these cases are subjected to further study.

Table 2.4: Analysis of outlier cases by different trimming methods

Trimming method	No. of trimmed cases	ALOS	% of outlier cases
universal >60 days	519	3.805	0.3
GM2	7788	2.938	5.1
GM3	4553	3.210	3.0
IR1.5	15141	2.834	10.0
IR2.0	12709	2.989	8.4

Source: Analysis of MS-DRG Grouper output.

The four trimming methods used in Table 2.4 were applied to the remaining cases and the resulting number of trimmed cases for each method is an indication of the variation in resource use within the particular DRGs. The IR1.5 method yielded a trim of around 10.0% of the available dataset. This is in line with the guidelines provided by Palmer and Reid, (2001) and the percentage of cases trimmed with this method is similar to that in other studies found in the literature. As a result the ALOS for the whole hospital fell to 2.8 days, or by around 25% implying that dealing with such outliers would lead to a reduction of around 1 day from the ALOS within the whole hospital; this represents a considerable saving in terms of resource use.

2.6.2 Statistical Analysis of Goodness of Fit

CV: The results presented in Table 2.5 stem from the derived DRG categories after the removal of extreme cases and those DRGs with 15 cases or less. This is referred to as *partial trimming* and it was found that 42% of such DRGs had a *CV* which was less than 1. When the IR1.5 trimming method was further applied to the data[19], the proportion of DRGs with a *CV* of less than 1 rose to 85%, being the highest

[19] The application of the other methods yielded the following percentages of cases with a CV < 1: GM2 – 78%, GM3 – 70% and IR2.0 – 82%.

proportion compared to the other trimming methods. The lesser the variation within the grouped DRGs, the better the performance of the classification system (Palmer et al., 2001). The IR1.5 method thus resulted in a significant number of DRGs (290) with a CV less than 1, highlighting the ability of the Grouper to produce homogeneous groups.

Table 2.5: Distribution of the DRGs with CV>1

Partial trimming (no of cases)		After IR1.5 trimming (no of cases)	
CV>5	1	CV 1-2	48
CV 3-5	7	CV<1	290
CV>2	35		
CV>1	195		
CV<1	143		

Source: Analysis of MS-DRG Grouper output.

The results suggest that following the trimming process an acceptable level of variation exists within the DRG groups and that trimming is a necessary task to improve homogeneity within DRG categories. The effect of trimming, especially by using the IR1.5 method, resulted in sizeable changes in the *CV* of the different DRG groups, helping to reduce the *CV* values by a significant amount. It was also noted that the impact of trimming on the *CV* for each of the DRG categories was also sensitive to the trimming method applied.

Aisbett et al., (2007) highlight that trimming would be useful if the removal of a small number of records greatly improves the Grouper's R^2 value and *CV* value for the specific groups. This requires the analysis of both the full set of data (inliers and outliers) and then specifically the inliers to gauge the differences in the coefficient of multiple determination as a result of trimming.

R^2: Table 2.6 shows R^2 values at different levels of trimming. In the absence of trimming, the DRG classification serves to explain 18.6% of the variation in resource use as measured by LOS. Trimming cases with a LOS greater or equal to 61 days improved R^2 to 21.1% and, further trimming by using the IR1.5 method, resulted in a further improvement in R^2 to around 25%. Nothwithstanding the variations in the number of trimmed cases resulting from the use of the different trimming methods applied in Table 2.6, the R^2 statistics obtained are relatively similar.

Table 2.6: R^2 under different trimming methods

	All cases	<60 days	GM2	GM3	IR1.5	IR 2.0
R^2	0.186	0.211	0.258	0.236	0.243	0.235
No of Cases	151,615	151,095	143,624	146,859	136,271	138,703

Source: Analysis of MS-DRG Grouper output.

The resultant trimmed R^2 values measure both the performance of the classification system (MS-DRG in this case) and also the characteristics of the patients trimmed. The results obtained suggest that the extra cases being trimmed under the choosen method (IR1.5) are in fact cases which should be removed from the dataset as they are not contributing to the performance of the classification system in explaining resource use. This helps to obtain cases within the DRG groups which are more comparable in terms of resource use.

Trimming increased the homogeneity within the DRG groups and the heterogeneity between groups. This further proves the ability of the derived DRGs to explain the differences in resource use. Based on the above results it can be concluded that the DRG classification can explain close to 30% of the total variation in LOS and therefore resource use between cases across the hospital. Further support of this conclusion was obtained by analyzing variations in R^2 across the different MDCs, an analysis which highlighted variations within different case specialties within the hospital.

Table 2.7 shows the degree of variance within each MDC[20], based on the allocation of cases to DRGs under different trimming scenarios. The percentage of cases trimmed under the full trimming scenario as a proportion of all cases under each MDC category is also provided. The percentage of cases trimmed varied between MDC categories and when all cases were taken together, an R^2 of around 19% was obtained. This highlights the fact that a significant proportion of cases within each MDC exhibited resource allocations which varied from the expected average of the particular MDC group. Figure 2.2 illustrates how R^2 varied across MDCs (after full trimming was performed). The higher R^2 values indicate that resource use was more predictable in some MDCs and that in these MDCs, variation in resource use can be better explained by the DRG classification.

In summary, it can be concluded that the trimming process yielded higher R^2 values in the majority of the MDCs analysed, with the exception of a few cases. It was also noted that there was a significant percentage point increase in R^2 for a number of MDC categories (14, 20, 21, 4, 5, and 16) after the IR1.5 trimming method had been applied to the data. One possible explanation for this is that once the removal of trim points was applied, the remaining cases in these particular MDCs had a high level of homogeneity within the group and higher levels of heterogeneity with the other groups. This resulted in a significant improvement in R^2 values. The difference observed arose from the particular patient characteristics being treated

[20] Table 2.8 in the appendix provides a description of the different MDC categories.

and also due possibly to differences in practice by physicians in the treatment of patients.

Table 2.7: R^2 by MDC categories under different trimming scenarios

MDC	All cases	Partial trimming[21]	Full trimming[22]	% of cases trimmed (full)[23]
	%	%	%	%
All	18.6	21.1	25.6	18.7
0	12.0	19.9	18.0	25.4
1	9.2	16.5	20.0	16.2
2	10.3	10.3	1.4	25.8
3	24.0	31.8	40.9	30.2
4	30.0	45.4	60.3	13.8
5	19.7	34.0	47.2	15.0
6	13.6	25.2	31.1	15.4
7	32.7	43.5	53.5	9.9
8	25.3	34.8	42.5	23.9
9	5.3	14.7	14.3	20.3
10	26.1	37.7	43.7	13.1
11	13.6	31.1	38.1	21.1
12	17.9	34.5	46.7	20.4
13	19.5	39.6	48.3	18.1
14	35.2	40.6	64.2	16.8
15	39.1	52.3	58.5	11.7
16	31.2	36.5	51.9	18.5
17	19.5	28.5	22.0	20.7
18	28.8	40.3	51.5	17.0
19	36.5	58.1	63.8	25.4
20	19.5	26.8	55.6	20.2
21	14.1	26.9	48.4	11.7
22	45.7	47.9	48.9	9.1
23	4.5	17.2	8.7	23.0
24	45.2	n/a	n/a	n/a
25	48.5	40.1	n/a	n/a

n/a represent values which cannot be calculated given the small amount of cases within such MDCs.
Source: Analysis of MS-DRG Grouper output.

An analysis of R^2 by MDC allows us to gauge whether the performance of the hospital is consistent across all types of cases. More than 70% of the derived MDCs had an

[21] Represents trimming of cases with more than 60 days LOS only.
[22] Represents trimming of cases with more than 60 days LOS, those with less than 15 cases within each DRG and using the IR1.5 method.
[23] Represents percentage of cases trimmed based on the full set of cases.

R^2 value higher than 0.4 after the trimming process. This indicates that in most of the MDCs the DRG classification can explain a significant part of the variation in LOS. Resource use is more predictable in some MDCs and less predictable in others. A low R^2 value was recorded for MDCs 0, 2, 9 and 23 after the trimming process. For each of these MDCs, a low R^2 value was recorded even when the untrimmed data was applied possibly indicating that in such cases the LOS variable might not be a good indicator of resource use.

A number of reasons could account for these low MDC values. One reason could be related to the particular set of grouping algorithms within these MDCs which did not specifically reflect the data of the Maltese health care sector. Another reason is that the nature of treatment within these particular MDCs may have been different in terms of procedures used and would vary on a case by case basis depending on other factors, such as differences in consultant and surgeon practice styles. Further analysis might therefore be needed to identify whether particular differences exist with the treatment of such categories in Malta and whether this is reflected in the assumptions taken into account by the MS-DRG grouping algorithm.

2.7 Discussion

Overall R^2 values of around 0.3 were found for LOS using trimmed data. The R^2 values obtained on untrimmed data were low (0.19) but this has to be viewed in line with the known quality limitations of the available hospital data. R^2 values also varied across the different MDC categories with some categories reaching levels close to 0.6 once trimming had been performed. Casas & Tomas (1993) and Reid et al., (1997) highlight that R^2 values based on untrimmed data may indeed be more than 20 percentage points lower after outliers are removed. The same studies by Casas & Tomas (1993) and Reid et al., (1997) suggest that cases removed as outliers have a substantial impact of around 25% on the ALOS. The adequate treatment of such outliers would imply an average reduction of around 1 day from the ALOS at the Maltese hospital. This is indeed a significant saving in terms of resource use within such a hospital setting.

The R^2 reported in this study compares well to that reported by Zhiping et al., (2004) in a Chinese hospital (0.12) using the Australian Refined DRGs (AR-DRG). Other studies by Closon & Roger (1989) report an R^2 value of 0.42 using Belgian data applied to the Health Care Financing Administration DRG (HCFA-DRG) Grouper software, whilst Casas & Tomas (1993) report R^2 values of around 0.4 using data from Ireland, Portugal, Switzerland and Spain.

Values for R^2 tend to vary between different Groupers and significant variations are also observed following the implementation of trimming methods to the available data. This conclusion is also evident in the work of Benton et al., (1998) who compare R^2 values arising from different Groupers. Reid et al., (1997) highlight the wide variation in R^2 at the MDC level, using both US DRGs and Australian DRGs. More recent studies compare specific locally designed systems with the US Groupers or other international available DRG systems. Sutch and Reid (2003) and Reid and Sutch (2008) quote a number of studies of this type, in particular work carried out on Welsh data using British Health Resource Groups (HRGs), US All Patient DRGs (AP-DRGs) and AR-DRGs. R^2 values ranging from 0.31 to 0.41 were reported. Furthermore, UK data applied to each of the three Groupers, yielded an overall R^2 of around 0.45.

However, a number of data deficiencies in the dataset used for this research need to be highlighted to illustrate the need for improvements in coding practices. Currently, coding at MDH is not carried out to fulfil the requirement of setting up and maintaining a DRG system. Therefore, shifting management thinking towards establishing such an objective would significantly help improve the data coding structures at the hospital. Furthermore, a significant proportion of available data (around 40%) lacked some important coding information. The success of a DRG system heavily depends on the quality of the coded data and the provision of additional years of data together with an improved coded dataset would help improve the accuracy of DRG Grouper outputs. The introduction of a well-defined purpose for current coding work would help improve the level of coding and ensure a more pragmatic picture of the activity levels within the hospital.

2.8 Conclusion

A health care Grouper software is considered to be useful if it partitions the hospital episode population in an informative way, both clinically in terms of inputs required and also in terms of the resources that are expected to be consumed (Aisbett et al., 2007). Homogeneous DRG categories provide knowledge to the policy maker on the average utilisation rate of episodes of care and serve as a good benchmark for comparison and evaluation purposes.

The *CV* and the R^2 coefficients obtained as a result of this study provide a suitable basis for recommending the use of DRGs in the Maltese health care system. The introduction of a DRG classification system can serve to explain some of the variation in the resource use within the hospital and is a positive step to further helping the management of resource use within the health care sector. The formulation of DRGs would encourage administrators to view the use and costs of hospital services along product lines based on DRGs and provide information on whether resources used for particular episodes of care are in line with what, on average, is expected. This therefore is an added tool, available to policy makers, for the better use of hospital resources. Furthermore, the implementation of a DRG system, apart from introducing a new purpose for the accurate collection of patient data, would also serve to make available to the hospital management authorities a greater range of information to control and monitor activities undertaken within the hospital.

This study evaluated the use of a casemix system to describe hospital activity in Malta. Viewing the output of MDH through the 'DRG lens' gives a clearer picture of what is being produced as a result of the investment made in this important and growing sector of the economy. Policy makers require accurate information on hospital activity to manage limited resources and monitor health care providers behaviour effectively (Ghaffari et al., 2008). This study concludes that one 'particular' Grouper software can be applied to the currently available data for the Maltese health care sector with relatively good results.

Experience has shown that the introduction of such as system into a country is a long and complex process requiring a change in the political and cultural setting of the hospital services. Results from this study show that a good foundation exists to start such a process and the creation of DRGs will be of benefit for management and implementation of policy in the Maltese health care sector. The use of DRGs for financing purposes might still be far away but the introduction of this system for policy and analysis purposes is to be encouraged. As a start, the information gained from the setting of a DRG system could serve as an instrument for the allocation of public hospital budgets funded by the national government.

In terms of further work, a study on the adoption of the grouping methodology to describe hospital output could be extended into a number of areas, initially by carrying out a comparison of the above results to those provided by other Grouper methodologies already applied in other countries. The application of the data to different Grouper software used by other countries is to be encouraged to identify the applicability and readiness of the Maltese data and thus ensure realibility and consistency in the obtained results. Furthermore, the availability of additional years of data and the expected improvement in the data recording systems should provide an added stimulus in realizing the benefits of such casemix systems. The statistical tools applied in this paper show that the results obtained are comparable to those observed in other countries, and that these methods can indeed be used to analyze the Maltese health care sector.

Considerable care should be taken in the interpretation of the results given the relatively small size of the dataset used in this study. A more detailed analysis of the outlier cases and their impact on ALOS and resource use would further improve the understanding of hospital output. Indeed, this study has shown that trimming options do affect the results and that other factors might have to be controlled for to fully understand differences in resource utilisation within hospital DRGs. Further work on the variation in outcomes within a particular defined DRG category, due to differences in consultant and surgeon volume levels is described in later chapters of this thesis. Variation in outcomes and practice patterns arising from differences in consultant job contract conditions are also presented.

The attention of governments to casemix use has been stimulated by the concerns of escalating health expenditure and the demand for transparent and efficient utilisation of resources in a period of increasing pressure to control health care spending. Wilke et al., (2001) stress that every medical relevance relating to illness carries with it an economic relevance and thus there is a growing need to define valid and reliable measures of output. This research is a first step in the laying down of an essential building block for the better understanding of the output currently being produced within the health care sector in Malta. DRG classification increases transparency of operations and projects a clearer picture of the hospital's activities. Defining the product categories of hospitals is a first and necessary important step. The introduction of a DRG system is also expected to have an effect on clinical behaviour and therefore have an effect on deviations from the optimum level of medical treatment. The introduction of a DRG classification system would serve to create an economic underpinning to management decisions taken within the hospital setting. All of these factors are ultimately expected to influence the overall outcome of medical interventions.

References

Aisbett, C., Wiley, M., McCarthy, B., and Mulligan, A. (2007). 'Measuring Hospital Case Mix: Evaluation of Alternative Approaches for the Irish Hospital System', ESRI Working Paper, no 192.

Anderson, A. (1984). 'Case mix methodologies: an overview.', In: Health Research and Educational Trust of New Jersey. *DRG evaluation (VolII). Case-Mix classification, Data and Management*. Princeton, pp. 1-22.

Aparo, U.L., Lorenzoni, L., Da Cas, R., Ncolai, P., Cristofani, G., and Puddu, P. (2009). 'An analysis of hospital productivity', *Casemix Quarterly,* vol.1(3).

Averill, R.F., Vertrees, J.C., Goldfield, N.I., Mullin, R.I., Fineran, E.C., Zhang, Z., Steinbeck, B., and Grant, T. (1998). 'The Evolution of Casemix Measurement Using Diagnosis Related Groups (DRGs)', 3M HIS Research Report, pg 5-98.

Bender, J.A. and McGuire, T.E. (1995). 'A focused look at the L3H3 exception policy', Proceedings of the 10th Patient Classification Systems/Europe Conference.

Benton, P.L., Evans, H., Light, S.M., Mountney, L.M., Sanderson, H.F., and Anthony, P. (1998). 'The development of Healthcare Resource Groups - version 3', *Journal of Public Health,* vol.20(3), pp. 351-358.

Berki, S. (1985). 'DRGs, Incentives, Hospitals and Physicians', *Health Affaris,* vol.Winter, pp. 70-82.

Bibbee, A. and Padrini, F. (2006). 'Balancing health care quality and cost-containment: the case of Norway',OECD working paper 481.

Busse, R., Schreyogg, J., and Smith, P. (2006). 'Editorial: Hospital case payment systems in Europe', *Health Care Management Science,* vol.9(3), pp. 211-213.

Casas, M. and Tomas, R. (1993). 'Diagnosis-Related Groups in Europe: Uses and Perspectives'. Springer-Verlag.

Closon, M.-C. and Roger, F. (1989). 'Case-mix and severity of cases for hospital financing and quality assessment', In: F. H. Roger-France, G. De Moor, J. Hofdijk, L. Jenkins (eds.): *Diagnosis Related Groups in Europe*. Ghent, Goff BVBA, pp. 43-55.

Coombes, J., Eckstein, G., and Gomes, R. (1995). 'Determination of AN-DRG trim points for NSW acute hospitals', Callaghan, (N.S.W.), Health Services Research Group, University of Newcastle.

Cots, F., Elvira, D., Castells, X., and Saez, M. (2003). 'Relevance of Outlier Cases in Case Mix Systems and Evaluation of Trimming Methods', *Health Care Management Science,* vol.6(1), pp. 27-35.

Darzi, A. (2008). 'High Quality Care for All: NHS Next Stage Review Final Report'. Department of Health (UK).

Dismuke, C. and Sena, V. (1999). 'Has DRG payment influenced the technical efficiency and productivity of diagnostic technologies in Portughese public hospitals? An empirical analysis using parametric and non parametric methods.', *Health Care Management Science,* vol.2(2), pp. 107-116.

Fetter, R.B., Shin, Y., Freeman, J.L., Averill, R.F., and Thompson, J.D. (1980). 'Case mix definition by diagnosis-related groups', *Medical Care,* vol.18(2 Suppl):iii, pp. 1-53.

Finger, A. (2000). 'Medicare's new red flag: same day readmissions.', *Medical economics,* vol.77(13), pp. 41-49.

Fischer, W. (2000). 'Are DRGs homogeneous with regards to resource consumption?'. 16th PCS/E proceedings, Groningen.

Ghaffari, S., Jackson, T.J., Doran, M., Wilson, A., and Aisbett, C. (2008). 'Describing Iranian hospital activity using Australian Refined DRGs: A case study of the Iranian Social Security Organisation', *Health policy (Amsterdam, Netherlands),* vol.87(1), pp. 63-71.

HOPE. (2006). 'DRGs as a financing tool', European Hospital and Healthcare Foundation, available at http://www.hope.ne/05eventsandpubications/ docpublications/77_drg_report/77_drg_report_2006.pdf (Accessed 12 December 2011).

Jonsson, B. (1996). 'Cost sharing for pharmceuticals', *Pharmacoeconomics,* vol.10(Suppliment 2), pp. 68-74.

Kahn, K.L., Keeler, E.B., Sherwood, M.J., Rogers, W.H., Draper, D., Bentow, S.S., Reinisch, E.J., Rubenstein, L.V., Kosecoff, J., and Brook, R.H. (1990). 'Comparing Outcomes of Care Before and After Implementation of the DRG-Based Prospective Payment System', *JAMA: The Journal of the American Medical Association,* vol.264(15), pp. 1984-1988.

Lichtig, L.K. (1986). 'Hospital information systems for casemix managment'. New York, Wiley.

Louis, D.Z., Yuen, E.J., Braga, M., Cicchetti, A., Rabinowitz, C., Laine, C., and Gonnella, J.S. (1999). 'Impact of a DRG-based hospital financing system on quality and outcomes of care in Italy', *Health Services Research.,* vol.34(1 Pt 2), pp. 405-415.

Mikkola, H., Keskimaki, I., and Hakkinen, U. (2002). 'DRG-related prices applied in a public health care system--can Finland learn from Norway and Sweden?', *Health policy (Amsterdam, Netherlands),* vol.59(1), pp. 37-51.

Ministry of Health - Government of Malta. (2008). 'National Report on Strategies for social protection and social inclusion 2008-2010'.

Moreno-Serra, R. and Wagstaff, A. (2010). 'System wide impacts of hospital payment reforms', *Journal of Health Economics,* vol.29(4), pp. 585-602.

Morrissey, M.A., Sloan, F.R., and Valvona, J. (1988). 'Shifting Medicare patients out of the hospital', *Health Affairs,* (Winter 1988), pp. 52-64.

Palmer, G., Reid, B., and Aisbett, C. (2001). 'Evaluation of the performance of diagnosis-related groups and similar casemix systems: methodological issues', *Health Services Management Research,* vol.14(2), pp. 71-81.

Reid, B., Palmer, G., and Aisbett, C. (1991). 'Choosing a DRG grouper for Australia: issues and options', *Australian Health Review,* vol.14(3), pp. 285-300.

Reid, B., Palmer, G., and Aisbett, C. (1997). 'Evaluating the performance of Australian National Diagnosis Related Groups', *Australian Health Review,* vol.23(2), pp. 20-31.

Reid, B. and Sutch, S. (2008). 'Comparing diagnosis-related group systems to identify design improvements', *Health policy (Amsterdam, Netherlands),* vol.87(1), pp. 82-91.

Rhodes, G., Wiley, M., Tomas, R., Casas, M., and Leidl, R. (1997). 'Comparing EU hospital efficiency using diagnosis-related groups', *The European Journal of Public Health,* vol.7(suppl_3), pp. 42-50.

Scheller-Kreinsen, D., Geissler, A., and Busse, R. (2009). 'The ABC of DRGs', *The Health Policy Bulletin,* vol.11(4), European Observatory of Health Systems and Policies, pp. 1-12.

Sutch, S. and Reid, B. (2003). 'An assessment of the performance of Healthcare Resource Groups', Proceeding of the 19th international casemix conference, pp. 261-266.

Wiley, M. (1999). 'Development and localisation of casemix applications for inpatient hospital activity in EU member states', *Australian Health Review,* vol.22(2), pp. 69-85.

Wilke, M.H., Hocherl, E.F.J., Scherer, J., and Janke, L. (2001). 'Introducing the new DRG-based payment system in German hospitals: a difficult operation?', *Health Economics in Prevention and Care,* vol.2(2), pp. 79-85.

Zhiping, G., Duckett, S.J., Legge, D.G., and Pei, L. (2004a). 'Describing Chinese hospital activity with diagnosis related groups (DRGs): A case study in Chengdu', *Health policy (Amsterdam, Netherlands),* vol.69(1), pp. 93-100.

Zhiping, G., Qin, Z., Wei, S., Zhengzhong, M., Jialin, J., and Minhui, H. (2004b). 'Study on feasibility of setting up DRG payment in China' EADN working paper no27.

Appendix

Table 2.8: A description of the MDC categories

MDC CODE	MDC DESCRIPTION
00	UNGROUPABLE
01	NERVOUS SYSTEM, DISEASES & DISORDERS
02	EYE, DISEASES & DISORDERS
03	EAR, NOSE, MOUTH, & THROAT, DISEASES & DISORDERS
04	RESPIRATORY SYSTEM, DISEASES & DISORDERS
05	CIRCULATORY SYSTEM, DISEASES & DISORDERS
06	DIGESTIVE SYSTEM, DISEASES & DISORDERS
07	HEPATOBILIARY SYSTEM & PANCREAS, DISEASES & DISORDERS
08	MUSCULOSKELETAL SYSTEM & CONNECTIVE TISSUE, DISEASES & DISORDERS
09	SKIN, SUBCUTANEOUS TISSUE & BREAST, DISEASES & DISORDERS
10	ENDOCRINE, NUTRITIONAL, AND METABOLIC, DISEASES & DISORDERS
11	KIDNEY AND URINARY TRACT, DISEASES & DISORDERS
12	MALE REPRODUCTIVE SYSTEM, DISEASES & DISORDERS
13	FEMALE REPRODUCTIVE SYSTEM, DISEASES & DISORDERS
14	PREGNANCY, CHILDBIRTH, & THE PUERPERIUM
15	NEWBORNS AND NEONATE CONDITIONS BEGAN IN PERINATAL PERIOD
16	BLOOD, BLOOD FORMING ORGANS, IMMUNOLOGICAL, DISEASES & DISORDERS
17	MYELOPROLIFERATIVE DISEASES & POORLY DIFFERENTIATED NEOPLASMS
18	INFECTIOUS & PARASITIC DISEASES
19	MENTAL DISEASES & DISORDERS
20	ALCOHOLDRUG USE AND ALCOHOLDRUG INDUCED ORGANIC MENTAL DISEASES
21	INJURIES, POISONINGS, AND TOXIC EFFECTS OF DRUGS
22	BURNS
23	FACTORS ON HEALTH STATUS & OTHER CONTACTS WITH HEALTH SERVICES
24	MULTIPLE SIGNFICANT TRAUMA
25	HUMAN IMMUNODEFICIENCY VIRUS INFECTIONS

Source: MS-DRG Grouper manual, Centers for Medicare and Medicaid Services, USA

Chapter

3

Volume and Competing Risk Outcomes under Different Consultant Contract Types: the Case of Malta

Abstract

This paper provides insight into the relationship between volume and outcomes, within the context of competing risk events. The majority of the literature on this subject has focused on analysing the volume-outcome relationship in a context of a number of hospitals but little is known of this relationship in the case of a single provider hospital. Therefore, the primary focus of this paper is to examine the prevalence of the volume-outcome relationship within the context of a single and unitary general public hospital. There is also the uncertainty of the nature of the hospital/physician volume effect and the channels through which such an effect might operate. In addition, this paper will contribute to the literature by analysing the role of the consultant job contract type and the impact that this has on the volume-outcome relationship.

Data from the Maltese health care system are evaluated from patients who underwent Percutaneous Transluminal Coronary Angioplasty (PTCA) in the period 2009 to 2011. The main event of interest is *failure* within the first 60 day period following the undertaking of the PTCA procedure, where *readmission* or *death* over the same period are treated as possible failure events. The procedure is deemed a *success* if no event occurs within the 60 day period following the undertaking of the PTCA. The cumulative incidence function for the event of interest is used in this study to estimate the competing risk model. These results are compared with other estimates obtained using the Kaplan-Meier method and the cause specific hazard function. A multinomial logistic

model is estimated to gauge for the robustness of the obtained results based on the survival analysis methods.

The results of the study support the view that there is some evidence of a difference in the hazard rate for the event of failure from the PTCA procedure which is dependent upon the volume of patients seen by consultants and surgeons and also by the job contract type of the consultants. The study finds that there is a lower risk of failure for patients who were under the care of consultants who practice exclusively within the public sector compared to those being under the care of *dual practice* contract consultants. There is also a higher hazard of treatment failure from the procedure if the patient is under the care of consultants with high patient volume levels or surgeons with low patient volume levels.

Furthermore, an increase in the volume of patients seen by consultants has a different impact on the hazard of treatment failure depending on whether the patient is being seen by consultants on *public-only* or *dual practice* contracts. As volume levels for consultants on *public-only* contracts increases the hazard of failing from the procedure increases. The hazard of failing from the procedure for patients under the care of consultants on *dual practice* contracts remains relatively unchanged, even when volume levels change.

The above conclusions are an important contribution towards understanding the relationship between volume and competing risk outcomes in the context of a single and unitary public hospital setting.

3.1 Introduction and Motivation

The investigation of the relationship between volume and survival outcome in the provision of health care services has attracted a lot of interest from researchers over the years. The majority of the literature finds a positive relationship between both hospital/physician volume and health care outcomes (Gandjour et al., 2003; Halm et al., 2000). Dudley (2000) states that the relationship between outcome and volume is so common and consistent over time that one would expect it always to hold in practice. Furthermore, if the procedure performed is complex then success is expected to be more dependent on the skill of the particular surgeon (Srinivas et al., 2006). There is though uncertainty as to the nature of this hospital/physician volume effect and very little is known as to the channels through which such an effect might operate. In particular, two different explanations have been proposed to explain the volume-outcome relationship: the *practice makes perfect theory* or the *selective referral theory, as* proposed by Luft et al., (1987).

The majority of the literature has focused on analysing the volume-outcome relationship in a context of competing hospitals but little is known of this relationship in the case of a single provider hospital. The primary focus of this study is therefore to understand the volume-outcome relationship within the context of a single and unitary general public hospital. This study contributes to the literature by analysing the role played by the consultant job contract type and the impact which this has on the volume-outcome relationship.

In line with the practice makes perfect hypothesis, the higher the volume levels of the consultant the higher would be the chance of outcome success following a hospital procedure. Furthermore, based on the selective referral hypothesis, consultants with favourable outcomes are more likely to have higher volume levels. Consultants working with the Government hospital are given an option to choose between two work plan types; *public-only,* whereby consultants cannot undertake any private practice and *dual practice* whereby consultants can practice in both the private and the public sector. Data on the job plan contract type of each of the consultants is provided by the hospital management for the purposes of this study. Consultants working only at the public hospital have all the necessary infrastructure to perform at the highest levels, whilst consultants on *dual practice* contracts have an incentive to achieve success in view of the expected impact on their private practice reputation given their performance within the public hospital.

Differences in the probability of achieving an event success or a failure event are thus studied in relation to the volume and job contract type of the consultant. A difference in the responsiveness of outcomes to volume changes is expected to exist in relation to the particular consultant job contract type. This difference might be related to the current capacity levels at which the particular consultants are operating at and/or to the ability of the particular consultant to alter his performance at the hospital.

To this end, the volume-outcome relationship will be evaluated in the context of competing events that follow the procedure for Percutaneous Transluminal Coronary

Angioplasty[1] (PTCA) after controlling for a number of characteristics pertaining to the consultants, surgeons and to the patients undergoing treatment. The study will review the role of a number of different volume measures and how each of these impact PTCA outcomes, taking possible competing risk events into account.

The study will use data from the Maltese health care system, which has one large publically funded acute general hospital, the Mater Dei Hospital, which provides free services at the point of use. In view of the available data, a distinction will be made between the *consultant,* who is primarily responsible for the patient case, and the *surgeon* who actually undertakes the procedure. Patients are assigned to consultants on the basis of the diagnosis made and the date when they are admitted to hospital.

Patients have little choice of the hospital they can attend, although they can decide to 'opt out' and visit a private hospital/clinic[2] if they so wish. The majority of private hospitals are run and staffed by the same consultants and surgeons who provide services at the public hospital.

The data cover the years 2009 to 2011 and are compiled on a patient level basis specifically for this study. The outcome effectiveness of undertaking PTCA is measured by the complications-free or event-free survival probability. The presence or occurrence of an event which is in competing risk to the event of interest, may preclude the incidence of the event of interest or may alter the probability of occurrence for this

[1] PTCA is a common cardiac procedure to open up partially blocked arteries in the heart. The physician threads a catheter through a large artery in the patient's arm or leg and into the artery which leads to the heart.

[2] A number of small private clinics/hospitals can be found in Malta. Activity levels in such sectors are relatively low and activity is serviced by a small private health insurance market.

event. The main event of interest is *failure* within the 60 day period following the PTCA procedure, where *readmission* or *death* within the 60 day period are treated as two possible failure events. The choice of the 60 day time period was based on the advice provided by the Clinical Performance Unit (CPU) of the hospital. This time period was considered to best reflect the time frame which captures procedural related complications and activity which even though might not be necessarily arising from procedure related complications is still considered to be related to the needs of the same patients undertaking the PTCA procedure.

According to the literature, the Cumulative Incidence Function (CIF) can be used to model competing risk events (Gooley et al., 1999; Prentice et al., 1978; Satagopan et al., 2004; Schwarzer et al., 2001). Other estimation techniques like the Kaplan-Meier (KM) method applied by Tai et al., (2001) and the Cause Specific Hazard Function (CSHF), used by Bakoyannis and Touloumi (2011), will also be applied for comparison purposes. Furthermore, to test the robustness of the results obtained from the survival analysis methods a multinomial logistic model of outcomes is also estimated.[3]

The results presented in this paper will show some evidence of difference in the hazard rate for the event of failure from the procedure which depends on the volume of patients seen by consultants and surgeons and also by the job plan contract type of the consultants. The study finds that an increase in volume has a different impact on the hazard rate of the event of interest depending on whether the patient is being seen by a consultant on a *public-only* or a *dual practice* job contract type.

---

[3] A detailed description of the results can be found in Appendix B to this chapter.

This paper is organized as follows: the following section, Section 3.2, provides an overview of the literature that relates to the volume-outcome relationship and gives special reference to competing risk events. Section 3.3 describes the available population dataset and highlights some preliminary findings. An overview of the methodologies for analysing competing risks data in the context of the possible strategies to model the covariates influencing outcomes is presented in Section 3.4 and Section 3.5 presents the key results of the study. Finally, Section 3.6 presents the study's conclusions.

3.2 Literature

Whilst the volume-outcome relationship is by no means universally accepted, it has in the literature been observed extensively in a variety of procedures. This section presents an overview of the main theories underpinning this relationship focusing on the fact that most studies distinguish between a volume-outcome relationship either at the hospital or at the physician level. Are doctors more effective in bigger hospitals with relatively high volumes? Do particular characteristics, related to the doctor, and do the working conditions of physicians affect the volume-outcome relationship? Answering these questions is crucial for the drafting of health care policy.

While it may be difficult to compare findings across a number of studies there does seem to be consensus among researchers that a positive relationship exists between procedure volume (at the level of the hospital and/or physician) and provider quality, as measured by a variety of patient outcomes. An extensive literature review undertaken by Halm et al., (2000) and Gandjour et al., (2003) supports the existence of a volume effect in most diseases and operation categories. At the end of this section a review is provided of the different measurement options of outcome applied in this area of study.

Theoretical underpinnings: The pioneering work by Luft et al., (1979) presents two hypotheses for the relationship between volume and outcome: *the practice makes perfect theory* and the *selective referral theory*. The practice makes perfect theory assumes that as physicians and/or hospitals perform more operations, they become more skilful and thus their positive outcome rates would be expected to improve. The performance of the physician is said to depend on the individual skills of the physician, the resources available to the physician at the hospital level and the 'spill-over' impact of both effects. Srinivas et al., (2006) conclude that for primary Percutaneous Coronary Intervention (PCI), physician experience, measured by the frequency levels of performing a certain task, modifies the hospital volume-outcome relationship and serves to offset the risks associated with the treatment of primary PCI in low volume hospitals. Hamad et al., (1988) show that in a single hospital setting, both low frequency and high frequency physicians can perform PTCA with success but for more complex cases, outcomes are likely to be better when the procedure is undertaken by more experienced physicians.

Gowrisankaran et al., (2006) found that there was also an element of *forgetting* when explaining the volume-outcome relationship given that the success rate of the physician may deteriorate when they do not perform regularly. Huesch and Sakakibara (2009) proposed that it's the most recent 'on the job' experience that is the most important factor affecting the outcome of the intervention. Furthermore, Ramanarayanan (2008) finds that the experience and competence gained by cardiac surgeons carrying out CABG[4] is not limited to the hospital in which the physician practises. Indeed, the outcomes of a physician in a particular hospital are said to improve from the added volume experience of practicing in another hospital or health care setting. This contrasts with earlier findings of Huckman and Pisano (2006) whereby improvements in performance are related to experience gained within the same hospital.

On the other hand, physicians and hospitals with better outcomes are likely to receive more referrals and more of these referrals may be more appropriate patients for certain procedures. Patients, if they have the option, are also inclined to show their preference for institutions or individuals who provide better outcomes, therefore suggesting the potential for selective referral. Indeed, Hamilton and Hamilton (1997) found that the volume-outcome relationship for hip fracture patients in Quebec hospitals reflected differences between hospitals that are fixed over time. This result is consistent with the selective referral hypothesis, identified by a number of authors, whereby high quality hospitals are likely to attract more patients (Farley and Ozminkowski, 1992; Hamilton and Hamilton, 1997; Hamilton and Ho, 1998). Tsai et al., (2006) obtain similar results when they studied the effect of hospital volume on the 30 day mortality of patients with congestive heart failure, suggesting that causality runs from outcomes

[4] Coronary Artery Bypass Graft.

to volume. It is thus the personal characteristics of the physician and the hospital which explain the impact of volume on outcome.

Hospital and/or physician volume: The establishment of a hospital volume threshold level tends to affect outcomes. Ritchie et al., (1993) use the rates of CABG interventions performed during the same hospital stay following a PTCA procedure, as a measure of outcome success and find that it is lower in hospitals in which more than 400 PTCAs are undertaken annually. Both hospital PTCA volume and cardiologist PTCA volume are found to be inversely related to a negative outcome. Shook et al., (1996) concludes that PTCA performed by high-volume operators was less likely to require emergency CABG and was also significantly associated with lower hospital morbidity and costs and shorter hospital Length of Stay (LOS). Of particular relevance is work by Hannan (1997) on PTCA patients and Jollis et al., (1994,1997) for acute myocardial infarction and coronary angioplasty who conclude that volume and short-term complications are inversely related to a cardiologist performing more than 75 cases per year.[5]

Patients treated by high volume physicians (after adjusting for potential confounding factors) have a much lower risk of mortality and of surgical complications associated with the treatment undertaken. This is the conclusion of both Billingsley et al., (2007) and Vakili et al., (2001). Furthermore, the risk of mortality is further reduced if such patients are treated in a high volume hospital. One would expect though that an

[5] The American College of Cardiology suggest hospitals undertaking a minimum of 200 PTCA procedures per year and an individual surgeon with a minimum of 75 PTCA cases per year.

increase in volume has a greater impact on the outcome of a small-volume hospital when compared to a large volume hospital.

The relevance of the volume-outcome relationship in a setting of a single hospital centre is less well studied in the literature. Ross et al., (2010) confirms a diminishing effect on volume beyond a particular threshold level for three specific conditions: acute myocardial infarction, heart failure and pneumonia. In a hospital setting where volume levels differ across physicians, Kawabuchi and Sugihara (2006) find a non-linear volume effect whereby the principle of 'the more the better', applies only up to a certain volume level. Congestion effects are then likely to set in.

By focusing on a publically funded health care system, Simunovic et al., (1999) found that patients treated by high volume hospitals had a better outcome when treated for pancreatic resection. There is emphasis on the fact that this relationship holds not only in the mixed public-private hospitals in the US, as studied by Glasgow and Mulvihill, (1996), Gordon et al., (1995) and Lieberman et al., (1995), but also in the publically funded hospital setting. In a public setting there is expected to be less financial and logistical barriers to care and better outcomes are linked not only to higher volume levels but also to better expertise and resources available in such public institutions. The mechanism by which physician and provider characteristics influence outcomes in such a setting is also expected to differ from other hospital settings.

It is evident that a certain amount of interaction between hospital volume and physician practice volume exists and that this impacts on health care outcomes. Although early studies like Kelly and Hellinger (1986) find that the volume-outcome relationship reflects hospital rather than physician characteristics, a more recent study by Birkmeyer et al., (2003) observed that links between hospital volume and operative mortality are largely influenced by surgeon volume, whereby the chance of survival, even in high volume hospitals, will most likely increase if treated by a surgeon with high volumes. In 135 studies reviewed by Halm et al., (2000), a statistically significant relationship between higher volume and better outcomes was found for 71% of hospital volume studies verses 69% for clinician volume studies. Other more recent studies assess simultaneously the impact of hospital and surgeon volume on outcomes and find that both high hospital and surgeon volume are linked to lower in-hospital mortality (Hannan et al., 2005; Joynt et al., 2011; Schrag et al., 2002; Schrag et al., 2003).

Indeed, particular physician preferences in the treatment of cases are expected to have an impact on outcomes. Being treated by a specialist in the field or by a more *aggressive* surgeon might have an impact on patient outcomes. A *conservative* surgeon might opt for limited surgery, making a compromise between control of symptoms and ultimate survival, whilst an *aggressive* surgeon might perform more high risk surgery, increasing the risk of complications in an attempt to improve on the chance of survival. McArdle and Hole (1991) reported an improvement in survival rates after colorectal surgery if the intervention was carried out by surgeons who had a special interest in the field.

Advances in medicine and technology are also expected to have an impact on the strength of the relationship between outcome and volume levels. Ho (2000, 2002) measured improvements in PTCA outcomes by reducing inpatient mortality and emergency bypass procedures and found that over time, the disparity in outcome between low and high volume hospitals has narrowed. The developments in technology may have reduced or eliminated the need to perform a high volume of procedures in order to obtain optimal outcomes. Even though technology has by far improved the outcome within hospitals, the impact of volume both in terms of physicians and hospitals has still been found to be important. McGrath et al., (2000) find that since the development of coronary stents, patients treated at high volume hospitals and by high volume physicians still tend to have better outcomes following a PCI. Patients treated by high volume physicians are less likely to require CABG emergency treatment and have a lower risk of mortality within 30 days of intervention. The impact of volume on outcome is therefore still strong and significant.

Measurement of outcomes: All possible outcomes need to be taken into account when considering the health consequences of a medical intervention. Varadhan et al., (2010) demonstrate that any analysis into the impact of an intervention within the health care sector must consider the full set of possible outcomes which can arise following the intervention. A competing risk is defined by Gooley et al., (1999) as "an event whose occurrence either precludes the occurrence of another event under examination or fundamentally alters the probability of occurrence of this other event". Competing risks are a common occurrence in medical research and the use of survival analysis techniques allow for the incorporation of the impact of all possible outcomes on the incidence of the particular event of interest.

The outcome of care, as well as being measured by in-hospital mortality, should also focus on post-operative conditions together with any occurrences of complications which result from the treatment undertaken. The outcome of the procedure performed can be viewed in terms of the state of health of the patient prior to and post the procedure. The study of outcomes, following a PTCA procedure, should not stop when the patient is discharged from hospital and survival analysis allow us to study and incorporate this time dimension into the analysis. The choice of the outcome variable also has an impact on the assessment made in relation to the success or failure of the intervention. Kawabuchi and Sugihara (2006) use death hazard ratios as a measure of outcome and conclude that there is no relationship between outcomes and PTCA hospital volume. This confirms a previous study by Kimmel et al., (2002) which found no association between volume and 30 day post-discharge events.

Since the seminal articles by Fix and Neyman (1951) and Prentice et al., (1978), the volume-outcome relationship in the context of competing events has been studied in the literature through the use of a number of survival analysis techniques mainly; the standard KM method, the CSHF and the CIF. In the presence of competing events, patients are at risk of more than one mutually exclusive event and the use of the CIF is encouraged by a number of authors (Andersen et al., 2012; Arriagada et al., 1992; Furstova and Valenta, 2011; Gaynor et al., 1993; Gooley et al., 1999; Kim, 2007; Melberg et al., 2010; Satagopan et al., 2004; Southern et al., 2006; Teixeira et al., 2013). In the case of more than one type of event (or failure) being present, if such events are somehow dependent on each other, then the KM estimate is found to be biased because the KM method censors all other events of interest as it assumes that they are independent from the primary event of interest (Arriagada et al., 1992; Tai et al., 2001). This was confirmed by Melberg et al., (2010) within the setting of a single hospital by studying the impact of competing risk events on the outcome of patients undergoing revascularization for coronary artery disease over a 10 year period.

Klein (2006) argues that in the study of competing risks of relapse and death in remission for bone marrow transplantation treatments, the CIF provides a natural way to identify the effect of covariates on the failure of treatment. When competing risks events are present the use of survival analysis methods is encouraged and the use of the CIF is recognised in the medical and statistical literature as the right tool to use (Gooley et al., 1999). Some authors make use of both the CSHF and the CIF function based on the proportional sub-distribution hazards model to evaluate the effect of covariates on the event of interest (Beyersmann et al., 2007; Teixeira et al., 2013).

Cumulative incidence models measure the effect of the covariate on the specific event cumulative probability, and this covariate affect may be derived from the direct effect of making the event of interest more or less likely to occur, or the indirect effect of making competing events more or less likely to occur. The CIF allows for the identification of risk factors for each of the competing events. On the other hand, the standard Cox hazard model is not designed to deal with the risk factors which contribute to a particular event in the presence of a competing risk. The CSHF ignores competing risks and thus could be inaccurate for this reason. The use of the CIF model is questioned though by Anderson et al., (2012) who argue that individuals who failed due to a competing risk should not be in the risk set used to estimate the sub-distribution function for the event of interest.

3.3 Data

The data used in this population based study comes from the Clinical Performance Unit (CPU) which is the unit responsible for collecting all clinical data for activity undertaken within the Mater Dei Hospital[6] (MDH). The unit of analysis is the patient and the analysis is performed for patients who entered hospital and underwent a PTCA procedure in the period 2009 to 2011.

The following ICD-9-CM[7] procedure codes where considered in this study: 0066, 3601, 3602, 3604, 3606, 3609, 0045, 0046, 0047 and 8856. The dataset contains 1,626 hospital visits and identifies both the *surgeon* who carried out the procedure and the *consultant* who was responsible for the individual patient care for each PTCA patient. The dataset provided the information on patient, surgeon and consultant characteristics used to identify the main covariates of the study. Table 3.1 presents a list of the principle variables used. Of particular relevance to this study is the information on the job plan contract type of each of the consultants seeing patients undertaking a PTCA procedure[8]. Consultants working with the Government hospital operate either under a, *public-only,* or a *dual practice* contract.

[6] Mater Dei Hospital is the only general acute public hospital in Malta.
[7] International Classification of Diseases – 9th revision – Clinical Modification (WHO).
[8] Data on the contract type of each of the surgeons is not provided for this study.

Table 3.1: Definition and description of the principal variables

Variable	Description of variable
General variables	
Surgeon	The professional responsible for carrying out the procedure
consultant	The professional responsible for the patient whilst in hospital for treatment
Age	Age of patient
LOS	The difference between the discharge and admittance date
diagtotal	Number of diagnoses per case
public-only	=1 for consultant exclusive public sector practice contracts =0 for consultant public and private practice contracts or *dual practice*
countconst	Consultant total number of cases
Outcome variables	
success	A procedure that is deemed a success if no other event occurs within the 60 day period following discharge date, or within the hospitalisation period
failure	Representing the event whereby a failure (either death or readmission) occurs within the 60 day period following discharge date, or within the hospitalisation period
death	Patient died within the 60 day period following discharge date and/or during the hospitalisation period
readmission	Patient readmitted to hospital within the 60 day period following discharge
Volume variables	
counts	A cumulative count of cases seen by the surgeon up until the date of admission of the patient
countc	A cumulative count of cases under the responsibility of a consultant up until date of patient admission
vol90s	Surgeon cumulative cases over the 90 days prior to the patient admission date
vol90c	Consultant cumulative cases over the 90 days prior to the patient admission date
cvol	=1 if countconst > 100 cases (*high* volume) =0 if countconst < 100 cases (*low* volume)
svol	=1 if surgeon volume > 225 cases (*high* volume) =0 if the surgeon volume < 225 cases (*low* volume)

Source: Based on data drawn from the provided hospital episode database.

3.3.1 Measures for outcome indicators

The event of interest in this study is the efficacy of the treatment, defined by the variable, *failure*, which is a measure of failing from the procedure within the 60 day period post the date of discharge or during the hospitalisation period. A patient may die, either during the hospitalisation period or within 60 days from discharge date, *death*, or be readmitted into hospital within 60 days after discharge, *readmission*. Both events would thus constitute an event *failure*. The time to the first observed event was calculated for each patient case and patient follow up was restricted by the latest available data[9] (December 2011).

PTCA related deaths are defined using the ICD-10 code provided from the Death Register made available by the Health Information Department within the Ministry of Health.[10] Deaths related to other causes are not considered given that the specific interest of the study is to look only at the impact of the PTCA procedure. Data is provided on a patient level basis and an encrypted identity card number allows for the identification of readmissions. The procedural codes of those readmitted are checked to establish whether the readmission event is due to the originally undertaken PTCA procedure. PTCA related readmissions are defined by observing the procedure codes[11] assigned to the readmitted patient cases. Patients might need to return to hospital because of other reasons (not related to the PTCA) but given that the focus is on the

---

[9] Data on the date of death were obtained from the Death Registry up to April 2012 and this was used to account for deaths post the 60 day discharge date.
[10] The cause of death codes (ICD-10) define deaths in relation to problems of the circulatory system.
[11] The procedural codes (0066, 3601, 3602, 3604, 3606, 3609, 0045, 0046, 0047 and 8856) are used to identify readmittance of patients due to PTCA related needs. An analysis of the data shows that there are no cases of patients who are readmitted to hospital and then die within the 60 day period.

specific PTCA procedure consequences the data used in the study only comprises patients readmitted for PTCA related factors.

3.3.2 Measures of volume

A number of volume measures are used in this study to test the robustness of volume indicators to explain the volume-outcome relationship. In particular, as highlighted by Ho (2002) and given the availability of the data, measures of cumulative PTCA volume by surgeon, *counts*, and by consultant, *countc*, are used as a measure of learning by experience. Alternative measures of volume, *vol90s* and *vol90c* are used to reflect the most recent activity levels (90 days) of the surgeon and the consultant preceding the date of patient admission to hospital. The dummy variables *cvol* and *svol* are constructed as indicators of the total volume levels of consultants and surgeons over the course of the three year activity within the hospital.

3.3.3 Description of data and preliminary evidence

Table 3.2 presents the principal characteristics of PTCA activity for the three years under study: volume levels, the mean age of patients, the Average Length of Stay (ALOS) and the volume and frequency rates for each of the events.

Table 3.2: Characteristics of PTCA activity

Year	No of cases	Mean age	ALOS	Event *success* cases	Event *death* cases	Event *readmission* cases
2009	447	62.03	3.98	402(89.9)	9 (2.0)	36(8.1)
2010	581	63.00	4.80	547(94.2)	12(2.1)	22(3.8)
2011	598	63.04	4.60	564(94.3)	7(1.2)	27(4.5)
Total	1626	62.75	4.50	1513(93.1)	28(1.7)	85(5.2)

Figures in parenthesis represent frequencies across rows.
Source: Analysis of hospital episode data.

It can be seen that the ALOS rises in 2010, compared to 2009, but falls marginally in 2011. A slight increase in the mean age of the patients and the total number of PTCA procedures undertaken over the three years is observed. The proportion of the event *success* increases over the period. The number of *death* and *readmission* cases registered, declines or is maintained at similar levels compared to 2009. The predominant cause of death is I21.9 (acute myocardial infarction) which accounted for more than 50% of the deaths falling within this category.[12] The average rate of the event *success* over the three year period is 93.1% of cases compared to less than a 2% rate for deaths and a rate of around 5% for re-admissions within the 60 day interval period.

3.3.4 Data descriptive statistics at the surgeon level

Table 3.3 presents details of the events of interest by volume of surgeon activity categorized into two groups: those dealing with less than 225 cases (defined as *low*)

[12] All other deaths fall within the 'Cerebrovascular' diseases category and the 'Ischaemic' heart disease category.

and those dealing with more than 225 cases (defined as *high*) over the three year period (undertaking more than 75 cases per year).

Table 3.3: Surgeons undertaking PTCA procedure (2009-2011)

Surgeon category	*volume of cases*	*success* cases	*death* cases	*readmission* cases	mean age	ALOS days
Low	471	90.7	2.1	7.2	60.4	5.4
High	1155	94.0	1.6	4.4	63.7	4.1
All	1626	93.1	1.7	5.2	62.7	4.5

Figures for the competing events represent relative percentages across the row. There are 3 surgeons with high volume levels and 9 with low volume levels.
Source: Analysis of hospital episode data.

Patients who are treated by *high* volume surgeons spend, on average, fewer days in hospital and the average age of such patients tends to be higher when compared to those treated by a surgeon in the *low* volume category. The proportion of success is highest among the *high* volume surgeon group suggesting a lower risk of failure for patients being seen by surgeons with high volume levels A patient seen by a *high* volume surgeon would on average have an overall lower risk of *death* or *readmission*, a lower overall LOS and a higher probability of a successful outcome.

3.3.5 Data descriptive statistics at the consultant level

Table 3.4 presents the frequency of events of interest by consultant with *low* or *high* patient volume levels. Consultants were also characterized by their job contract plan; *public-only* or *dual practice* contracts.[13] The data show that all consultants on *public-*

[13] Data was not available for the number of hours spent practicing in the private sector for *dual practice* contract consultants.

only contract arrangements have *low* patient volume levels. The mean age and ALOS for each category of consultants is also shown.

Table 3.4: Consultants responsible for the PTCA procedure (2009-2011)

Consultant	number of consultants	volume of cases	*success cases*	*death cases*	*readmission cases*	mean age	LOS days
All	35	1626	93.1	1.7	5.2	62.7	4.5
low	31	337	94.4	1.8	3.9	63.8	6.9
high	4	1289	92.7	1.7	5.6	62.5	3.9
public-only	4	87	97.7	-	2.3	61.6	5.8
dual practice	31	1539	92.7	1.8	5.4	62.8	4.4

Figures for the competing events represent relative percentages across the row. All *public-only* consultants fall within the low volume category.
Source: Analysis of hospital episode data.

Similar rates of death are observed for patients under the care of *low* and *high* volume consultants. Slightly higher rates of *success* are observed for patients under the care of *low* volume consultants. This result differs from that derived for surgeons and from that expected by the practice makes perfect hypothesis. A possible explanation could be the fact that consultants on exclusive public sector contracts have higher success rates compared to *dual practice* consultants and all consultants employed on a *public-only* contract fall within the *low* volume category. Consultants with *low* volume levels are perhaps better able to manage the cases under their responsibility which possibly leads to higher rates of success.

Patients under the care of consultants with *public-only* contracts have lower rates of *readmission* compared to patients under the care of consultants on *dual practice* contracts. Patient *readmission* rates are higher within the *high* volume group whilst the mean LOS is lower compared to the *low* volume group. This is evidence to suggest that there is a role for the consultant job contract type to explain volume-outcome relationships. There are no cases of *death* recorded within the category of patients under the care of consultants on exclusive public sector contracts.

3.4 Methodology

This section describes the empirical framework used to model competing risk events in relation to the patient, consultant and the surgeon characteristics, all of which are expected to have an impact on the event of interest.

The majority of the literature that deals with the use of competing risk events, concludes that the use of the KM method may not be the most appropriate method to use in such circumstances. Teixeira et al., (2013) and Kim (2007) note that the KM technique and its complement (*1-KM*) have been applied assuming that all other competing events, apart from the event of interest, are censored and thus are non-informative to the analysis of the event of interest. In such cases this would imply that such patients would still be in a position to get the event, given that they are treated as censored, and this would lead to overestimating the probability of failure from the event of interest.

The majority of the literature distinguishes between the use of the CSHF and the CIF to model competing risk data. Indeed, many studies, (Furstova and Valenta, 2011; Gooley et al., 1999; Kim, 2007; Klein, 2006; Melberg et al., 2010; Pintilie, 2007; Putter et al., 2007; Satagopan et al., 2004; Southern et al., 2006; Varadhan et al., 2010) recommend the use of the CIF to deal with issues of competing risk data given that such a function also incorporates relevant information arising from other competing risk events. This involves the estimation of the likelihood of one event taking into account information with respect to other events. Bakoyannis (2012) defines the cause specific hazard as "the instantaneous failure rate from a specific cause *i* given that no failure from any cause has yet occurred", treating observations with failures from all other causes but the event of interest as censored.

Subjects would be censored if they experience an event rather than the event of interest (competing risk). Thus for example when analysing the risk of getting an event readmission, subjects who get the event death would be treated as censored. Under the CSHF method, patients who achieve other events rather than the event of interest within the 60 day period are censored (this though allows them to achieve the event of interest in future days). Patients would also be censored if the study ends before the passage of the 60 day period and the event of interest does not occur.

The CIF is on the other hand based on a risk set which is not *natural* as it includes not only subjects who have not yet failed but also subjects who have failed from other causes. This proposition allows for a patient experiencing a competing risk event to be censored but in an informative manner. In estimating the CIF, the occurrence of a competing risk event is an event in itself and is thus informative to the determination of the probability of failing from the particular event of interest. This would ensure that the impact from each of the competing events on the event of interest is taken into account and that any effects which may arise from the relationships between the covariates of the model are included. The effect of covariates on the CIF for competing risks would be derived directly from the CIF. In view of such characteristics, the CIF has been recognised in both the medical and the statistical literature as the right tool to use in such circumstances (Gooley et al., 2001).

The KM and CSHF assume that censoring does not affect potential failure times. The assumption of independence of different event types may not be clinically meaningful, particularly in this study whereby the individual is studied over time and the events which occur following the patients' hospitalization are considered to be a result of the PTCA procedure performed. The time to the event of interest and the time to the competing events are correlated up to a certain degree. If patients, who have experienced a competing event are treated as censored then the possibility of achieving the event of interest for such patients would still exist, potentially leading to an overestimation in the probability of this event.

3.4.1 Modelling based on cause-specific hazards

The cause specific hazard is denoted by $h_j(t)$ which expresses the joint probability distribution of failure time T and failure C for a cause j in the knowledge that there are other possible competing events. The adopted function censors those who have experienced other events prior to the event of interest. The function also models the effect of covariate effects, X, on the cause specific hazard. The addition of factors X affecting the instantaneous failure rate from a specific cause of type j is given by the conditional cause specific hazard: $h_j(t/X)$

$$h_j(t/X) = \lim_{\Delta t \to 0} \frac{P(t \le T < t + \Delta t, C = j \mid T \ge t, X)}{\Delta t} \tag{1}$$

In addition to estimating the cause specific hazard of the event of interest, the CSHF across different groups is compared. It is of particular interest in the context of this study to look at patients under the care of consultants and surgeons with different volume levels and consultants with varying job contract conditions. Furthermore, the study incorporates the effect of covariates, such as the age of the patient, and the hospital LOS, on the event of interest. Within this framework, as discussed by Prentice et al., (1978) a semi-parametric Cox regression model (assuming proportional hazards) with a completely flexible unspecified baseline hazard is defined.[14] The effect of covariates on the cause specific hazard for cause *j* given covariates vector *X* is modelled by:

$$h_j(t/X) = h_{j,o}(t)\exp\left(B_j^T X\right) \tag{2}$$

The first term of the right hand side of the function expresses the baseline hazard (where all covariates are set to zero). The term $exp(B_j)$, represents the relative change in the cause specific hazard for the j^{th} cause due to a one unit increase in the corresponding covariate. This identifies any possible factors which could explain the occurrence of the event of interest in the presence of competing risks.

A Cox proportional hazards model is estimated separately for each competing event. As stated by Bakoyannis and Touloumi (2011), cause specific hazard functions, being

[14] The assumption of proportionality is tested as it is important to understand that there could be a changing effect of a covariate over time rather than a constant hazard assumption over time.

instantaneous risk functions, do not quantify the overall benefit or harm of an exposure (be it a treatment or a particular characteristic) to the patient. The inferences from this approach need to be evaluated cautiously because the assumption of independent competing events is strongly needed when estimating the CSHF. Covariate effects in the CSHF pertain to the event of interest only, without consideration to how the covariates act on competing events. The CSHF ignores the impact from competing risk events and thus could lead to inaccurate conclusions. The extent to which this model characterizes the covariate's influences solely on risk of the event of interest depends on how causes of failures may be inter-related. Interpreting the cause specific hazard as an increase or a decrease in apparent risk has to be treated with caution.

3.4.2 Modelling the cumulative incidence function

In view of the strong assumptions of independence between events to allow for the interpretation of the CSHF, an alternative measure of the distribution of risk; the sub-hazard distribution function is advocated in the literature. This serves to deal with the limitations raised by Klein (2006) that the CSHF does not directly model the effect of covariates on the CIF. Fine and Gray (1999) developed a model that builds on the Cox proportional hazards model and directly links the regression coefficients with the CIF whereby all competing events are assumed to have an impact on the coefficients. In such a case, there is a one-to-one relationship between the sub-hazard and the cumulative distribution. Furthermore, it can be noted that the effect of a covariate on the CSHF may differ from the effect on the CIF.

Competing risk models have the advantage of taking into account the fact that the presence of an event which is in competing risk to the event of interest, may preclude the incidence of the event under study and may thus alter the probability of occurrence for this event. It is thus more appropriate to estimate event rates by evaluating a CIF by taking into account other events within a competing risk framework. When competing risks are present then the way in which covariates are associated with the cause-specific hazard may not coincide with the way these covariates are associated with the cumulative incidence. A covariate that has no direct influence on the hazard of a primary event can still be significantly associated with the cumulative probability of that event, if the covariate influences the hazard of a competing event. Some effect will also arise from the association between covariates and the cause specific hazard for the competing events.

The CIF gives the probability of a subject failing due to a cause *j* in the presence of all the competing events before (or up to) time *t*:

$$F_j(t) = p(T < t, J = j) \quad (3)$$

The sub-distribution hazard for event *j* is defined as the probability for a subject to fail from cause *j* at a time *t*, given the subject experienced no failure prior to *t* or that if failure occurred, it is from another cause prior to *t*:

$$h_j(t / X) = \lim_{\Delta t \to 0} \frac{P((t \leq T < t + \Delta t, C = j \mid T \geq t) or (T < t, C \neq j), X)}{\Delta t} \quad (4)$$

The risk set is thus a *non-natural* set as it still includes subjects who fail from another cause apart from *j* and are thus not physically available to fail together with subjects who have not yet failed. The CIF is estimated for each of the events of interest whereby individuals experiencing competing events are still held within the risk set so that they can be adequately counted as having no chance of failing from the event of interest. The weights used to keep such subjects in the risk set decrease over time as the likelihood for these subjects to be censored increases. Furthermore, when patients fail because of competing events, the covariate values for such subjects continue to be used in subsequent risk calculations. Lau et al., (2009) provides an illustration comparing the different risk sets for both cause specific and the sub-distribution hazard. Both risk sets would be equal until observing the first competing event and would be smaller in all time points for the sub-distribution hazard thereafter.

The proposed model by Fine and Gray (1999) for the sub-distribution hazard is a Cox–type semi-parametric proportional hazards model:

$$h_j^{sd}(t / X) = h_{j,o}^{sd}(t)\exp\left(B_j^{sdT} X\right) \qquad (sd \text{ denotes sub-distribution}) \quad (5)$$

The first term on the right hand side of this function denotes a baseline sub-hazard function with all covariates set to zero. *X* is the covariate vector and the values of *B* are the covariate effects obtained in relation to competing risks *j*. This model uses the partial likelihood principle[15] such that one can assume a constant difference between the cumulative incidence functions independent of time *t*. The CIF allows for the

[15] The log(-log) transformation.

incorporation of the relevant information arising from other competing risk events when estimating the likelihood of the event of interest. Sub-distribution hazard rates are assumed to be proportional for all the included covariates. Based on this model, the CIF becomes:

$$CIF_j(t/X) = 1 - \exp\left[-\int_0^t h_j^{sd}(s/X)ds\right] \quad (6)$$

The CIF for competing risk *j* is therefore a function of the sub-hazard only for cause *j,* with the integral on the right hand side of the function being the cumulative sub-distribution hazard function (Cleves et al., 2010). The CIF is estimated directly from the regression coefficients obtained by the method proposed by Fine and Gray (1999) and the effects of covariates on the cumulative incidence function could thus be derived directly from the model for, $h_j^{sd}(s/X)$.

3.5 Results and Discussion

This section presents the results of the impact of a number of covariates on the event of interest, primarily different consultant and surgeon volume levels. In addition, variations in the relationship between consultant volume levels and the events of interest due to varying consultant job contract conditions are included. Results based on the CSHF and the CIF using both non-parametric and semi-parametric methods are provided. The conclusions derived from the estimated multinomial logistic regression model are presented in Appendix B of this chapter.

3.5.1 *Non-parametric estimation*

3.5.1.1 The Cause Specific Hazard Function

Figure 3-1 illustrates the CSHF of the event *failure*, representing the risk of failing from the PTCA procedure.[16] The results show that the hazard of failing from the intervention within the 60 day period increases as *t* approaches the 60 day point mark with as expected, higher risk in the first few days following the procedure. Indeed, Figure 3-2 illustrates the hazard estimate for the event *death*, and shows the relatively high risk of dying in the immediate days following the undertaking of the procedure . The hazard from this event falls as *t* increases. Figure 3-3 shows the hazard estimates for the event *readmission* which increases with *t* up to around the 50 day point.

Figure 3-1: Hazard for event *failure*

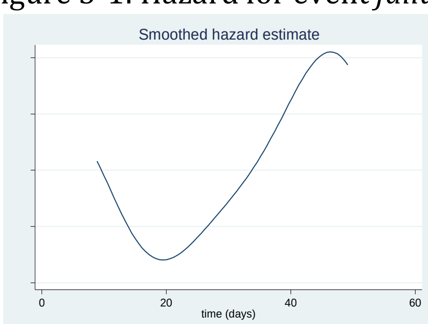


Figure 3-2: Hazard for event *death*

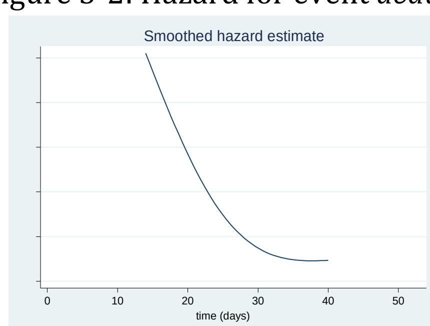
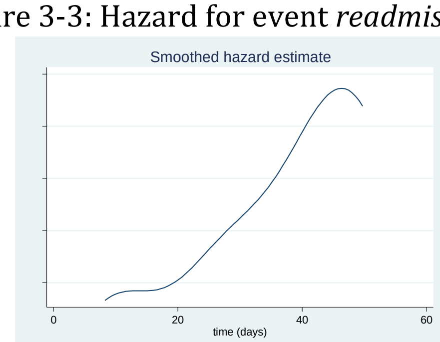


Figure 3-3: Hazard for event *readmission*



---

[16] The hazard estimate for the risk of failing from the undertaken procedure is composed from the addition of the hazard of events *death* and *readmission*.

In addition to estimating the CSHF for the different failure events, results are presented for differences in the hazard rates for the event *failure*[17] when different consultant and volume characteristics are considered. In particular, three groups of characteristics are identified: patients under the control of consultants with a *low/high* volume level of patient activity, patients being seen by surgeons with a *low/high* volume level of patient activity and patients being seen by consultants with different job plan contracts. It can be concluded from Figure 3-4 that the rate of failing from the treatment procedure is higher if the patient is seen by consultants on *high* volume levels (*cvol=1). The hazard of failure for patients being seen by surgeons at higher volume levels (Figure 3-5) compared to patients seen by low volume surgeons varies as t changes.* Figure 3-6 represents differences due to the job plan contract type of the consultant.[18]

Figure 3-4
By consultant volume levels (CSHF)

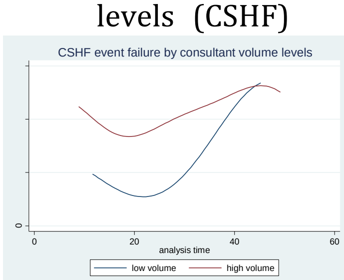


Figure 3-5
By surgeon volume levels (CSHF)

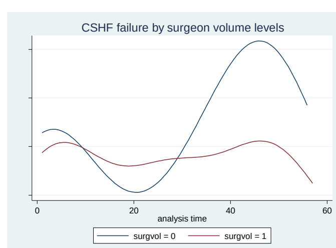
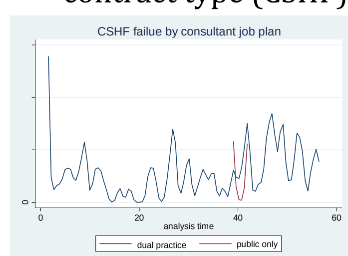


Figure 3-6
By consultant job plan contract type (CSHF)



[17] The other representations for the separate failure events *readmission* and *death* are included in Appendix A to this chapter in Figure 3-19 and Figure 3-20 respectively.

[18] This primarily represents the 'readmittance' activity and the fact that there are a relatively small number of consultants on a *public-only* contract type. There are no cases of patients who experience the event *death* and are under the care of *public-only* contract type consultants.

The results of the log-rank test for equality of the hazard associated with the event of interest for the different groups are presented in Table 3.5. The results show that surgeon volume has a marginally significant effect on the cause specific hazard for the event *failure*. However, no significant effect of differences in the consultant and surgeon characteristics were observed for the hazard functions of the other specific events.

Table 3.5: Log rank test for equality of the cause specific hazard

p-value	Event *failure*	Event *death*	Event *readmission*
Consultant volume	0.3231	0.9214	0.2273
Surgeon volume	0.0785*	0.4234	0.1163
Consultant contract type	0.1144	0.2085	0.2767

* Significant at the 10% level of significance.

3.5.1.2 The Cumulative Incidence Function

Figures 3-7, 3-8 and 3-9 show the CIF for the different events of interest. The cumulative incidence of failing from the PTCA procedure within the 60 day period increases as t gets closer to 60 days.[19] The treatment failure probability can be separated into the failure from getting the event *readmission* or *death*. The cumulative probability of failing from each of the events, *death* and *readmission,* within 60 days also increases as the 60th day following discharge day is approached.

[19] The results obtained from the CIF are compared to those obtained from the use of the KM method and it is found that as expected the (1-KM) method overestimates the incidence of the event *failure*.

Figure 3-7
CIF for event *failure*

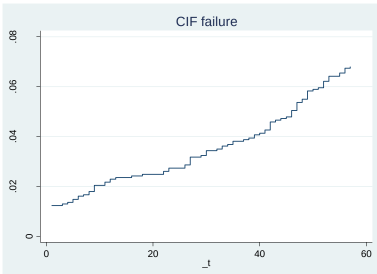


Figure 3-8
CIF for event *death*

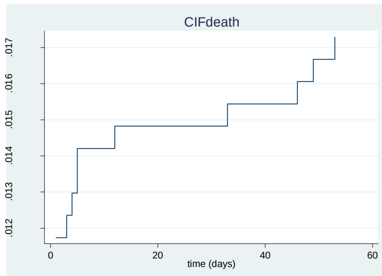
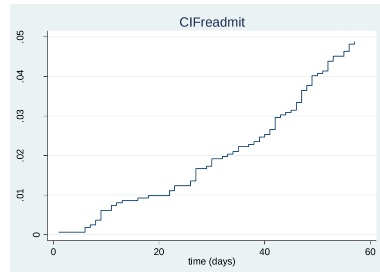


Figure 3-9
CIF for event *readmission*



The impact of differences in the consultant and surgeon volume levels together with the impact of differences in the consultant job plan contract type on the CIF for the failure event[20] is shown in Figures 3-10, 3-11 and 3-12. The incidence of failing following the PTCA procedure is higher if the patient is under the care of consultants with high volume levels. The incidence of failure for patients being seen by surgeons at higher volume levels (Figure 3-11) compared to patients seen by low volume surgeons varies as 't' changes. Figure 3-12 represents differences due to the job plan contract type of the consultant.[21]

[20] The results for the events *readmission* and *death* are included in Appendix A to this chapter in Figure 3-21 and Figure 3-22 respectively.
[21] This primarily represents the 're-admittance' activity and the fact that there are a relatively small number of consultants on a *public-only* contract type. There are no cases of patients who experience the event *death* and are under the care of *public-only* contract type consultants.

Figure 3-10
CIF by consultant volume

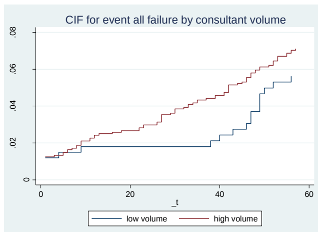


Figure 3-11
CIF by surgeon volume

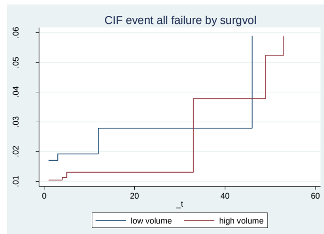
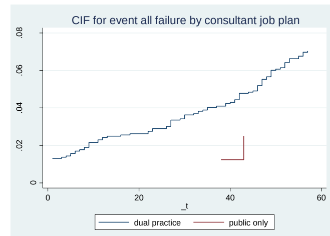


Figure 3-12
CIF by consultant job contract type



Using the *p* values obtained from the Pepe and Mori (1993) test, as presented in Table 3.6, it can be concluded that there is a statistical difference in the CIF curves based on the volume levels of the consultant for the event *readmission*.

Table 3.6: Pepe and Mori (1993) test comparing the cumulative incidence for the two groups: consultant and surgeon volume levels

p-value	Event *failure*	Event *death*	Event *readmission*
Consultant volume	0.1196	0.9606	0.0028**
Surgeon volume	0.4285	0.4166	0.8343
Consultant contract type	0.00002***	0.00001**	0.01988**

*** p<0.01, ** p<0.05, * p<0.1.

The differences in the CIF based on the volume levels of the surgeon were found to be statistically insignificant whilst the difference in the CIF curve based on the consultant contract type is accepted for all event types.

3.5.2 Semi-parametric estimation: the CSHF and the CIF

This section discusses the CSHF and the CIF taking into account differences in consultant and surgeon characteristics whilst also controlling for patient related factors of *age* and *LOS*.[22] Covariates are chosen on the basis of a number of regression estimations and are dependent on those factors which are expected to affect hazard and incidence rates. Results are also presented for the compounding interaction effects of the consultant job contract category on the relationship between the event of interest and the different measures of volume. The results from each of the semi-parametric estimation methods are compared for robustness purposes to the obtained estimates based on a multinomial logistic model.[23]

3.5.2.1 Modelling the CSHF

This section provides the results for the CSHF estimated using the Cox multivariable regression model for the *failure* event. The events *death* and *readmission* constitute a failure and a separate CSHF for each of these possible failure events is estimated. The results are presented using different sets of volume variables to capture the robustness of the volume measures used. The effect of the consultant job plan contract typ*e* on the cause specific hazard is included together with an interaction term to study whether differences in the consultant contract type modify the relationship between the volume

[22] The total number of diagnosis for the particular patient, *diagtotal*, to represent the severity of illness for the patient, was also tested in the various estimations undertaken.
[23] A full description of the results obtained using a multinomial logistic model is provided in Appendix B to this chapter. The results obtained using this method serve to overall confirm the results obtained through the CSHF and the CIF.

variable and the event of interest. Tables 3.7, 3.8 and 3.9 present the results for the different CSHF's obtained using the different volume variables under consideration.

The results in Table 3.7 show that there is a lower (and significant) risk of failing from the procedure for patients seen by consultants on *public-only* job contracts compared to patients under the control of *dual practice* contract consultants (the hazard rate of failing falls by 72%). The hazard rate for the volume variables *countc* and *counts* are both insignificant and close to 1 indicating that the rate of failure is expected to stay fairly flat as volume levels increase. As expected the covariates for *age* and *LOS* positively affect the hazard rate for the event *death*. The proportional hazard assumption[24] is clearly accepted for the estimation is columns 3-6 whilst only accepted marginally in the first two columns of the same table.

Table 3.7: Cause Specific Hazard at particular event of interest (countc and counts)

	(1)	(2)	(3)	(4)	(5)	(6)
Event	Failure	Failure	Re-admit	Re-admit	Death	Death
Variable						
public-only	0.281*	0.326	0.390	0.443	-	-
age	1.003	1.004	0.984	0.984	1.068***	1.070***
LOS	1.041***	1.043***	1.025	1.027	1.060***	1.063***
countc	0.999		0.999		0.998	
counts		0.999		0.999		0.999
Observations	1,619	1,619	1,619	1,619	1,619	1,619

Hazard ratios are reported, *** p<0.01, ** p<0.05, * p<0.1.
The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.

[24] Obtained using the *estat phtest* command in Stata. Data provided in Appendix C to this chapter.

Table 3.8: Cause Specific Hazard at particular event of interest (vol90c and vol90s)

Variable / Event	(1) Failure	(2) Failure	(3) Re-admit	(4) Re-admit	(5) Death	(6) Death
public-only	0.355	0.325	0.466	0.439	-	-
age	1.003	1.003	0.983	0.983	1.066***	1.066***
LOS	1.046***	1.044***	1.030*	1.028*	1.069***	1.065***
vol90c	1.004		1.002		1.007	
vol90s		1.000		0.999		1.002
Observations	1,619	1,619	1,619	1,619	1,619	1,619

Hazard ratios are reported, *** p<0.01, ** p<0.05, * p<0.1.
The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.

Table 3.9: Cause Specific Hazard at particular event of interest (cvol and svol)

Variable / Event	(1) Failure	(2) Failure	(3) Re-admit	(4) Re-admit	(5) Death	(6) Death
public-only	0.437	0.294*	0.599	0.402	-	-
age	1.004	1.005	0.984	0.985	1.068***	1.069***
LOS	1.050***	1.042***	1.034*	1.026	1.070***	1.062***
cvol	1.433		1.447		1.388	
svol		0.703*		0.721		0.678
Observations	1,619	1,619	1,619	1,619	1,619	1,619

Hazard ratios are reported, *** p<0.01, ** p<0.05, * p<0.1.
The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.
The baseline, cvol=0 indicates consultant *low* volume, svol=0 indicates surgeon *low* volume.

Table 3.8 presents the results when the volume variables representing the most recent activity of surgeons and consultants prior to the hospital admittance date of the patient, *vol90c* and *vol90s* are considered.[25] There is no statistically significant relationship between the event of *failure* and the most recent activity volume levels of both

[25] The proportional bazard assumption is rejected in columns 1,2 whilst the assumption holds for column 3-6, (obtained using *estat phtest* command in Stata). The relevant data is provided in Appendix C to this chapter.

consultants and surgeons. As expected, the hazard rate for the event of *death* is significant for both changes in *age* and *LOS*.

This results in Table 3.9 show that the hazard rate of failing is lower if a patient is under the care of a consultant on a *public-only contract*. There is a lower hazard rate of failing from the procedure if the patient is seen by a surgeon on a high volume level, possibly suggesting that the practice makes perfect hypothesis holds in this case. On the other hand, there is a higher hazard rate of failing from the procedure (however insignificant) if the patient is under the care of a consultant with high volume levels. The proportional hazard assumption[26] is accepted for the estimation is columns 3-6 whilst rejected in the first two columns of the same table.

We included the interaction term in order to determine whether the consultant job plan can modify the effect of volume on the hazard for each of the events of interest. The results are presented in Table 3.10. We found that if volume (represented by *countc*) is increased by one unit and all other variables are held constant, the hazard of failing remains practically the same for those patients under the care of consultants working on *dual practice* contracts. If volume is increased by one unit and the patient is under the responsibility of a consultant on a *public-only* contract, then the hazard rate of failing is equal[27] to 1.107. This rate varies for different levels of volume due to the interaction effect. The rate of failing from the procedure undertaken for a change in volume is slightly higher for those patients seen by consultants on *public-only* contracts

[26] Obtained using the *estat phtest* command in Stata. Data provided in Appendix C to this chapter.
[27] Obtained as the exp((-0.001+0.103)*counc). Values represent the coefficients corresponding to the hazard rates provided in Table 3.10.

compared to those patients under consultants with *dual practice* contracts. If a patient is moved from a consultant with a *dual practice* contract to a consultant with a *public-only* contract, at a volume level set to zero and holding all other variables constant, the hazard of failing decreases (by around 93%). However, the effect of the consultant contract type on the hazard rate of failing changes with changes in volume levels due to the interaction effect.[28]

Table 3.10: Cause Specific Hazard at particular event of interest - with interaction terms

Event / Variable	(1) Failure	(2) Failure	(3) Re-admit	(4) Re-admit	(5) Death	(6) Death
age	1.003	1.003	0.983	0.983	1.068***	1.066***
LOS	1.041***	1.047***	1.026	1.030*	1.060***	1.069***
public-only	0.067	0.081*	0.081	0.089	-	-
countc	0.999		0.999		0.998	
countc x *public-only*	1.108		1.121		0.864	
vol90c		1.003		1.002		1.007
vol90c x *public-only*		1.214*		1.243**		1.000
Observations	1,619	1,619	1,619	1,619	1,619	1,619

Hazard ratios are reported, *** p<0.01, ** p<0.05, * p<0.1.
The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.

Similar results (column 2) are obtained when using the volume indicator, *vol90c*, which represents the number of cases seen by the consultant in the last 90 days prior to the patient's admission to hospital. An insignificant hazard rate of failing is obtained with respect to changes in volume for patients who are under the care of consultants on *dual practice* contracts. If volume is increased by one unit and the patient is under the responsibility of a consultant on a *public-only* contract, then the hazard rate of failing is

[28] exp((-2.696+(0.103*countc)). Values represent the coefficients corresponding to the hazard rates provided in Table 3.10.

equal[29] to 1.217. This rate varies for different levels of volume due to the interaction effect. A significant hazard rate of failing for patients under the care of consultants on *public-only* contracts is recorded when the volume variable, *vol90c*, is used. If a patient is moved from a consultant with a *dual practice* contract to a consultant with a *public-only* contract, at a volume level set to zero and holding all other variables constant, the hazard of failing decreases (by around 92%). The results also show that the likelihood of the event *death* is significantly affected by the *LOS* and the *age* covariates. The proportional hazards assumption[30] is rejected (at the 90% level of significance) for the representations in column 1 and column 2 whilst accepted for all other columns in Table 3.10. The robustness of these results has been assessed in relation to estimates obtained from a multinomial logistic model[31]. The results obtained from the CSHF are broadly in line with those derived from the multinomial logistic model.

The marginal effect of changes in the volume indicator, *countc* is not constant and varies for different values of volume. The predictions[32] of the hazard for the event of interest at different volume levels of *countc,* at each of the different consultant job plan contracts, are presented in Figure 3-13 and Figure 3-14.

[29] Obtained as the exp((0.003+0.194)*counc). Values represent the coefficients corresponding to the hazard rates provided in Table 3.10.
[30] Obtained using *estat phtest* command in Stata. Data provided in Appendix C to this chapter.
[31] A detailed description of the results is presented in Appendix B to this chapter.
[32] Based on equation (1) in Table 3.10.

Figure 3-13: The hazard function of failing from the procedure for different volume levels for consultants on *public-only* contracts[33]

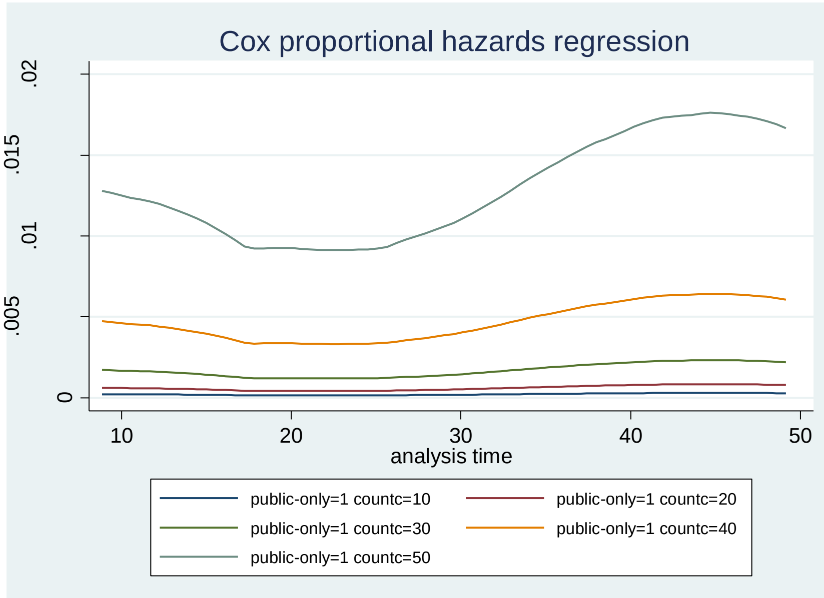
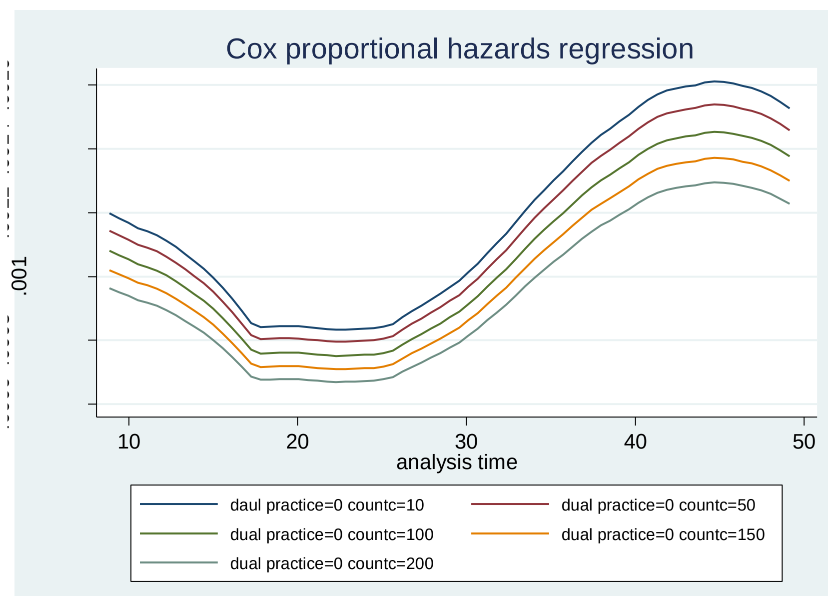


Figure 3-14: The hazard function of failing from the procedure for different volume levels for consultants on *dual practice* contracts[34]



[33] The different volume levels applied reflect the current volume levels of the consultants operating under the *public-only* contract type.
[34] The different volume levels applied reflect the current volume levels of the consultants operating under the *dual practice* contract type.

The results in Figure 3-13 show that as volume levels for consultants on *public-only* contracts increase, the hazard of failing from the procedure increases. This reflects the positive interaction term for patients under the care of consultants on *public-only* contracts. The hazard of failing for patients under the care of consultants on *dual practice* contracts (Figure 3-14) is only affected marginally when volume levels for these consultants on such contract type increase. This suggests that changes in failure outcomes are more sensitive to changes in volume levels of consultants on *public-only* contracts.

3.5.2.2 Modelling the CIF

Tables 3.11, 3.12 and 3.13 present the results of a series of CIFs based on the approach proposed by Fine and Gray (1999) to model each of the events using various volume measures: *countc* and *counts,* vol90c and *vol90s,* and *cvol and svol.* Table 3.14 shows how the consultant job plan contract type modifies the effect of the volume covariates on the outcome events. Variables for patient *age* and *LOS* at the hospital are considered as the main covariates in the different specifications.[35]

As shown in Table 3.11, the sub-hazard for the volume variables *countc* and *counts* are both insignificant and close to 1 indicating that as the volume of the consultant and surgeons increase by one unit the incidence rate for failure is expected to stay fairly flat, when keeping all other variables constant. Furthermore, assuming the same level of

[35] The total number of diagnosis for the particular patient, *diagtotal*, which represents the severity of the patient case, was also tested in the various estimations undertaken.

volume for the consultant in column 1, the sub-hazard rate of failing from the procedure is lower for patients being seen by consultants on *public-only* contracts compared to *dual practice* contracts. The sub-hazard of failing from the procedure is also lower (however insignificant) for patients being seen by consultants on *public-only* contracts compared to dual practice contracts when assuming the same level of volume for the surgeon in column 2. The study found that the incidence of the event *failure* is affected by the LOS covariate whilst the specific event of failure *death* is affected by the *age* and the *LOS* covariates.

Table 3.12 presents the results obtained when the number of cases seen by the consultant *vol90c* and the number of cases seen by the surgeon, *vol90s* are taken as the volume measures. Insignificant sub-hazard values were obtained for the volume measures so that the most recent activity volume levels of the consultant and the surgeon did not have an impact on the incidence rate of failing following a PTCA procedure. Keeping all variables constant, the sub-hazard rate of failing from the procedure is lower (however insignificant) for patients being seen by consultants on *public-only* contracts compared to *dual practice* contracts. In line with the previous table, the results in table 3.12 show that the incidence of the event *failure* is affected by the *LOS* covariate whilst the specific event of failure *death* is affected by the *age* and the *LOS* covariates.

Table 3.11: Cumulative Incidence Function at particular events of interest (countc and counts)

Event Variable	(1) Failure	(2) Failure	(3) Re-admit	(4) Re-admit	(5) Death	(6) Death
age	1.004	1.004	0.982*	0.983	1.068***	1.0699***
LOS	1.041***	1.043***	1.021*	1.023*	1.060***	1.0626***
public-only	0.281*	0.326	0.370	0.436	-	-
countc	0.999		0.999		0.998	
counts		0.999		0.999		0.999
Observations	1,619	1,619	1,619	1,619	1,619	1,619

Coefficients reflect Sub-Hazard Ratios,*** p<0.01, ** p<0.05, * p<0.1.
The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.

Table 3.12: Cumulative Incidence Function at particular events of interest (vol90c and vol90s)

Event Variable	(1) Failure	(2) Failure	(3) Re-admit	(4) Re-admit	(5) Death	(6) Death
age	1.003	1.003	0.982*	0.982*	1.067***	1.067***
LOS	1.047***	1.044***	1.026**	1.024**	1.069***	1.065***
public-only	0.355	0.325	0.459	0.432	-	-
vol90c	1.004		1.003		1.008	
vol90s		0.999		0.999		1.001
Observations	1,619	1,619	1,619	1,619	1,619	1,619

Coefficients reflect Sub-Hazard Ratios,*** p<0.01, ** p<0.05, * p<0.1.
The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.

Table 3.13: Cumulative Incidence Function at particular events of interest (cvol and svol)

Event Variable	(1) Failure	(2) Failure	(3) Re-admit	(4) Re-admit	(5) Death	(6) Death
age	1.004	1.005	0.983	0.984	1.068***	1.069***
LOS	1.050***	1.042***	1.030**	1.022*	1.070***	1.062***
public-only	0.437	0.294*	0.587	0.395	-	-
cvol	1.433		1.438		1.390	
svol		0.703*		0.725		0.676
Observations	1,619	1,619	1,619	1,619	1,619	1,619

Coefficients reflect Sub-Hazard Ratios,*** p<0.01, ** p<0.05, * p<0.1.
The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.
The baseline, cvol=0 indicates consultant *low* volume, svol=0 indicates surgeon *low* volume.

The volume variable is in Table 3.13 measured using a constructed dummy variable which distinguishes between high and low volume consultant and surgeons operating at the public hospital. There is a lower hazard rate of failing from the procedure if the patient is seen by a surgeon on the high volume level, *svol*, possibly suggesting that the practice makes perfect hypothesis holds in this case. On the other hand, there is a higher hazard rate of failing from the procedure (however insignificant) if the patient is under the care of a consultant with high volume levels, *cvol*. The results in Table 3.13 (column 2) show that keeping all variables constant, the sub-hazard rate of failing is lower and significant if a patient is under the care of a consultant on a *public-only* contract. As expected, the cumulative incidence rate of the event *death* was found to be affected by both *age* and *LOS*.

Table 3.14: Cumulative Incidence Function at particular events of interest-with interaction terms

Event Variable	(1) Failure	(2) Failure	(3) Re-admit	(4) Re-admit	(5) Death	(6) Death
age	1.003	1.003	0.982*	0.982*	1.068***	1.067***
LOS	1.042***	1.047***	1.022*	1.027**	1.0604***	1.069***
public-only	0.067**	0.081***	0.136**	0.146**	-	-
countc	0.999		0.99		0.998	
countc x *public-only*	1.108**		1.072*		0.993	
vol90c		1.003		1.002		1.008
vol90c x *public-only*		1.214***		1.150***		1.000
Observations	1,619	1,619	1,619	1,619	1,619	1,619

Coefficients reflect Sub-Hazard Ratios,*** p<0.01, ** p<0.05, * p<0.1.
The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.

Table 3.14 introduces into the study the effect that the consultant job plan has on the relationship between volume and the incidence of the event of interest. The specification in column 1 shows that if volume is increased by one unit and the patient is under the control of a consultant on a *dual practice* contract, with all other variables kept constant, the incidence rate of failing from the PTCA procedure is practically unchanged. If volume is increased by one unit and the patient is under the responsibility of a consultant on a *public-only* contract, then the incidence rate of failing is equal[36] to 1.107. The magnitude of this change varies for different levels of volume due to the interaction effect. The incidence rate of failing increases significantly for patients being seen by consultants on *public-only* contracts for a unit change in volume compared to patients being under the control of consultants on *dual practice* contracts. The sub-hazard for failing from the procedure for patients under the care of consultants with exclusive public sector contracts, at zero volume levels and keeping all other variables constant, is lower for patients under the care of consultants on *public-only* contracts compared to *dual practice* contracts. The magnitude of this change varies and depends on the particular volume level due to the interaction effect.[37]

Results based on *vol90c* (column 2) indicate that the effect of a one unit change in volume on the incidence rate of failing for patients under the care of consultants on *public-only* contracts is equal[38] to 1.218. A small positive sub-hazard of failing for patients is obtained with respect to volume changes when consultants overseeing the patients are on *dual practice* contracts. The sub-hazard of failing, at zero volume levels

---

[36] Obtained as the exp((-0.0011+ 0.1029)*countc). Values represent the coefficients corresponding to the sub-hazard ratios provided in Table 3.14.

[37] exp((-2.6955+(0.1029*countc)). Values represent the coefficients corresponding to the sub-hazard ratios provided in Table 3.14.

[38] Obtained using exp((-0.0008333+0.0779643)*vol90c). Values represent the coefficients corresponding to the sub-hazard ratios provided in Table 3.14.

and keeping all other variables constant, for patients under the care of consultants with exclusive public sector contracts is lower when compared to those patients under the care of consultants on *dual practice* contracts. The sub-hazard of failing for patients under the care of consultants with *public-only* contracts is around 8% of that for patients under the care of consultants who can carry out private practice.

The results show that there is a significant sub-hazard of failing from the event *re-admission* for patients with *public-only* contracts. Patients under the care of consultants who have a *public-only* contract have a lower sub-hazard for the event re-admission compared to those patients who are under the care of *dual practice* contract consultants. Furthermore, we find positive and significant values for the sub-hazard of the interaction term in columns 3 and 4. If volume is increased by one unit and the patient is under the care of a consultant on a *public-only* contract then the incidence rate of being re-admitted is higher compared to the same volume increase of one unit for consultants with *dual practice* contracts. The robustness of these results has been assessed in relation to estimates obtained from a multinomial logistic model.[39] The results obtained from the CIF are broadly in line with those derived from the multinomial logistic model.

The Fine and Gray (1999) model assumes that the effects of the covariates are proportional to the sub-hazard of the particular event of interest. The assumption of proportional hazards was tested for a possible changing effect of the covariates over

[39] A detailed description of the results is presented in Appendix B to this chapter.

the time period. The plot of the Schoenfeld residuals[40] (based on the results presented in column 1 in Table 3.14) shown in Figure 3-15, was used to test the proportional hazards assumption underpinning the CIF. Results show that there is an indication of no violation of the proportionality assumption. The assumption of proportionality was also tested by introducing time varying coefficients for all of the covariates in the model. Figure 3-16 illustrates the Schoenfeld residuals after the time varying coefficients are taken into account. Some of the tested time varying coefficients turned out to be statistically significant thus suggesting the rejection of the proportionality assumption.

Figure 3-15: Schoenfeld residuals

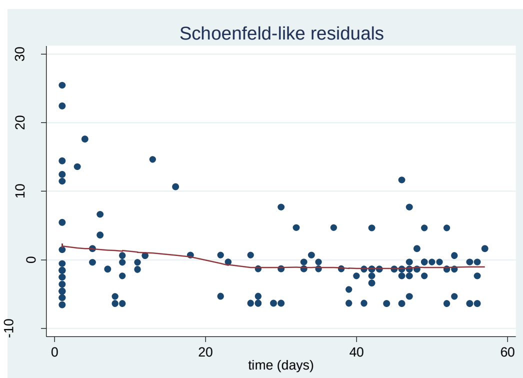
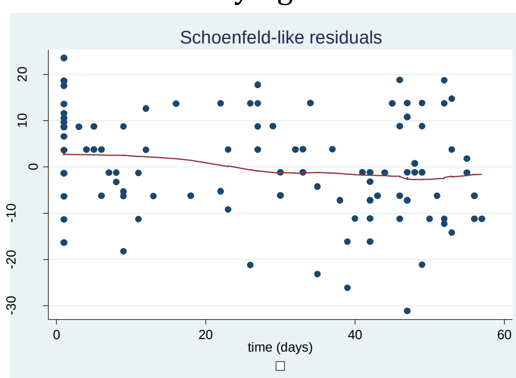


Figure 3-16: Schoenfeld residuals with time varying coefficients



The marginal effect of changes in the volume indicator, *countc* is not constant and varies for the different values of volume. Figure 3-17 and 3-18 show predictions[41] of the incidence rate for the event of interest *failure* at different volume levels of *countc, for* each of the consultant job plan contract types. We found evidence of differences in the

[40] Based on Grambsch and Therneau (1994).
[41] Based on equation (1) in Table 3.14.

CIF for the event of interest, at different levels of volume for particular job plan contract types.

Figure 3-17: Volume variation for consultants on *public-only* contracts[42]

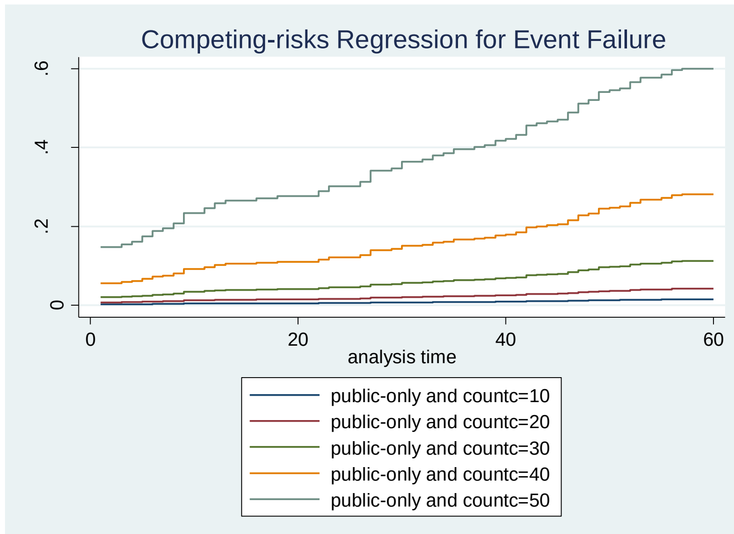
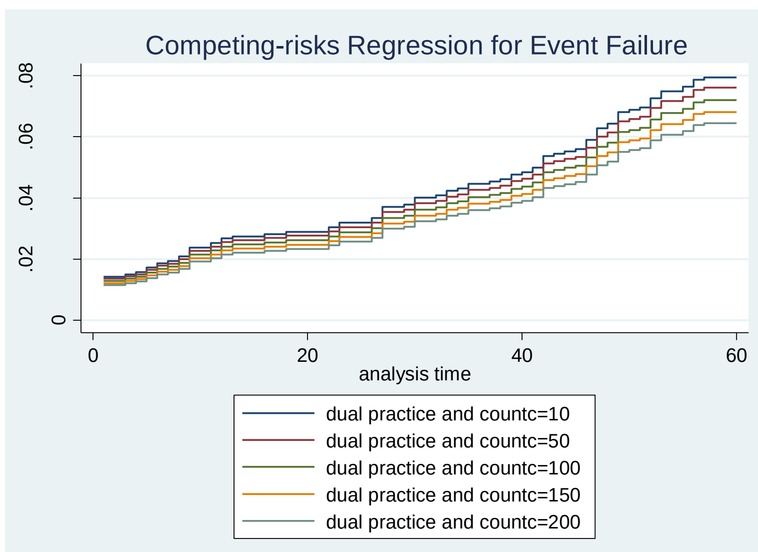


Figure 3-18: Volume variation for consultants on *dual practice* contracts[43]



[42] The different volume levels applied reflect the current volume levels of the consultants operating under the *public-only* contract type.
[43] The different volume levels applied reflect the current volume levels of the consultants operating under the *dual practice* contract type.

Figure 3-17 shows that the incidence of failure increases if the volume of consultants on *public-only* contracts increases. This is in line with the results obtained on the interaction term for the consultant job plan contract type which shows that the impact of a change in volume on the incidence rate of failing for patients under the care of consultants with exclusive public sector contracts, is higher. Figure 3-18 provides evidence that changes in the volume levels of consultants who undertake private practice has little impact on the incidence rate of failing from the procedure undertaken. The incidence of failing for patients under the care of consultants on *dual practice* contracts only changes marginally when volume levels of these consultants change.

3.6 Conclusion

This study focused on understanding the volume-outcome relationship in the context of competing events following a PTCA procedure. The impact of patient focused covariates such as *age* and *LOS* is included in the analysis. Furthermore, we studied whether the consultant job plan contract type can modify the effect of changes in volume on the hazard rate of each of the events of interest.

Both non-parametric and semi-parametric survival analysis methods were used in the study together with the use of a multinomial logistic model to test for the robustness of the obtained results. Results based on the non-parametric CSHF and the CIF show that there is a slightly higher rate of failure following the PTCA procedure for patients who are under the care of consultants with high volume levels. The results obtained from the non-parametric methods, in relation to the variation in failure rates due to difference in job plan contract type, are to be treated with caution and one acknowledges the contrainst due to the significantly limited dataset available.

The results obtained from the semi-parametric models show that there is a lower risk of failure from the PTCA procedure for patients being under the care of consultants with *public-only* job contracts compared to patients under *dual practice* consultants. Some evidence of higher risk of failure for patients under the care of consultants with high volume levels is also noted. Furthermore, a change in volume was found to have an even higher impact on the sub-hazard rate for failing if patients were under the responsibility of consultants practicing exclusively in the public sector. This could indeed reflect the capacity constraints within which consultants working at the public hospital operate, whereby added activity for such consultants could lead to an impact on overall outcomes. On the other hand, one finds some evidence of a lower risk of failure for patients seen by surgeons with high volume levels, possibly implying some evidence of the presence of the practice makes perfect hypothesis. Also, and not surprisingly, our findings show that the likelihood of death of a patient undergoing a PTCA procedure is affected significantly by their age and LOS at hospital.

The results based on the multinomial logistic estimations confirm the results obtained from the CSHF and the CIF functions. One finds that there is a lower risk for a patient to register a failure event after the PTCA procedure if the patient is under the care of a consultant on a *public-only* contract with the public hospital. Patients being seen by consultants on high volume levels are more likely to have an event *failure* during the 60 day period. Moreover, a change in volume was found to have an even higher impact on the incidence for failing if patients were under the responsibility of consultants practicing exclusively in the public sector compared to *dual practice* consultants.

The survival models used assume homogeneity where all individuals are subject to the same risks embodied in the hazard function. When we introduce covariates we relax this assumption by introducing observed sources of heterogeneity. There could though also be unobserved sources of heterogeneity that are not being captured by the covariates. A problem of concern is the omission of variables (counfounders) that affect the event of interest variable and that such variables are likely to be correlated with the included control variables. Indeed, there could be an intrusive effect of omitted variables and as stated by Heckman and Singer (1981) the estimates of the parameters of duration models, are influenced by the distribution of unobservables as this could lead to estimates which poorly describe the true behavioural models generating duration data.

Furthermore, in this study we have controlled for a range of patient and consultant characteristics which could possibly influence the volume-outcome relationship, particularly, patient age, severity of illness and consultant job contract type. One recognises that there are though a number of other factors which consultants put into the balance when making a choice between a dual practice contract and a public-only contract with the public hospital. The observed impact on outcomes may thus be due to factors other than the direct variation arising from the particular job contract type. The data for consultant job contract type may be thus capturing factors which affect the actual decision of the consultant to choose the particular type of job plan contract.

The identification of the volume-outcome relationship could in the context of this study be partly controlled by the fact that there is only one single public hospital on the island which deals with the health care needs of the Maltese population. Consultants cannot move to other public hospitals and all consultants considered in this study work within the same public sector structure. This reduces the possible biases which can arise when patients could be referred to hospitals with different characteristics thus possibly affecting outcomes. Furthermore, the size of the sample used within this study is constrained by the fact that the sample includes all the existing PTCA activity carried out by consultants within the hospital over the three year period, thus serving to reduce potential biases in sample selection.

It is indeed possible that consultants with different skills and work/leisure preferences have a tendency to opt for a particular type of contract type and thus outcomes are also influenced by such factors rather than just the particular job contract type. Consultants might base their choice of job contract type also on what they perceive to be the mostly sought out form of contract type requested and desired by their patients. Patient preferences will thus have an impact on consultant job contract type. Furthermore, one could consider data in relation to years of experience at the job (to measure skills) and an indicator of work/leisure preferences to help disentangle the confounding impact of such factors on outcomes. Determining the possible reasons underlying the decision by consultants to work under the conditions of a particular type of job contract could add to a better explanation of the drivers of the relationship between outcomes and the consultant job contract type. This is important especially for policy related reasons. Further work in this area would be a possible extension to this research work.

These conclusions contribute towards a better understanding of the existence, nature and possible channels for explaining the relationship between volume and competing risk outcomes within a unique public hospital setting like that of Malta. Furthermore, this study helps provide evidence on some of the factors which are likely to have an impact on treatment effectiveness. In particular, further investigation into the behavioural patterns of consultants and surgeons within such a hospital setting would serve to obtain a better understanding of the contribution of health professionals towards outcome effectiveness. A possible further extension for this study could be to incorporate additional information on the characteristics of the surgeons working within the hospital and thus to analyse their particular impact of patient outcomes. One would expect that even within a single hospital setting, differences in the behaviour of consultants and surgeons at an individual level exist and such differences could have an impact on outcome. The broadening of this study to include other procedures within the hospital, apart from PTCA, would help set up effective policy initiatives within the health care sector.

References

Al, C., Norton, E., (2003). 'Interaction terms in logit and probit models', *Economics Letters*, vol.80(1): 123-129.

Andersen, P.K., Geskus, R.B., de Witte, T., and Putter, H. (2012). 'Competing risks in epidemiology: possibilities and pitfalls', *International Journal of Epidemology,* vol.41(3), pp. 861-870.

Arriagada, R., Rutqvist L.E, Kramar, A., and Johansson, H. (1992). 'Competing risks determining event-free survival in early breast cancer', *British Journal of Cancer,* vol.66, pp. 951-957.

Bakoyannis, G. and Touloumi, G. (2011). 'Practical methods for competing risks data: a review', *Statistical Methods in Medical Research,* vol.21(3), pp. 257-272.

Bakoyannis, G.T.G. (2012). 'A practical guide on modeling Competing risk data', CASCADE colloboration, www.cascade-colloboration.org (Accessed 21st September 2013).

Beyersmann, J., Dettenkofer, M., Bertz, H., and Schumacher, M. (2007). 'A competing risks analysis of bloodstream infection after stem-cell transplantation using subdistribution hazards and cause-specific hazards', *Statistics in Medicine,* vol.26(30), pp. 5360-5369.

Billingsley, K.G., Morris, A.M., Dominitz, J.A., Matthews, B., Dobie, S., Barlow, W., Baldwin, L.-M., and ldwin, L.M. (2007). 'Surgeon and hospital characteristics as predictors of major adverse outcomes following colon cancer surgery: understanding the volume-outcome relationship', *Archives of Surgery,* vol.142(1), pp. 23-31.

Birkmeyer, J.D., Stukel, T.A., Siewers, A.E., Goodney, P.P., Wennberg, D.E., and Lucas, F.L. (2003). 'Surgeon volume and operative mortality in the United States', *New England Journal of Medicine,* vol.349, pp. 2117-2127.

Cleves, M., Gutierrez, R.G., Gould, W., and Marchenko, Y.V. (2010). 'An introduction to survival analysis', Third Edition. Stata Press.

Cornfield, J. (1957). 'The estimation of the probability of developing a disease in the presence of competing risks', *Am Journal Public Health Nations Health*, vol47(5), pp. 601-607.

Dudley, R. (2000). 'Selective referral to high-volume hospitals: Estimating potentially avoidable deaths', *JAMA,* vol.283(9), pp. 1159-1166.

Farley, D.E. and Ozminkowski, R.J. (1992). 'Volume-Outcome Relationships and Inhospital Mortality: The Effect of Changes in Volume over Time', *Medical Care,* vol.30(1), pp. 77-94.

Fine, J. and Gray, R.J. (1999). 'A Proportional Hazards Model for the Subdistribution of a Competing Risk', *Journal of the American Statistical Association,* vol.94(446), pp. 496-509.

Fix, E. and Neyman.J. (1951). 'A simple stochastic model of recovery, relapse, death and loss of patients', *Human Biology,* vol.23(3), pp. 205-241.

Furstova, J. and Valenta, Z. (2011). 'Statistical Analysis of Competing Risks: overall survival in a group of chronic myeloid leukemia patients', *European Journal of Biomedical Informatics,* vol.7(1), pp. 2-10.

Gandjour, A., Bannenberg, A., and Lauterbach, K.W. (2003). 'Threshold volumes associated with higher survival in health care: a systematic review', *Medical Care,* vol.41(10), pp. 1129-1141.

Gaynor, J.J., Feuer, E.J., Tan, C.C., Wu, D.H., Little, C.R., Straus, D.J., Clarkson, B.D., and Brennan, M.F. (1993). 'On the Use of Cause-Specific Failure and Conditional Failure Probabilities: Examples From Clinical Oncology Data', *Journal of the American Statistical Association,* vol.88(422), pp. 400-409. American Statistical Association.

Glasgow, R.E. and Mulvihill, S.J. (1996). 'Hospital volume influences outcome in patients undergoing pancreatic resection for cancer', *West J Med,* vol.165(5), pp. 294-300.

Gooley, T.A., Leisenring, W., Crowley, J., and Storer, B.E. (1999). 'Estimation of failure probabilities in the presence of competing risks: new representations of old estimators.', *Statistics in Medicine,* vol.18, pp. 695-706.

Gooley, T.A., Leisenring, W., Crowley, J., and Storer, B.E. (2001). 'Why Kaplan-Meier fails and cumulative incidence succeeds when estimating failure probabilities in the presence of competing risks.', In: Handbook of statistics in clinical oncology, pp. 513-523, New York.

Gordon, T.A., Burleyson, G.P., Tielsch, J.M., and Cameron, J.L. (1995). 'The effects of regionalization on cost and outcome for one general high-risk surgical procedure', *Ann.Surg,* vol.221(1), pp. 43-49.

Gowrisankaran, G., Ho, V., and Town, R.J. (2006). 'Causality and the volume outcome relationship in Surgery', *Stanford university*, unpublished work.

Grambsch, P.M. and Therneau, T.M. (1994). 'Proportional hazards tests and diagnostics based on weighted residuals', *Biometrika,* vol.81(3), pp. 515-526.

Halm, E.F., C Lee, C., and Chassin, M.R. (2000). 'Is volume related to outcome in health care? A systematic review and methodologic critique of the literature', *Annals of Internal Medicine,* vol.137(6), pp. 511-520.

Hamad, N., Pichard, A.D., Lyle, H.R., and Lindsay, J.Jr. (1988). 'Results of percutaneous transluminal coronary angioplasty by multiple, relatively low frequency operators: 1986-1987 experience', *Journal of American Cardiology,* vol.61(15), pp. 1229-1231.

Hamilton, B.H. and Hamilton, V.H. (1997). 'Estimating surgical volume outcome relationships applying survival models: accounting for frailty and hospital fixed effects', *Health Economics,* vol.6(4), pp. 383-395.

Hamilton, B.H. and Ho, V. (1998). 'Does Practice Make Perfect?: Examining the Relationship between Hospital Surgical Volume and Outcomes for Hip Fracture Patients in Quebec', *Medical Care,* vol.36(6), pp. 892-903.

Hannan, E.L., Racz, M., Ryan, T.J., McCallister, B.D., Johnson, L.W., Arani, D.T., Guerci, A.D., Sosa, J., Topol, E.J. (1997). 'Coronary Angioplasty Volume-Outcome Relationships for Hospitals and Cardiologists'. *Journal of the American Medical Association*, vol.277(11), pp 892-898.

Hannan, E.L., Wu, C., Walford, G., King, S.B., III, Holmes, D.R., Ambrose, J.A., Sharma, S., Katz, S., Clark, L.T., and Jones, R.H. (2005). 'Volume-outcome relationships for percutaneous coronary interventions in the stent era', *Circulation,* vol.112(8), pp. 1171-1179.

Heckman, J, Singer,B. (1982). 'The identification problem in econometric models for duration data', In Werner Hildenbrand(ed) *Advances in Econometrics*, Cambridge University Press.

Ho,V. (2000). 'Evolution of the volume-outcome relation for hospitals performing coronary angioplasty', *Circulation,* vol. 101, pp. 1086-1811.

Ho, V. (2002). 'Learning and the evolution of medical technologies: the diffusion of coronary angioplasty', *Journal of Health Economics,* vol.21(5), pp. 873-885.

Huckman, R.S. and Pisano, G.P. (2006). 'The Firm Specificity of Individual Performance: Evidence from Cardiac Surgery', *Management Science,* vol.52(4), pp. 473-488.

Huesch, M.D. and Sakakibara, M. (2009). 'Forgetting the learning curve for a moment: how much performance is unrelated to own experience?', *Health Economics,* vol.18(7), pp. 855-62.

Jollis, J.G., Peterson, E.D., DeLong, E.R., Mark, D.B., Collins, S.R., Muhlbaier, L.H., and Pryor, D.B. (1994). 'The relation between the volume of coronary angioplasty procedures at hospitals treating Medicare beneficiaries and short-term mortality', *The New England journal of medicine,* vol.331(24), pp. 1625-1629.

Jollis, J.G., Peterson, E.D., Nelson, C.L., Stafford, J.A., DeLong, E.R., Muhlbaier, L.H., and Mark, D.B. (1997). 'Relationship Between Physician and Hospital Coronary Angioplasty Volume and Outcome in Elderly Patients', *Circulation,* vol.95(11), pp. 2485-2491.

Joynt, K.E., Orav, E.J., and Jha, A.K. (2011). 'The association between hospital volume and processes, outcomes, and costs of care for congestive heart failure', *Annals of Internal Medicine,* vol.154(2), pp. 94-102.

Kawabuchi, K. and Sugihara, S. (2006). 'The volume outcome relationship in Japan: the case of Percutaneous Transluminal Coronary Angioplasty (PTCA) volume on Mortality of Acute Myocardial Infarction (AMI) Patients', *NBER*.

Kelly, J.V. and Hellinger, F.J. (1986). 'Physician and Hospital Factors Associated with Mortality of Surgical Patients', *Medical Care,* vol.24(9), pp. 785-800.

Kim, H.T. (2007). 'Cumulative Incidence in Competing Risks Data and Competing Risks Regression Analysis', *Clinical Cancer Research,* vol.13(2), pp. 559-565.

Kimmel, S.E., Sauer, W.H., Brensinger, C., Hirshfeld, J., Haber, H.L., and Localio, A.R. (2002). 'Relationship between coronary angioplasty laboratory volume and outcomes after hospital discharge', *American Heart Journal,* vol.143(5), pp. 833-840.

Klein, J.P. (2006). 'Modelling competing risks in cancer studies', *Statistics in Medicine,* vol.25(6), pp. 1015-1034.

Lau, B., Cole, S.R., and Gange, S.J. (2009). 'Competing risk regression models for epidemiologic data', *American Journal of Epidemology,* vol.170(2), pp. 244-256.

Laudicella, M., Li Donni, P., Smith, P.C., (2013). 'Hospital readmission rates: signal of failure or success?', *Journal of Health Economics*, vol32(5), pp. 909-21.

Lieberman, M.D., Kilburn, H., Lindsey, M., and Brennan, M.F. (1995). 'Relation of perioperative deaths to hospital volume among patients undergoing pancreatic resection for malignancy', *Ann.Surg,* vol.222(5), pp. 638-645.

Luft, H.S., Bunker, J.P., and Enthoven, A.C. (1979). 'Should operations be regionalized? The empirical relation between surgical volume and mortality', *The New England Journal of Medicine,* vol.301(25), pp. 1364-1369.

Luft, H.S., Hunt, S.S., and aerki, S.C. (1987). 'The volume-outcome relationship: practice-makes-perfect or selective-referral patterns?', *Health Services Research,* vol.22(2), pp. 157-182.

McArdle, C.S. and Hole, D. (1991). 'Impact of variablility among surgeons on postoperative mortality and morbidity and ultimate survival.', *British Medical Journal,* vol.302(6791), pp. 1501-1505.

McGrath, P.D., Wennberg, D.E., Dickens, J.D., Siewers, A.E., Lucas, F.L., Malenka, D.J., Kellett, M.A., and Ryan, T.J. (2000). 'Relation Between Operator and Hospital Volume and Outcomes Following Percutaneous Coronary Interventions in the Era of the Coronary Stent', *The Journal of the American Medical Association,* vol.284(24), pp. 3139-3144.

Melberg, T., Nygard, O.K., Kuiper, K.K., and Nordrehaug, J.E. (2010). 'Competing risk analysis of events 10 years after revascularization', *Scandanavian Cardiovascular Journal,* vol.44(5), pp. 279-288.

Pepe, M.S. and Mori, M. (1993). 'Kaplan-Meier, marginal or conditional-probability curves in summarizing competing risks failure time data', *Statistics in Medicine*, vol12(8), pp.737-751.

Pintilie, M. (2007). 'Analysing and interpreting competing risk data', *Statistics in Medicine,* vol.26(6), pp. 1360-1367.

Prentice, R.L., Kalbfleisch, J.D., Peterson, A.V., Flournoy, N., Farewell, V.T., and Breslow, N.E. (1978). 'The Analysis of Failure Times in the Presence of Competing Risks', *Biometrics,* vol.34(4), pp. 541-554.

Putter, H., Fiocco, M., and Geskus, R.B. (2007). 'Tutorial in biostatistics: competing risks and multi-state models', *Statistics in Medicine,* vol.26(11), pp. 2389-2430.

Ramanarayanan, S. (2008). 'Does Practice Make Perfect: An Empirical Analysis of Learning-by-Doing in Cardiac Surgery'. Working paper series, University of California.

Ritchie, J.L., Phillips, K.A., and Luft, H.S. (1993). 'Coronary angioplasty: statwide experience in California.', *Circulation,* vol.88(6), pp. 2735-2743.

Ross, J.S., Normand, S.L., Wang, Y., Ko, D.T., Chen, J., Drye, E.E., Keenan, P.S., Lichtman, J.H., Bueno, H.Ã., Schreiner, G.C., and Krumholz, H.M. (2010). 'Hospital volume and 30-day mortality for three common medical conditions', *The New England journal of Medicine,* vol.362(12), pp. 1110-1118.

Satagopan, J.M., Ben-Porat, L., Berwick, M., Robson, M., Kutler, D., and Auerbach, A.D. (2004). 'A note on competing risks in survival data analysis', *British Journal of Cancer,* vol.91(7), pp. 1229-1235.

Schrag, D., Panageas, K.S., Riedel, E., Cramer, L.D., Guillem, J.G., Bach, P.B., and Begg, C.B. (2002). 'Hospital and surgeon procedure volume as predictors of outcome following rectal cancer resection', *Annals of Surgery,* vol.236(5), pp. 583-592.

Schrag, D., Panageas, K.S., Riedel, E., Hsieh, L., Bach, P.B., Guillem, J.G., and egg, C.B. (2003). 'Surgeon volume compared to hospital volume as a predictor of outcome following primary colon cancer resection', *Journal Surgical Oncology,* vol.83(2), pp. 68-78.

Schreyogg, J., Stargardt, T., (2010), 'The trade-off between costs and outcomes: the case of acute myocardial infarction', *Health Services Research*, vol.45(6Pt1), pp. 1585-601.

Schwarzer, G., Schumacher, M., Maurer, T.B., and Ochsner, P.E. (2001). 'Statistical analysis of failure times in total joint replacement', *Journal of Clinical Epidemiology,* vol.54(10), pp. 997-1003.

Shook, T.L., Sun, G.W., Burstein, S., Eisenhauer, A.C., and Matthews, R.V. (1996). 'Comparison of percutaneous transluminal coronary angioplasty outcome and hospital costs for low-volume and high-volume operators', *Journal of American Cardiology,* vol.77(5), pp. 331-336.

Simunovic, M., To, T., Theriault, M., and Langer, B. (1999). 'Relation between hospital surgical volume and outcome for pancreatic resection for neoplasm in a publicly funded health care system', *Canadian Medical Association Journal,* vol.160(5), pp. 643-648.

Southern, D.A., Faris, P.D., Brant, R., Galbraith, P.D., Norris, C.M., Knudtson, M.L., and Ghali, W.A. (2006). 'Kaplan-Meier methods yielded misleading results in competing risk scenarios', *Journal of Clinical Epidemiology,* vol.59(10), pp. 1110-1114.

Srinivas , V.S., Hailpern, S.M., Koss E, Monrad, E.S., and Alderman, M.H. (2006). 'Effect of physician volume on the relationship between hospital volume and mortality during primary angioplasty', *Journal of the American College of Cardiology,* vol.53(7), pp. 574-579.

Tai, B.C., Machin, D., White, I., and Gebski, V. (2001). 'Competing risks analysis of patients with osteosarcoma: a comparison of four different approaches', *Statistics in Medicine,* vol.20(5), pp. 661-684.

Teixeira, L., Rodrigues, A., Carvalho, M.J., Cabrita, A., and Mendonca, D. (2013). 'Modelling competing risks in nephrology research: an example in peritoneal dialysis', *BMC Nephrology,* vol.14(110).

Tsai AC., Votruba, M., Bridges JF., and Cebul, R.D. (2006). 'Overcoming bias in estimating the volume-outcome relationship', Health research and educational Trust*, Health Services Research*, vol.41(1), pp.252-64.

Vakili, B.A., Kaplan, R., and Brown, D.L. (2001). 'Volume-outcome relation for physicians and hospitals performing angioplasty for acute myocardial infarction in New York state', *Circulation,* vol.104(18), pp. 2171-2176.

Varadhan, R., Weiss, C.O., Segal, J.B., Wu, A.W., Scharfstein, D., and Loyd, C. (2010). 'Evaluating health outcomes in the presence of competing risks: a review of statistical methods and clinical applications', *Medical Care,* vol.48(6(Suppl)), pp. 96-105.

Appendix A

This appendix includes the non-parametric CSHF and CIF for each of the events of interest: *failure*, *death* and *readmission.* The impact of changes in consultant volume levels, surgeon volume levels and consultant job plan contract type on the CSHF and CIF are presented.

Figure 3-19: Cause Specific Hazard Function – for the event *readmission*

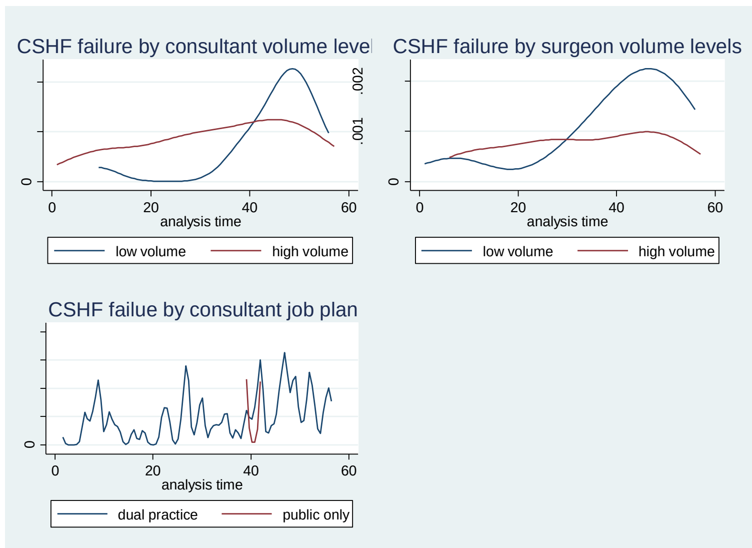


Figure 3-20: Cause Specific Hazard Function – for the event *death*

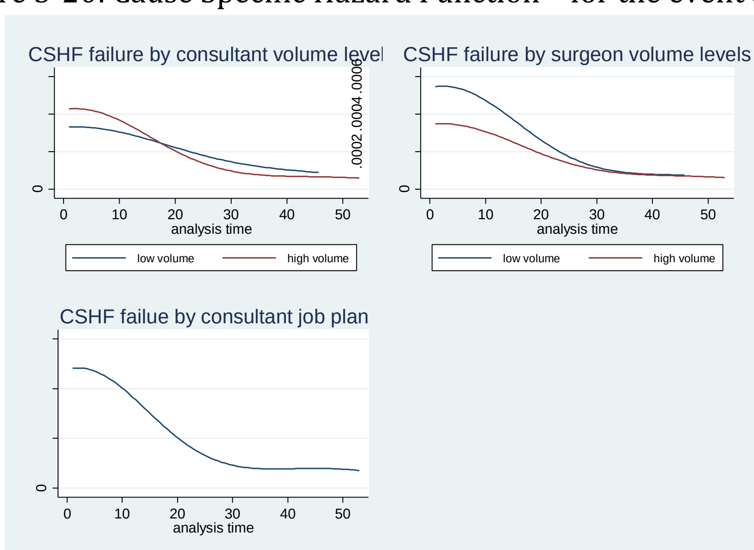


Figure 3-21: Cumulative Incidence Function – for the event *readmission*

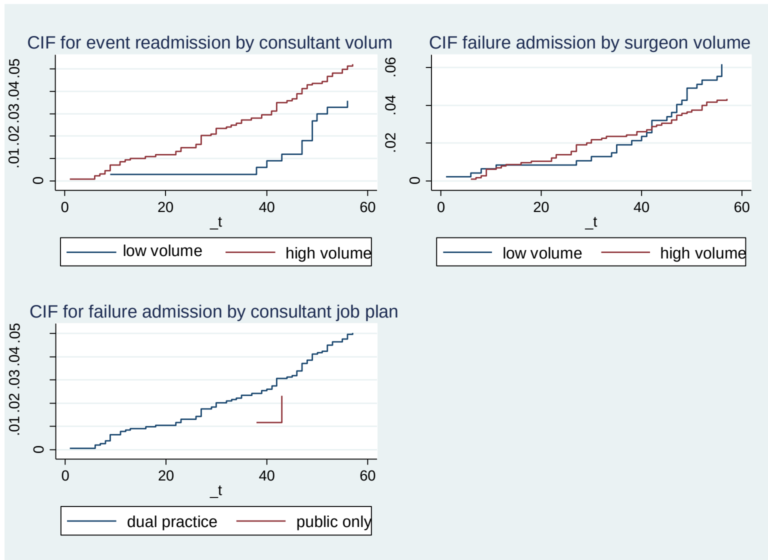
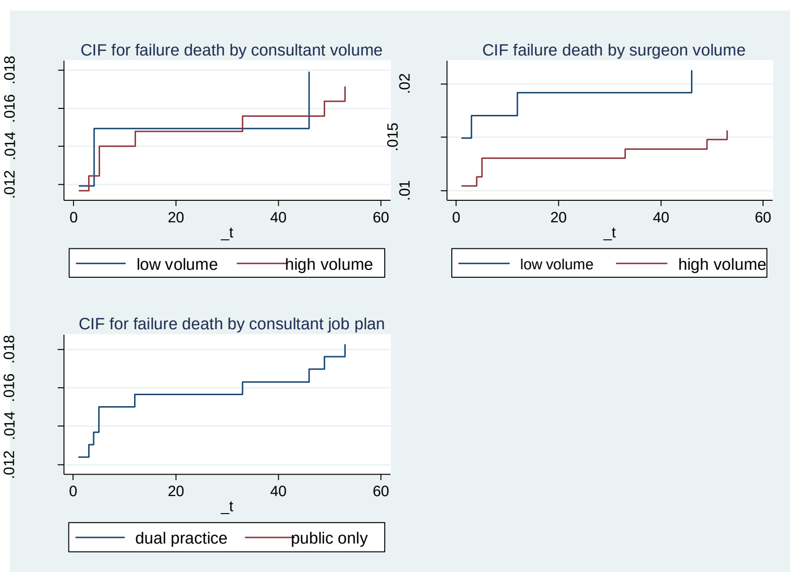


Figure 3-22: Cumulative Incidence Function – for the event *death*



Appendix B

The results obtained from a multinomial logistic model for outcomes within the 60 day period following the undertaking of the PTCA procedure are presented in Tables 3.15, 3.16 and 3.17. This is carried out to check for the robustness of the results obtained using the survival analysis estimation methods. To this effect results are compared to the estimation results obtained from the use of the CSHF and the CIF presented in Sections 3.5.2.1 and 3.5.2.2. Tables 3.15, 3.16 and 3.17 present the result of outcomes at 60 days for the estimations carried out using the different volume measures: *countc* and *counts*, *vol90c* and *vol90s* and *cvol* and *svol* respective. This type of modelling structure also allows for the control of the possible influence of additional variables on the *failure* event.

The first two columns within the tables represent the event *failure* which could constitute either of the events *death* or *readmission*. Columns 3-6 within each of the tables represent separate estimations for each of the possible events which could lead to failure.

The results in Table 3.15 confirm that there is a lower risk for a patient failure event after the PTCA procedure if the patient is under the care of a consultant on a *public-only contract* with the public hospital. As the volume of the consultant increases (countc) and the volume of the surgeon (counts) increases given that the odds ratio is very close to one then there is not likely to be any difference to the probability of failure. The relative risk of the particular failure from event *death* is expected to be higher as *age* and *LOS* increase.

Similar results are obtained in Table 3.16 which presents the relative risk of failure when the volume variables representing the most recent activity of consultants and surgeons prior to the patient hospital admittance date (*vol90c* and *vol90s*) are considered. These results are also comparable to the results in Table 3.8 using the CSHF and Table 3.12 using the CIF within the main text of the Chapter.

The results in Table 3.17 confirm the results obtained in Table 3.9 and Table 3.13 whereby the event *failure* for a patient is less likely if a patient is under the control of *public-only* contract consultants. Patients being seen by consultants on high volume levels are more likely to have an event *failure* (although insignificant). Patients being seen by surgeons on high volume levels are less likely to have a failure event during the 60 day period. The results also confirm that the relative risk of failure from death is expected to be higher as age and LOS of patients increase.

Table 3.15: Multinomial logistic regression (countc and counts) - odds ratios

	(1)	(2)	(3)	(4)	(5)	(6)
Event	Failure	Failure	Re-admit	Re-admit	Death	Death
Variable						
age	1.002	1.003	0.984	0.985	1.068***	1.070***
LOS	1.051***	1.054***	1.034*	1.036*	1.078***	1.081***
public-only	0.234**	0.286*	0.313	0.379	-	-
countc	0.998*		0.999		0.998	
counts		0.999*		0.999		0.998
Constant	0.066***	0.060***	0.169***	0.155***	0.000***	0.000***
Observations	1,626	1,626	1,626	1,626	1,626	1,626

*** p<0.01, ** p<0.05, * p<0.1

The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.

Table 3.16: Multinomial logistic regression (vol90c and vol90s) - odds ratios

	(1)	(2)	(3)	(4)	(5)	(6)
Event	Failure	Failure	Re-admit	Re-admit	Death	Death
Variable						
age	1.001	1.002	0.983	0.984	1.067***	1.066***
LOS	1.057***	1.055***	1.038*	1.037*	1.088***	1.084***
public-only	0.298*	0.282*	0.376	0.370	-	-
vol90c	1.002		1.000		1.008	
vol90s		0.998		0.997		1.002
Constant	0.050***	0.056***	0.140***	0.150***	0.000***	0.000***
Observations	1,626	1,626	1,626	1,626	1,626	1,626

*** p<0.01, ** p<0.05, * p<0.1

The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.

Table 3.17: Multinomial logistic regression (cvol and svol) - odds ratios

	(1)	(2)	(3)	(4)	(5)	(6)
Event	Failure	Failure	Re-admit	Re-admit	Death	Death
Variable						
age	1.002	1.004	0.984	0.986	1.068***	1.069***
LOS	1.061***	1.053***	1.044**	1.035*	1.088***	1.080***
public-only	0.380	0.252*	0.507	0.334	-	-
cvol	1.424		1.436		1.412	
svol		0.668*		0.678		0.657
Constant	0.037***	0.062***	0.096***	0.157***	0.000***	0.000***
Observations	1,626	1,626	1,626	1,626	1,626	1,626

*** p<0.01, ** p<0.05, * p<0.1

The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.
The baseline, *cvol*=0 indicates consultant *low* volume, *svol*=0 indicates surgeon *low* volume.

The effect on the occurrence of a failure event, arising from differences in the consultant job plan type, given a change in volume is studied by adding an interaction term to the analysis. Results are shown in Table 3.18. These results can be compared to those in Table 3.10 (based on the CSHF) and Table 3.14 (based on the CIF). Overall similar conclusions can be made when interpreting such results. If volume is increased by one unit and the patient is under the care of a consultant with a dual practice contract, with all other variables kept constant, the risk of failure is practically unchanged. If volume is increased by one unit and the patient is under the care of a consultant on a *public-only* contract then the relative risk of failure is expected to increase by a factor of 1.069[44]. The impact of a change in volume on the relative risk of the event *failure* is found to be higher for patients under the care of *public-only* consultants. The relative risk of patients from the event *failure* for patients under the care of consultants with *public-only* contracts is less compared to those being under the care of *dual practice* consultants (at zero volume levels and keeping all other variables constant). In view of the interaction term applied within the multinomial logistic regression model, the magnitude of this change varies at the particularly set volume levels.

[44] Obtained as exp(-0.002+0.069)* countc). Values represent the coefficients corresponding to the sub-hazard ratios provided in Table 3.18.

Table 3.18: Multinomial logistic regression output (with interactions)

Event Variables	(1) Failure	(2) Failure	(3) Re-admit	(4) Re-admit	(5) Death	(6) Death
age	1.002	1.002	0.984	0.983	1.068***	1.067***
LOS	1.052***	1.058***	1.034*	1.039**	1.078***	1.088***
public-only	0.087	0.090*	0.125	0.115	-	-
countc	0.998*		0.999		0.998	
countc x *public-only*	1.072		1.066		1.006	
vol90c		1.002		0.999		1.008
vol90c x *public-only*		1.162		1.156		1.022
Constant	0.066***	0.050***	0.168***	0.139***	0.000***	0.000***
Observations	1,626	1,626	1,626	1,626	1,626	1,626

Coefficients reflect odds ratios, *** $p<0.01$, ** $p<0.05$, * $p<0.1$
The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.

Similar conclusions are derived from the results in columns 3-6 of Table 3.18 for the different particular failure events of *death* and *readmission* when compared to results in Tables 3.10 and 3.14. The public-only variable coefficients are though not statistically significant in most of the cases.

Appendix C

Table 3.19: Proportional Hazard test based on the 'estat phtest' command in Stata.

p value	Table 3.7	Table 3.8	Table 3.9	Table 3.10
Equation 1	0.0567*	0.0524*	0.0174**	0.0814*
Equation 2	0.0543*	0.0253**	0.0360**	0.0573*
Equation 3	0.9272	0.9049	0.1387	0.9239
Equation 4	0.9242	0.6843	0.3506	0.8030
Equation 5	0.9328	0.9951	0.9928	0.9743
Equation 6	0.6706	0.8029	0.9905	0.9990

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Chapter

4

Modelling Consultant Practice Heterogeneity arising from Contract Type within the Context of an Acute General Public Hospital

Abstract

This paper reviews the behaviour of consultants working in the context of very specific incentives and work practices within the Maltese Islands' only public hospital. It focuses on the extent to which the consultants' job plan affects practice variation. We exploit the data available on the investigations carried out at patient level in the undertaking of the Percutaneous Transluminal Coronary Angioplasty (PTCA) procedure. These investigations were further subdivided into two categories: those performed within the first two days of hospital stay and those performed during the remaining hospital stay period. Practice variation within each of the two categories is analysed in this paper. The analysis found that the job contract type of the consultant had a significant effect on practice variation even after controlling for patient casemix. Furthermore, the impact on practice variation due to differences in job contract type was significantly different when the investigations were compared before and after the first two days of hospital stay.

4.1 Introduction and Motivation

Variations in practice patterns have important consequences for the quality and cost of health care treatment. Medical variation is the difference in the use of health care services or resources between individuals or groups of patients who share a similar health status or medical condition. In the literature, such variation in practice patterns has often been linked to patients not receiving the best possible care or receiving extra care, which in turn leads to inappropriate and inefficient use of available resources. Is variation always necessarily bad? Some variation might be expected in a health care setting but it is the magnitude of the variation which causes concern. Wennberg and Gittlesohn (1973) have argued that a reduction in variation will lead to improved efficiency, have an impact on patient welfare and thus have an impact on health outcomes.

This paper studies the behaviour of consultants working in the context of very specific incentives and work practices created within Mater Dei Hospital (MDH), the only publically funded general acute hospital on the Islands of Malta. The consultant is the hospital professional who is primarily responsible for the patient care. Although many studies focus on variation between hospitals or between different geographic areas, this study will look at variation at the individual consultant level within a single hospital. Work by (Ellis and McGuire, 1986; Harper et al., 2001) and Jacobs et al., (2006) shows a preference for comparison between departments rather than whole hospitals to ensure that similar specialty activities are analysed. In this study it will be assumed that consultants decide on treatment practices after considering all the conditions and characteristics of the health care system within which they practice. The importance of the consultant contract job type in explaining practice heterogeneity will be the study's primary focus.

Data from the Maltese health care system is used in this study. Patients are assigned to consultants, who take on the responsibility for the patient's case on the basis of the diagnosis made and the date in which they are admitted to hospital. Patients have no choice of the hospital they can attend, although they can decide to 'opt out' and visit a private hospital/clinic[1], if they so wish.

In this paper, practice variation will be measured using the *number of investigations* ordered and performed on a patient who undertook a PTCA proceduce during their hospital stay. PTCA is a procedure used to widen narrowed arteries. In the procedure the cardiologist inserts a catheter with a deflated balloon at its tip into the narrowed artery. The balloon is then inflated, helping to compress the plaque and thus enlarge the blood vessel. Blood could then flow more easily. In comparison to open heart surgery this procedure is less invasive and in general considered to be less expensive. The choice of the PTCA procedure for this study was made for a number of reasons. PTCA is a relatively standard procedure for patients with fairly similar characteristics. This makes it a good candidate for testing for the effect of variation in treatment which could be possibly linked to medical practice. Given that it's a fairly standard procedure one would expect the treatment to be offered to such patients to be relatively similar.

[1] A number of small private clinics and hospitals can be found in Malta. Activity levels in such sectors are relatively low and it very much depends on a small private health insurance market.

There is in fact a very limited number of ICD-9-CM specific codes which are used to define this procedure. This should make the group easier to manage, thus treating a less diverse set of patients limiting heterogeneity in the treatment process across this particular procedure. Furthermore, PTCA is a procedure performed by highly trained and specialised doctors whose abilities and good judgement are considered to be of importance to the success of the intervention.

The consultant responsible for the admitted patient has direct control over the number of investigations carried out at the hospital. Variation in such decisions could reflect a consultant's strategic behaviour due to their type of contract and/or the consultant's treatment preferences. The relationship between the specific consultant contract conditions and the number of investigations ordered by the consultant will be tested in the empirical part of this paper.

This paper will add to the literature by analysing how practice variation differs by consultant job contract type in relation to the different time periods during which the patient is in hospital. The available data allow for the number of investigations carried out at the hospital to be further subdivided into activity undertaken in the first two days of the patient's stay in hospital and that carried out in the remaining time the patient spends at the hospital. These sub-samples of data will be used to identify whether differences in the number of investigations within each period vary with different consultant job contract types.

The activity undertaken in the first two days following admission is considered by the consultant to be an *absolutely necessary* part of the procedure that is being performed.

Some of the investigations performed in the remaining part of the hospital stay, though important for the success of the procedure, might reflect behaviour variation in consultant practice. This paper will test how much of this variation is due to the particular job contract type of the consultant. A two-stage multilevel modelling approach will be adopted given that patient level data can be clustered within consultant categories.[2] This will identify variation arising from patient characteristics and/or from consultant related differences. An Estimated Dependent Variable model (EDV) will be used to identify reasons for consultant related variations. This level of analysis ensures that the causes behind such variations are clearly identified, which otherwise might not be the case if the analysis is undertaken at an aggregated level.

The results presented in this paper will show that both the job contract type and the volume levels of the consultant significantly affect practice variation. The existence of higher practice variation linked to consultant characteristics in the post two day hospital stay period is also verified. The fact that patients are being seen by consultants who can dual practice increases the practice variation especially within the post 2 day hospital stay period. There is a compounding impact of the job plan contract type on the relationship between volume and practice variation whereby the responsiveness of practice variation to volume changes is positively stronger for consultants working exclusively with the public hospital.

The paper is organized as follows: the following section, Section 4.2, provides an overview of the literature relating to consultant behaviour variation within a number

[2] A one-stage multilevel model is also estimated to check for the robustness of the results.

of institutional settings, identifying the main sources of variation and how these impact on consultant behaviour. Section 4.3 introduces a theoretical model specification which will be used to motivate the empirical approach, described in Section 4.4. Section 4.5 describes the available population based dataset highlighting some descriptive information. This is followed by a discussion of the key results derived from the estimations. The final section, Section 4.7, presents the study's conclusions.

4.2 Literature

In this section a review of the literature that relates to the identification of the most important factors that influence practice variation is presented. The modelling of these factors within a utility function for the consultant, as reviewed by a number of authors, is then critically assessed. This section closes with an assessment of the expected consultant behavioural responses that result from changes in the main determinants of utility.

Determinants of practice variation: The literature on health care practice variation is extensive and diverse, and primarily seeks to identify the determinants of care variation. The interaction between the patient and the consultant partly explains behaviour but there is also an impact from the joint interaction among consultants, patients and the hospital. The variation in hospital and consultant utilization rates across different areas of clinical practice has been widely documented and three key variables are identified as influencing the consultants' strategic behaviour: consultant and patient characteristics, current practices within the organization and environmental setting characteristics.

Grytten and Sorensen (2003) show that practice variation by diagnosis and treatment type explains more than 40% of the variation in expenditure on laboratory investigations, consultations and expenditure on specific procedures in Norway. Variation is said to persist even when age, gender, and socio-economic variables are controlled for. Practice patterns are the result of the training and clinical experience gained by consultants and there are a number of consultant practice patterns which persist over different clinical scenarios, as highlighted by O'Neill and Kuder (2005). In a paper by Mercuri et al., (2012), practice variations exist even in very small geographic areas despite adjustment for casemix and the fact that consultants operate using similar resources. In a number of other studies which deal with variations in practice style within particular centres (de Jong et al., 2006; Hayward et al., 1994; Westert et al., 1993), mixed results are obtained with consultants choosing a patient Length of Stay (LOS) which is in line with the Average Length of Stay (ALOS) of the particular hospital in which they operate.

The majority of the literature focuses on patient and consultant behaviour under different institutional frameworks. Early studies in the '60s and '70s linked variation to patient context and circumstantial differences (Andersen and Newman, 1973; Anderson and Mooney, 1990; Ro, 1969; Wennberg and Gittlesohn, 1973) whereas more recent research (Allard et al., 2011; Biglaiser and Ma, 2007; Hennig-Schmidt et al., 2011; O'Neill and Kuder, 2005) has associated variation to the differences in behavioural and distributional factors related to remuneration systems and workload levels for consultants rather than differences related to patient related illness. The consultant treatment and referral decisions serve to influence the patient's expected utility and the prevalence of treatment variation leads most scholars to conclude that uncertainty is the main factor leading to variation from what constitutes the *optimal* treatment process.

The consultant is an agent of an uninformed patient and Wennberg (1985) claims that uncertainty is the most important factor influencing physician behaviour. This uncertainty could be related to the initial health status or unknown patient preferences, described by Pauly (1980) as *irreducible uncertainty,* or even the absence of information on the consequences of health care treatment that is shared equally by the doctor and the patient. The unexplained residual variance in the function used to explain practice differences would serve as a proxy for the uncertainty attributed to physician *practice style*. However, practice style maybe just one of all possible divergences and thus one cannot attribute all variation to this factor.

Indeed, consultants are assumed to not only consider their own welfare but to balance their own gained welfare against the welfare of their patients in their decision making. Ellis and McGuire (1986) build a model on the premise that benefits to the patient and to the hospital are the main arguments in the utility function of the consultant. The role of ethics in the treatment of patients is given priority by Evans (1974) whilst financial returns are considered as the main variable influencing consultant behaviour by Pauly and Redish (1973) and Pauly (1980). The concern for the social good as a determinant of utility gained by a consultant is also considered by Eisenberg (2002).

A consultant's instructions and advice are an institutional reality in the health care sector and Farley (1986) concludes that the pattern of medical treatment is largely determined by consultants and not by the result of marginal choices made by patients. Eisenberg (2002) finds that around 90% of the health care expenditure is a direct result of consultant decisions in the choice of clinical practice. Although this does not prevent patients from being involved in the decision making process, in most cases it means that the course of treatment is not controlled by the patient even though it is the patient who decides when to seek and start treatment. This puts consultants in a position where they can actually allocate resources. Indeed, consultants adopt their own individual strategies in their endeavour to maximize their own utility function against a number of constraints.

Particularly in view of the given size of the market, it is likely that the consultant undertaking the procedure in question holds information on other consultants offering the same or similar services. Competition among experts serves to limit the potential exploitation of the consumer (Winand, 1997; Wolinsky, 1994) and as noted by Pauly and Satterthwaite (1981), the amount of information which patients have of the practice of the consultants in relation to the number of consultants available in the market also affects consultant decisions. It is likely that in cases where there are only a few consultants operating in a particular field, each of the consultants would have a reputation which is well known within the community and most patients would also be well informed about such consultant characteristics. Consultants expect to be judged by their patients on the basis of the benefit or gain they expect to receive from treatment.

Modelling of consultant behaviour: Consultants have different beliefs, work within different hospital environments and job contract conditions and consider their own income streams in their decision making. Nevertheless, they are all expected to take into account the benefit to the patient in their decision making. Consultant behaviour has over the years mainly been modelled on income, leisure and inducement, whereby income and leisure are expected to positively impact utility whilst inducement is expected to impact it negatively (Pauly and Redisch, 1973; Evans, 1974; Pauly, 1980; Wilensky and Rossiter, 1983; Ellis and McGuire, 1986; McGuire and Pauly, 1991; Blomqvist and Leger, 2005; Gruber and Owings, 1996; Grytten and Sorensen, 2003; Rickman and McGuire, 1999)). De Jaegher and Jegers (2000) add to this that the number of patients in the clientele of the consultant should be considered as part of the utility function of the consultant.

Papers by Evans (1974) and Fuchs (1978) model consultant behaviour by maximizing a utility function which includes income and inducement as arguments of interest. McGuire and Pauly (1991) model consultant behaviour by adding leisure to the utility function. Target income behaviour and profit maximization lie at opposite ends of a spectrum of income effects. Evans (1974) defines inducement as the "persuasive activity of the consultant to shift the patient's demand curve according to the consultants self-interest".

Indeed, inducement could be a source of additional income to the consultant. However, because of ethical considerations, this is assumed in most work to negatively affect overall consultant utility. In this context inducement means a loss in utility and consultants balance the marginal utility of income against the disutility of exploiting patients to obtain it (Evans, 1974; Pauly, 1980; Wilensky and Rossiter, 1983). There is a moral cost of providing inappropriate medical treatment and there is an expected cost due to peer review, loss of reputation and loss of future patients. According to Dranove (1988), there are decreasing returns to consultant inducement because patients can begin to disbelieve their consultant and decide to shift to another consultant.

On the determinants of doctor's behaviour and dual practice: It is of interest to evaluate the response of consultant behaviour to changes in the factors which affect utility. The behaviour of the consultant (to hospitalize, to treat or to take time to diagnose carefully) is influenced by both the financial and non-financial incentives of the job. Do consultants actually seek to achieve some level of target income? If consultants are employed on a fee per patient system then greater competition will make consultants further exploit the 'information advantage' i.e. their knowledge and information they have to hand, to see more patients (Feldman and Sloan, 1989; Rice and Labelle, 1989). On the other hand, if consultant income is independent of output, there is no incentive to induce activity within the same public hospital. Some level of inducement to the private sector may still be evident although difficult to measure. Indeed, one needs to distinguish between volume (first visit) inducement and intensity response inducement. Wennberg et al., (1982) assumes that practice style affects only intensity demand, a finding confirmed by the work of Mitchell and Sass (1995) and Mitchell et al., (2000), whereby supplier inducement effects were found for the provision of laboratory investigations and ancillary services.

Consultants on a fixed salary are expected to offer the most comprehensive treatment option to their patients by engaging in the provision of a complete set of possible investigations together with the basic necessary procedures. This is in contrast to providing only a minimum number of investigations over and above the basic necessary procedures. The provision of the most comprehensive treatment serves to enhance the consultant's image and reputation as highlighted by Gonzalez (2004) apart from the fear of dealing with a malpractice lawsuit, as identified by Gal-Or (1999). It is common in some hospital settings that health care authorities set out treatment guidelines to support the diagnosis of patients so as to ensure a streamlined level of service. Carlsen et al., (2003) finds some contradicting evidence, whereby the effect of payment schemes on the number of consultations and the number of ordered laboratory investigations is not found to be as strong.

Consultant behaviour is also expected to be influenced by the particular and specific conditions of work that defines the relationship between the consultants and the provider of health care services. In the literature there are mainly two distinguishable substitution effects that affect the labour supply of doctors: one is between labour and leisure and the other between public and private sector health care practices. Of particular relevance to this study is the latter given that the majority of consultants working at the public hospital have *dual practice* contracts. This particular contract type, *dual practice*, means that the consultant, apart from being an agent of the patient and the public hospital, has also to consider his own return from private practice in decisions to maximize his utility.

A number of papers have dealt with consultants' response to the form of consultant job contract type (Blomqvist, 1991; Ellis and McGuire, 1990; Ellis and McGuire, 1986). In their review of the dual practice literature, Socha and Bech (2011) did not find conclusive evidence as to the effect of dual practice on public health care although the negative effects seem to be apparent. In fact, dual practice consultants are expected to have good income opportunities within the private sector and might decide to concentrate more of their efforts in their private sector activity with negative implications on availability, costs and quality for the public health care sector.

Patients treated at the public hospital are assigned to consultants who are either engaged by the hospital on a *public-only* contract or on a *dual practice* contract. The consultant is the key decision maker who selects the level of services provided to a patient. The income earned by the consultant from the public hospital is independent of the number of cases seen or the number of investigations ordered. However, consultants who *dual practice* are also interested in their future private income stream and this is expected to be affected by their overall success and reputation gained from the activity undertaken at the public hospital. Gonzalez (2004) concludes that when allowing *dual practice*, consultants use the public sector to improve their reputation and thus earn more revenue from their private practice.

.

The prevalence of *dual practice* contracts may indeed present a puzzle as it is at odds with basic labour supply theories which predict that individuals normally prefer to work for longer hours in their higher paid job instead to undertaking multiple jobs (Lang 1994). Indeed, theories of incentive design predict that employers seek to ensure that their workers do not divert their attention to other tasks or jobs (Holmstom 1999). There must be some form of complementarity in allowing *dual practice* in terms of benefits for the consultants and the government/hospital alike. A number of models of dual practice reviewed by Eggleston (2006) conclude that this depends on the ability of the government to monitor activity within the hospital whereby the potential social costs of dual practice have to be set against the costs of enforcing stricter restrictions.

The possibility of dual practice gives a kind of performance based incentive to consultants to increase their effort levels. When allowing private practice the government is giving out a salary together with a non-wage 'benefit' to the consultants. This ensures that the government recruits good quality doctors at relatively modest rates. The possibility of private practice is seen by consultants as an opportunity to increase strategic influence, clinical autonomy and the realisation of individual aspirations as a clinician. Higher skilled consultants potentially anticipate higher earnings from private practice and thus the possibility of *dual practice* for high skilled consultants is more valued if the individual thinks that he is in a better position to have a lucrative private clinic clientele. As consultant skill increases the marginal benefit of private practice is expected to increase.

The temptation to earn private profits also serves to attract the best consultants to the public hospital, thus enhancing the quality of the public service. Gonzalez (2004), finds that there is an incentive for *dual practice* doctors to enhance their reputations by practising 'over provision' of health care services to earn a good reputation and thus support their private practice. The consultants private practice could indeed be using public facilities to treat private pay patients (Gruen 2002, Ferrinho 2004). Furthermore, results of tests carried out at the public hospital could be discussed at the private clinic (Mitchell JM 1995 and MsGuire 2002). Such a practice is not without its drawbacks as consultants could practice *skimping* during their work hours in the public hospital. Physicians using public resources to treat their private clients would clearly undermine the efficiency of the overall public health care service apart from gaining from a cost advantage over those who only practice privately.

Bir and Eggleston (2003) consider access enhancement whereby public providers practicing dual practice make use of the public services to treat their private patients. Furthermore, in the context of free public care provision and the availability of private care, Brekke and Sorgard (2007) conclude that allowing dual practice would lead to the use of the public health services by those who would have otherwise purchased care from the private sector. This has the effect of *crowding out* private financing and creates a burden on society as a result of the additional taxation which will be needed to finance the increase in public expenditure. *Moonlighting* by consultants might though lead to a more efficient health care service within the public sector, as savings made as a result of patients deciding to use private treatment, could be used to improve the quality of the public health care (Biglaiser and Ma, 2007).

Bloor et al., (2004) looked at the variation in activity rates of consultant surgeons working in the National Health Service (NHS) in the UK. Consultants with freedom to carry out private practice had high NHS activity rates compared to those hired on a full-time contract. Furthermore, if private sector activity is reduced this would not necessarily increase NHS activity and this could become a possible barrier to increasing productivity within the NHS. Consultants only involved in public health care provision are assumed by Socha (2010) to be more mission oriented and thus better represent the interests of the public. Gonzalez (2005) concludes that dual practitioners are usually assumed to try to maximize income which favours longer waiting times in order to boost their private practice income through *cream skimming*. The reputation of such consultants in the public hospital is also at stake and such consultants are inclined to provide the best service at the public hospital to improve their reputation.

The current literature acknowledges that the availability of patient level data allows the partitioning of practice variation into that which arises from individual factors and that which is attributable to providers within which such individuals are clustered. The methods adopted in the treatment of hierarchical structured data vary in relation to the aims of the study and the limitations imposed by the available data sources. The use of a single level model which ignores the natural clustering of the data structure assumes that units are independent. This approach ignores that the patient clustering imposes some correlation structure on the data and this invalidates classical OLS assumptions leading to inefficient estimates. Moreover, a single level approach to the analysis of such hierarchically structured data will fail to exploit the full richness of the information contained within and between the various levels of data. However, the particular sample size considerations of this dataset might suggest that there could be potential gains of 'borrowing strength' across groups and thus seek to increase efficiency for the obtained estimates through the use of a single level model.

Another approach suggested by Rice and Leyland (1996) is to undertake an aggregate level analysis by using group level mean values. Amongst the drawback with this method are the problems envisaged by the so called *ecological fallacy*[3] and the fact that statistical estimates can be unreliable resulting in large standard errors (often due to the high collinearity between explanatory variables). It is also of concern that the aggregation of variables tends to provide unreliable estimates when the number of individuals within the higher level groups are small as inference would be based on small samples. This is a cause of concern in view of the sample properties of the dataset used in this study. Olsen and Street (2008) recommend the use of panel data estimators within a multilevel modelling framework given that the number of consultants is too small to apply models at the consultant level with averaged patient data.

Multilevel models address the estimation problems that arise from the correlation structure of clustered data, attributing variability in the dependent variable to the specific hierarchical levels of the model. The application of multilevel modelling

[3] Ecological fallacy refers to the case where aggregate level associations are wrongly inferred to exist at the individual level.

techniques in the field of health economics has become more common following the work of (Rice and Jones, 1997; Rice and Leyland, 1996). Individual patients (micro units) can be clustered within particular groups (macro units) on the basis of the common characteristics which they possess. Individuals under the control of a particular consultant are expected to receive similar treatment compared to patients with similar conditions but under the responsibility of different consultants.

The use of the multilevel modelling technique allows the researcher to investigate the nature of the between group variability and the effects of group level characteristics on individual outcomes thus capturing more information from the hierarchically clustered available data (Hvenegaard et al 2009). Various types of estimators have been suggested in the literature to estimate the consultant effect parameter[4] with the random or fixed component effects commonly used to obtain the consultant effect. The choice of whether a random effects or a fixed effect component model is to be used requires careful consideration and the choice of method may be determined by the data generating process and also the type of inference sought. In this regard, the fixed effects model allows for the separation of the effect of patient characteristics from consultant characteristics which could be correlated with patient characteristics (Cookson and Laudicella 2011).

---

[4] Schmidt and Sickles (1984) propose the use of a fixed effects model, Kumbhakar and Lovell, (2000) propose the use of a random effects model estimated by GLS, Pitt and Lee, (1981) propose the use of a random effects model estimated by Maximum Likelihood. The pros and cons of each of these model specifications are summarized in a table provided by Jacobs et al., (2006) pp. 74.

Laudicella et al., (2010) find that the health care resource group of the patient captures much of the variation in the costs of treatment in the NHS, however, other factors are also important in explaining such differences. In fact, more information could therefore be obtained from the available patient level data if the hierarchically clustered data is taken into account.

4.3 Theoretical model

In order to motivate our empirical question of whether there is variation in practice at a consultant level due to contract variation, we model a basic utility maximization problem of a consultant providing services within the unique public hospital. The consultant is employed in the public sector and can practice some form of dual practice i.e. working a number of hours per week working in the private sector. Consultants operate under a contract, *C,* that can be a full time contract with the public hospital or part time, which allows them to practise in the private sector. The consultant has a number of decision variables under his control, primarily the *number of investigations, T,* that are ordered to be performed at the hospital, and the level of effort the consultant devotes to the health care system, *e*.

As the consultant's private sector activity cannot be observed, the focus of the study will be on consultant behaviour within the public sector. The consultant obtains utility, $U()$, from his income, *Y,* where $U_Y>0$, and a reputation gained from his rate of success in the provision of health care services, *S,* where $U_S>0$. His income is contingent on the type of contract he has, *C.* Success, *S*, is dependent positively on the number of investigations, *T,* the consultant's experience measured as volume of patients treated, *V,* patient casemix characteristics, M, and the effort exerted, *e.* Finally, effort, *e(),* creates disutility, U_e<0 and depends positively on volume, $e_V>0$, the consultant type of contract, *C* and on patient casemix characteristics, *M*. Thus, a function of utility can be expressed as:

$$U = U(Y(C), S(T, V, M, e), e(V, \mathrm{C}, M))$$

The constraint the consultant faces is the number of hours/activity he can work in each sector and the amount of leisure time he desires. The type of contract chosen by the consultant significantly affects the number of hours available to treat patients within the private sector. Thus, the consultant chooses an optimal number of investigations *T** and *e** in the public sector that maximizes his utility subject to his time and income constraints, ie., those solving the problem

$Max_{T.e}\, U\big(Y(C), S(T, V, M, e), e(V, C, M)\big)$, subject to time and income constraints which also depend on the type of contract, *C.*

The First Order Conditions, (FOC), of this maximization problem[5] implicitly define reduced forms for how the optimal number of investigations, *T*,* and the effort level, *e*,* depend on the exogenous variables, i.e. *C,V,M.*

[5] i.e., $\frac{dU}{dT} = \underset{+}{U_s} \underset{+}{\frac{\delta S}{\delta T}} = 0$ and $\frac{dU}{de} = \underset{+}{U_s} \cdot \underset{+}{\frac{\delta S}{\delta e}} + \underset{-}{\frac{\delta U}{\delta e}} = 0$, where the signs below each expression indicate the sign of the marginal utility terms.

As we do not observe effort e , the empirical section associated with this study will concentrate on estimating the reduced form of the consultant's number of investigations, T^*, which depends on C and V, given the consultant and patient characteristics. This also takes into account the particular characteristics of the Maltese health care setting whereby the number of investigations ordered by the consultant, T* is expected to be the prime variable available to the consultant to impose his sytle of treatment onto the patient. To illustrate this we will use data on patients who have received PTCA treatment within the public hospital. The focus will be the number of investigations ordered by the consultants and administered during the patients' stay in hospital.

4.4 Empirical approach

This section describes the empirical approach used to investigate consultant practice variation by using the number of investigations performed at patient level. This involves looking at the importance of consultants' contract type as the prime determinant to explain practice variation. The clustering of patients under different consultants imposes a correlation structure on the data, reflecting the shared experiences of being treated by the same consultant. Two randomly selected individuals treated by the same consultant would be expected to receive more similar treatment than two individuals being treated by different consultants. The variation in treatment between consultants could also depend on the fact that consultants also treat a different mix of patients.

The possibility of using patient level data ensures that more detailed information rather than just DRG related data could be used to obtain information on the patient (Hvenegaard et al., 2009). The use of the available administrative (not sample) patient level data helps to deal with differences in patient characteristics whilst offering analytical advantages in terms of analysis of consultant characteristics affecting practice variation. Inference about the variables of interest will be more robust if based on individual rather than aggregated data as standard errors will be more precisely estimated (Rice and Leyland 1996).

Given an individual based dataset, a two stage model is used to reflect the literature on the use of Estimated Dependent Variable (EDV) models in applied political studies, as presented by (Karen and Shively, 2005; Laudicella et al., 2010; Lewis and Linzer, 2005) as well as a number of studies relating to pupil achievement within the education system[6], as reviewed by Rice and Leyland (1996).

A first stage regression of the number of investigations will be undertaken at patient level against a number of patient characteristics which are likely to explain and determine the number of investigations for the patient:

$$y_{ij} = \alpha + \beta_x x_{ij} + \mu_j + \varepsilon_{ij} \qquad \text{(1}^{\text{st}} \text{ stage regression)}$$

where y_{ij}, is the number of investigations carried out for patient i as ordered by consultant j and this is the variable of interest used to measure variation in practice style. The vector of patient characteristics, x_{ij}, comprises the patient's age and a

[6] Karen and Shively (2005) and Lewis and Linzer (2005) discuss the use of EDV models within a two-stage approach providing consistent and efficient estimates.

measure of severity represented by the Charlson Comorbidity Index. The term μ_j is a consultant time-invariant fixed effect[7] that captures unobserved heterogeneity across consultants over and above that described by the explanatory variables in the regression. This term provides each consultant's average contribution to the number of investigations per case after controlling for the explanatory factors which also affect the number of investigations undertaken.

The proportion of variance attributed to the consultant can be judged as an indicator of the degree of influence of the consultant on the number of investigations recommended. A second stage regression will deal with factors which explain such variation:

$$\hat{\mu}_j = \pi_1 + \pi_2 \chi_j + \pi_3 C_j + v_j \qquad \text{(2}^{\text{nd}}\text{ stage regression)}$$

where $\hat{\mu}_j$ represents the consultant fixed effects estimated from the first stage regression, χ_j is a vector of variables capturing particular consultant characteristics, apart from the job contract type of the consultant, C_j, which is reviewed separately in this study. A number of different measures of consultant volume levels constitute the vector χ_j. Of primary interest to this study is the relevance of the consultant job contract type, C_j, to explain variations in $\hat{\mu}_j$. The use of the Efron robust standard error is implemented to deal with the heteroscedastic sampling errors of the estimated dependent variable which might result in biased standard errors for the second stage

[7] The purpose of the estimation process is to generate inferences about individual consultants. The table provided by Jacobs et al., (2006) provides a summary of the advantages and disadvantages of the different model specifications.

regression. The use of Efron heteroscedastic standard errors applied in this study are said to yield consistent standard error estimates even in small samples (Lewis and Linzer 2005)[8], even though the OLS estimator may be inefficient.

The approach described above makes use of the multi-level structure of the available data to analyze data grouped into hierarchical macro units (consultants), with a number of micro units (patients) within each, to obtain the consultant fixed effects. The subscript *j* introduces an important characteristic to the regression as it allows the intercept and coefficients to vary across consultants to ensure that independently distributed consultant fixed effects are obtained (Laudicella et al., 2010). The two stage model has the advantage of allowing for the partition of the overall variation into that due to differences in patients and that arising from differences in the consultant propensity to prescribe investigations or tests. In this instance, higher level effects are not viewed as nuisance parameters but are of central importance to the analysis. This would lead to a better exploration of the relationship between individuals and the contexts in which they belong. Another advantage of the two-stage model is that it parallels closely the data generation process, which consists of aggregating patient data into consultant clusters.

Two assumptions are important to note with regards to the data used in the model. Separability is assumed between the variables of interest to ensure that the influence of patient characteristics can be singled out and removed from the influence of

[8] Efron standard errors are based on the Jacknife technique of Efron (1982) and are typically more accurate than the White standard errors (McKinsie and White 1985) in samples smaller than 250 observations (Long and Ervin 2000). Lewis and Linzer (2005) recommend the use of the Efron residuals in their tested scenarios undertaken using a small sample (n=30).

consultant characteristics on the number of investigations per case. Furthermore, it is assumed that consultants share the same *number of investigations* function. This assumption can be used to describe how variables at the consultant level influence the consultant decisions with regards to the number of investigations and thus provide information on the factors which determine the variation in the average number of investigations $\hat{\mu}_j$ across different consultants.

The use of the two-stage model is in this paper evaluated against the use of a one-stage multilevel model[9] which includes group level predictors as explanatory variables. Jusko and Shively (2005) show that in general there are potential efficiency gains from estimating both the bottom and top level parameters in a single stage model. Lewis and Linzer (2005) state that this especially holds when the amount of information available to estimate the bottom level effects in each top level group is small.

Any efficiency gains from running a single stage model would have to be evaluated against the possible loss in the efficiency of parameter estimates within the two stage model. It could well be that the extra effort to fit the more complex single stage hierarchical model would provide less of benefit in comparison to the simple two stage Estimated Dependent Variable (EDV) model. The two–step strategy is seen by Jusko and Shively (2005) as a go between those who are concerned with the fact that the complexities of the individual (patient) cases must be taken care of together with the need to draw broad comparisons across consultants.

[9] $y_{ij} = \alpha + \beta_x x_{ij} + \beta_z z_j + \mu_j + \varepsilon_{ij}$ whereby 'y_{ij}' is the dependant variable, 'x_{ij}' is the explanatory variable at the patient level, 'z_j' is the explanatory variable at the consultant level, 'μ_j' is the residual fixed parameter at the consultant level and 'ϵ_{ij}' is the residual parameter at the patient level.

One can raise questions about the accuracy of the estimates and the associated standard errors given the relatively small sample sizes usually encountered in such studies. Of particular concern to the data described above is what constitutes a sufficient sample size for accurate estimation and evaluation purposes. Issues related to sample size within a multilevel regression model have been given particular attention in the literature[10].

Although the results in the literature are not completely in agreement with each other, they all seem to conclude that in general, the estimates from regression coefficients (OLS) are generally unbiased although less efficient because of the generally larger sampling variance. A large number of groups appears to be more important than a large number of individuals per group. As to the variances, estimates of the lowest level variances are generally accurate but the group level variances are sometimes underestimated. Busing (1993) and Van der Leeden and Busing (1994) suggest a sample size of 100 to achieve accurate group level variances.

Kreft (1996) suggests a rule of thumb, the 30/30 rule, whereby a sample of at least 30 groups with at least 30 individuals per group is recommended. Within the context of a one stage multilevel model this advice seems to be sound if the interest is in the fixed parameters. If the interest is in the random part, i.e. the variance and covariance components and their standard errors, the number of groups should be considerable higher. This advice is also repeated in other more recent work by (Bell et al., 2010)

[10] Mass and Hox (2004, 2005), Snijders (2005), Bell et al., (2010), Hox (2010), Snijders and Bosker (2012).

which bases such conclusion on the work of Hox (1998), Maas and Hox (2002 and 2004). Hox (1998) also recommends that at least 20 observations at level 1 and 50 groups at level 2 are necessary when examining interactions across levels. Similar conclusions are given by Maas and Hox (2004) whereby if one is only interested in the fixed effects of the model, then even ten groups can lead to good estimates.

The question of what constitutes a sufficient sample size for accurate estimation will remain. Increasing the size of the number of groups is usually difficult in practice due to cost related issues and due to the fact that for some studies this is impossible given that all the groups possible are already incorporated within the database. The size of the sample within this study is constrained by the fact that this study, which focuses on the particular PTCA activity undertaken within the hospital already includes all the existing consultants within the hospital. Given that the objective of the paper is to study the behaviour of consultants in Malta, the size of the hospital automatically leads to a restriction on the possible size of the sample. Although one would ideally adhere to these sample size guidelines, the nature of the research in question makes these sample size suggestions difficult to achieve. Olsen and Street (2008) show that the comparison of organisations, when the number of organisations is small, is possible and better inference could be made if patient level data is used. There is indeed an element of compromise which needs to be achieved between the accuracy of the estimates and the possible use of such estimates for policy recommendation purposes.

4.5 Data and variables

The data used in this study come from the Clinical Performance Unit (CPU) within the principal hospital of Malta. The database used is an administratively provided data set of all the PTCA activity undertaken at the public hospital. The unit of analysis is the patient and the analysis is performed for patients who entered hospital and underwent a PTCA procedure between 2009 and 2011. The following ICD-9-CM procedure codes were considered in this study: 0066, 3601, 3602, 3604, 3606, 3609, 0045, 0046, 0047 and 8856. Table 4.1 provides the main descriptive statistics for the variables under consideration.

Table 4.1: Definitions and sample statistics of the main variables

Variable	Definition	Mean	Std.dev.
inv	Number of investigations for patient *i*	35.3	32.1
invf2d	Investigation in first 2 days at the hospital	23.1	14.1
inva2d	Investigation after first 2 days at the hospital	16.8	26.7
countconst	Consultant total number of cases	299	176
countc	A cumulative count of cases under the responsibility of a consultant up until date of patient admission	150.4	133.7
vol90c	Consultant cumulative cases over the 90 days prior to the patient admission date	26.1	17.4
age	Age of patient	62.7	10.5
diagtotal	Number of diagnoses per case	1.12	0.94
CharlsonI	Charlson Comorbidity Index		
public-only	=1 for exclusive public sector practice contracts =0 for public and private practice contracts or *dual practice*		
cvol	= 1 if countconst > 100 cases (*high* volume) = 0 if countconst < 100 cases (*low* volume)		

Source: Based on data drawn from the provided hospital episode database.

The dataset identifies the consultant for each PTCA patient and contains a total of 1,626 hospital visits. The dataset provides information at patient level for *age,* together with a number of consultant characteristics used to identify the main covariates of the study. Data are also available at a patient level for the number of diagnoses per case, *diagtotal,* the number of investigations[11] carried out at a patient level, *inv,* and the date when each of the investigations was undertaken. This gives the ability to distinguish between those investigations undertaken in the first two days of hospital admission, *invf2d,* and those investigations carried out during the rest of the patients' hospital stay, *inva2d*.

These patients are clustered into 35 different consultants, four of which have a *public-only* type of contract compared to the rest who can *dual practice*. Consultants choose between two work plan options, *public-only,* whereby consultants cannot undertake any private practice and *dual practice* whereby consultants can practice in both the private and the public sector. Data on the job plan contract type of all the consultants undertaking PTCA activity is provided for this study.

All consultants carrying out the PTCA treatment at the public hospital are included in the dataset and thus the complete population of the PTCA activity undertaken at the hospital is considered. The analysis of the number of investigations ordered and carried out at the hospital on a patient level shows that there is considerable variation among patients in the number of investigations ordered by consultant with the average being 35 but with a large standard deviation of 32.

[11] Table 4.13 in Appendix B of this chapter identifies the main investigations/tests carried out for PTCA patients. Data on the resources allocated to each of the investigations was not made available for this study and thus no particular weighting structure was applied to the different investigations carried out for the patients undertaking a PTCA.

Based on advice provided by hospital practitioners, the investigations that take place within the first two days of a patient's hospital stay are considered as *absolutely necessary* and variation between consultants in the ordering of such investigations is not expected to be large. On the other hand, investigations carried out during the rest of the patient's hospital visit, though important for the success of the procedure, might reflect an element of consultant practice behaviour. The average number of investigations within the first two days of hospital stay is of 23 with a standard deviation of 14.1. The average number of investigations within the post 2 day hospital stay is of 17 with a standard deviation of 26.7. The variation noted in the number of ordered investigations within the post 2 day period is relatively higher.

The number of investigations ordered by the consultant is also expected to be dependent on volume levels and a number of volume measures were constructed to test the robustness of this relationship. The four volume measures applied in the study are defined in Table 4.1. An analysis of the available data shows that there are no consultants on *public-only* contracts with *high* volume levels, *cvol=1*, whilst consultants on *dual practice* contracts fall both within the *high* and the *low* volume categories. There are 4 consultants with *high* volume levels. The average age of patients undertaking this procedure is around 63 years and the majority of patients treated are males (77% males and 23% females). The ALOS of patients admitted for this procedure is 4.5 days. Most of the cases treated have just one diagnosis which reflects the fact that patients treated are clear identifiable cases which are in need of a PTCA intervention. The Charlson Comorbidity Index, referred to as *CharlsonI* is used as a measure of severity of patient illness. The concentration of volume within a few number of consultants is expected given the 'small' nature of the Maltese hospital setting and the particular procedure under study. The procedure under study (PTCA) is a relatively complex intervention which requires a considerable degree of skill by the doctor.

The consultant responsible for the patient has direct control over the number of investigations performed for the admitted patient and differences in this variable are an indication of the behaviour of the consultant on how to best treat the patient. Physicians are not constrained by the hospital in their treatment recommendations and will consistently apply their 'rules' of practice when treating similar patients (Weinstein 2004). Information on the variation in individual physician behaviour is thus approximated by using the variation in the number of investigations ordered by the consultants. The number of investigations ordered is thus one of the few channels available for the consultant whereby his treatment style could be imposed on the overall process of treatment undertaken by the patient. This could help explain some element of practice variation once the differences in patient characteristics are controlled for.

Patients admitted for the PTCA procedure remain under the care of the same consultant throughout their entire hospital visit which ensures that treatment can be attributed to just one consultant. It is important to note that all investigations considered in this study are undertaken at the public hospital. It has not been possible to differentiate between investigations undertaken in the public sector but which could have instead been carried out in the private sector once the patient is discharged from hospital.

Table 4.2 shows data for the mean number of diagnosis per patient, the Charlson Comorbidity Index, the total mean number of investigations performed and the investigations performed in the pre- and post- 2 day hospital stay period. The mean LOS for the different category type of patients is also provided. Table 4.2 also shows variations for patients who are under the care of consultants with varying job contracts and volume characteristics.

Table 4.2: Descriptive statistics for variables of interest by consultant job plan

Row		Mean no of diagnosis	Charlson Index	Mean no of investigations (total stay)	Mean LOS	invf2d	inva2d
1	All patients	1.12	0.407	35.3	4.5	23.07	16.82
2	*public-only* & low volume*	1.21	0.402	32.8	5.78	22.01	14.08
3	*dual practice*-All	1.11	0.407	35.4	4.43	23.14	17.02
4	*dual practice* & high volume	1.07	0.408	36.2	3.86	24.33	16.32
5	*dual practice* & low volume	1.34	0.404	32.7	7.34	19.0	19.59
6	low volume-All	1.31	0.403	32.8	6.98	19.73	18.18

Source: Analysis of the hospital episode database.
*All consultants falling under the *public-only* category have *low* volume levels.

Patients being seen by consultants on *public-only* contracts (row 2) compared to those on *dual practice* contracts (row 3) tend on average, to have a higher number of diagnoses and a higher overall average LOS. The overall number of investigations ordered for such patients tends on average, to be lower. It is particularly interesting to note that whilst there is minimal difference in the number of investigations performed within the first two days of admittance (22.01 compared to 23.14) between patients seen by consultants on different job plans, this difference widens when the number of investigations performed at the hospital in the period post the first two days of admittance is considered (14.08 compared to 17.02).

Consultants on *dual practice* contracts who have a *high* volume of patients, as defined by *cvol* (row 4) tend to order more investigations on average and their patients spend fewer days in hospital compared to patients seen by *low* volume consultants also on *dual practice* contracts (row 5). Furthermore, patients under the care of consultants on *dual practice* contracts and with *high* volume (row 4) have, on average, the highest mean number of investigations for all categories. The consultant contract type tends to influence the time period when investigations are performed (rows 5 against row 2), where more investigations, on average, take place in the first two days following hospital admittance for patients under the care of *public-only* contract consultants. Patients under the care of consultants with *low* patient volume levels but who have the possibility of undertaking private practice, have more investigations performed towards the end of their hospital stay.

Figure 4-1 shows the variation in the number of total investigations carried out during the full hospital stay for patients grouped by consultant and also the variation by job contract type. Figures 4-2 and 4-3 separate the investigations into those carried out prior and post the 2 day hospital stay mark, showing also the variation by job contract type in each case.

Figure 4-1: Variation in number of total investigations by consultant (left) and by consultant job plan contract type (right)

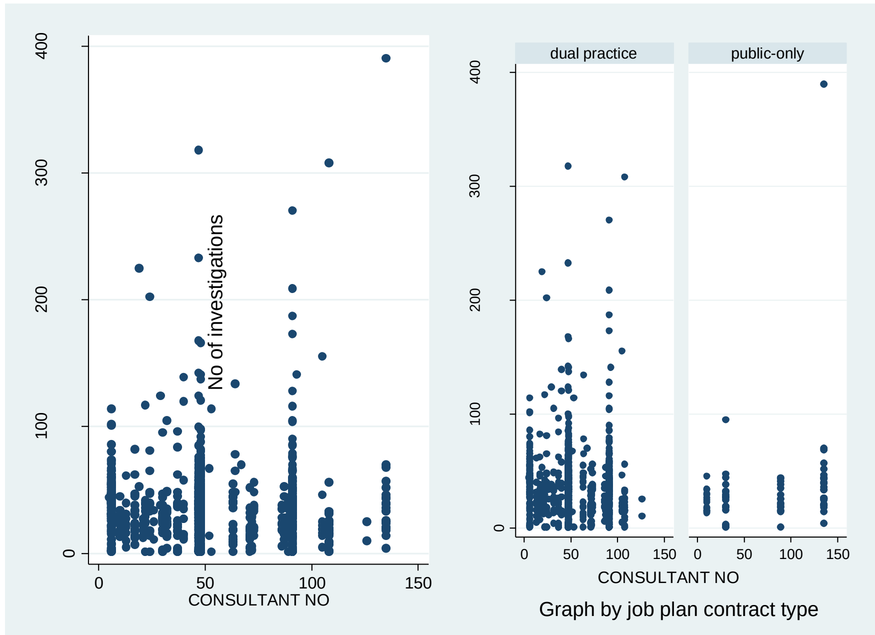


Figure 4-2: Variation in number of investigations for all consultants (left) and by consultant job plan type (right) for the first 2 days of hospital stay

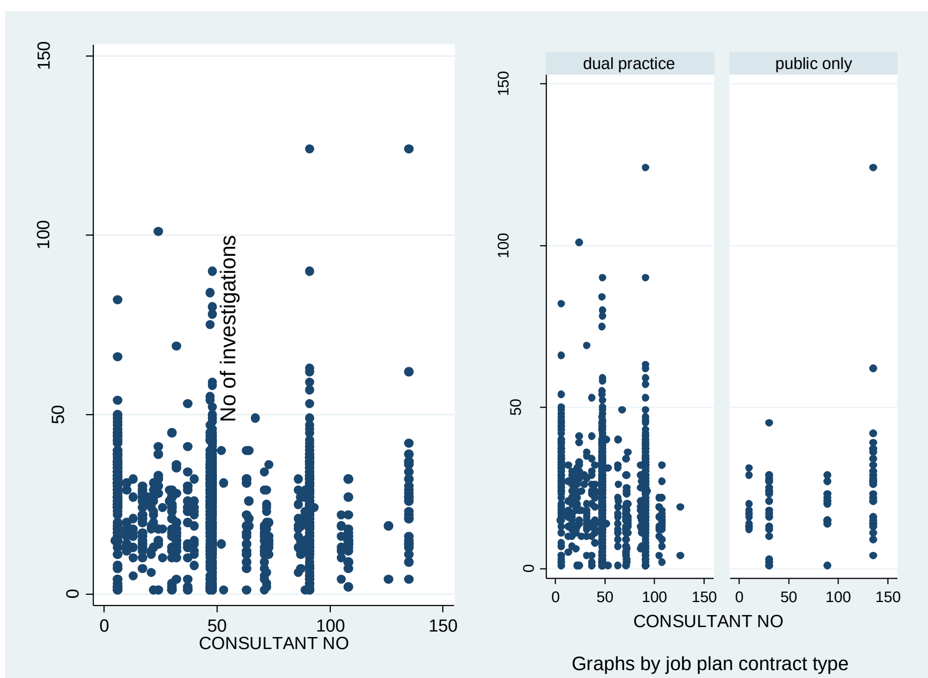
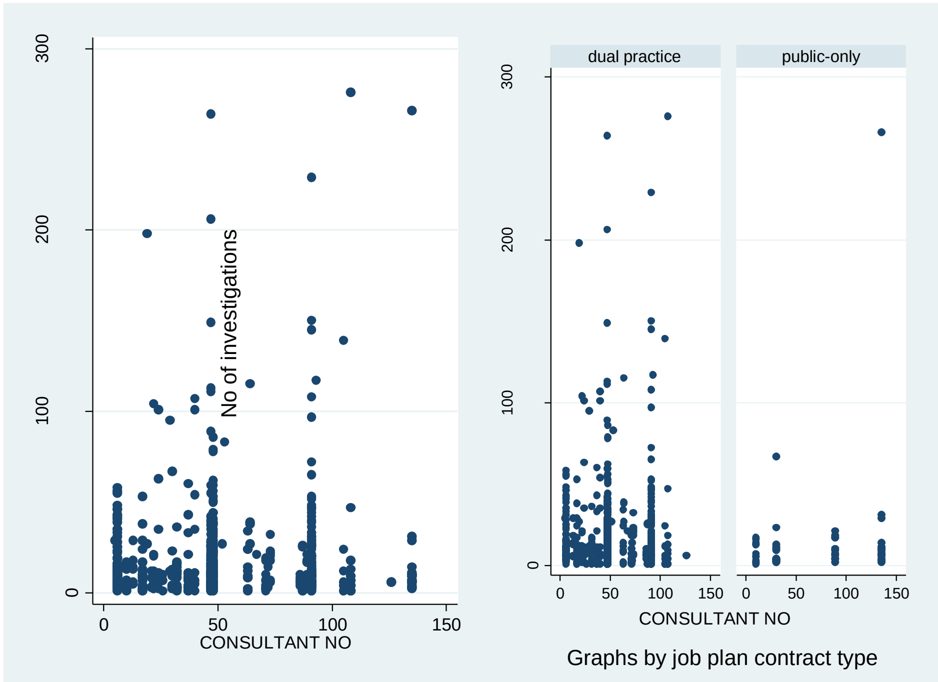


Figure 4-3: Variation in number of investigations for all by consultants (left) and by consultant job plan type (right) after the first 2 days of hospital stay



Variation is present in all the representations although as expected the variation in the number of investigations per patient in the post 2 day hospital admission period is greater. Indeed, part of this variation might be explained by the behavioural characteristics of the consultants themselves particularly the job contract type of the consultant. There is an indication of a greater degree of variation in the number of investigations for patients seen by consultants on *dual practice* contracts.

4.6 Results and Discussion

The initial set of results in this section present the 1st stage regression used to obtain estimates of the consultant fixed effects. The 2nd stage regression results are then estimated; first considering all the investigations carried out during the patient stay at the hospital and then re-estimated to check the robustness of the results when different investigation period dates are considered. The compounding effect of the consultant job plan on the relationship between volume and practice variation is estimated considering all investigations that took place over the full hospital stay period and then re-estimated to incorporate the different dates when the investigations were performed.[12] The resulting marginal effects on practice variation arising from changes in volume for different job plan contract conditions are also presented.

4.6.1 First stage regression

Table 4.3 shows the 1st stage regression results for three different investigation period dates: all cases taken together, *inv;* those performed within the first two days, *invf2d;* and those performed after the first 2 days, *inva2d*.

[12] The results of the estimated one-stage multilevel model used to check for the robustness of the results obtained from the two-stage estimation method are included in Appendix C of this chapter.

Table 4.3: First stage regression results

	(1)	(2)	(3)
Dependent variable	inv	invf2d	inva2d
age	0.186**	-0.099***	0.385***
CharlsonI	17.870***	10.388***	4.483**
Constant	13.798**	23.452***	-10.088
Observations	1,162	1,160	864
R-squared	0.105	0.199	0.032
sigma_u	29.64	6.839	27.96
sigma_e	29.90	12.36	25.57
rho	0.496	0.234	0.544

*** p<0.01, ** p<0.05, * p<0.1.

Table 4.3 shows that age and the measure of severity are statistically significant in explaining the number of investigations for PTCA patients. Column (2) shows that age is negatively related to having investigations performed within the first 2 days and this is possibly related to the fact that differences could exist in the number of investigations ordered due to age concerns. The variable *rho,* is of particular interest as it gives a measure of the variation of $\hat{\mu}_j$, or the extent of variation from the average for different consultants[13].

From column (1) it can be noted that around 50% of the variation in the number of investigations is due to consultant level variation rather than the characteristics of patients. When investigations during the first two days of admission are considered, column (2), the variation shown by the difference in consultants is around 25% compared to around 54% for the investigations performed after the first two days, column (3). There is a greater proportion of variation in the number of investigations

[13] where $\hat{\mu}_j$ represents the consultant fixed effects estimated from the first stage regression.

occurring at the consultant level in the post 2 day hospital stay period. The investigations ordered and carried out by the consultant during the first two days of hospital stay are as expected very much dependent on the patient characteristics. To this effect a higher R^2 value is reported in column (2) compared to column (3), and this is substantiated by the lower 'rho' value representing variation due to consultant factors reported in column (2). It is in this context that investigations carried out during the first two days of patient hospital stay are referred to as 'absolutely necessary' so as to indicate that these tests are absolutely necessary in terms of the particular treatment needs of the patient. Absolutely necessary tests are tests which are considered to be indispensable for the overall treatment of the patient and such test would be expected to be carried out in the first days of the patient hospital stay.

If the consultant had to order tests which might also be important for the treatment of the patient but which are not 'absolutely necessary' then the hypothesis here is that these are ordered in the post 2 day period. In the post 2 day period the role of consultant decisions in the choice of investigations ordered and carried out is expected to be higher. One would expect more flexibility in the choice of tests/investigations for consultants in the post 2 day hospital stay period.

4.6.2 Second stage regression – all investigations

In the 2nd stage regression the influence of particular consultant characteristics on the variance in $\hat{\mu}_j$ was analyzed. The estimate of the fixed effect component resulting from the 1st stage regression equation was retained and used as an input for the 2nd stage regression results, as presented in Table 4.4. A number of volume indicators, *cvol, countconst, countc and vol90c,* were applied to test the robustness of the volume measure in explaining consultant practice variation.

Table 4.4: Second stage regression results – all investigations

	(1)	(2)	(3)	(4)
Dependent variable	$\hat{\mu}_j$	$\hat{\mu}_j$	$\hat{\mu}_j$	$\hat{\mu}_j$
public-only	-0.5016	-1.6432	-0.9989	-1.1670
cvol	-0.1342			
countconst		-0.0049**		
countc			-0.0046**	
vol90c				-0.0385**
Constant	0.1313	1.3879*	0.6830	0.9653
Observations	1,162	1,162	1,162	1,162
R-squared	0.0002	0.0113	0.0055	0.0070

*** p<0.01, ** p<0.05, * p<0.1,
The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.

The results presented in Table 4.4 show that the job contract type of the consultant, is not statistically significant in explaining the variance in $\hat{\mu}_j$ when investigations over the entire hospital stay are considered. This indicates that the variation arising from consultant practices as identified in the 1st stage regression is not due to the differences in the job contract type of the consultant. However, some evidence is found of lower practice variation when consultant volume levels increase.[14]

4.6.3 Second stage regression – with dated investigation categories

Table 4.5 shows variation in consultant practice when the number of investigations carried out on patients are divided into those which are performed in the first two days

[14]A statistically insignificant coefficient is obtained when the *cvol* indicator is used as a measure of volume.

of hospital stay (columns 1-4) and those performed in the post 2 day hospital stay period (columns 5-8). It can be seen that there is a statistically significant effect of the consultant contract type on the variation in $\hat{\mu}_j$ in column (1). However, the consultant job plan contract type variable is statistically insignificant when other volume indicators are taken into account, suggesting that the job plan type *of the* consultant does not explain practice variation in the first two days of hospital stay. Whilst recognizing that the consultant job plan type variable is not significant in explaining variation in the investigations carried out during the first two days of hospital stay, a significant and positive relationship is found between volume levels of the consultants and practice variation.

Table 4.5: Second stage regression results – investigations (<2 and >2 days)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	<2 days	<2 days	<2 days	<2 days	>2 days	>2 days	>2 days	>2 days
Dependent variable	$\hat{\mu}_j$	$\hat{\mu}_j$	$\hat{\mu}_j$	$\hat{\mu}_j$	$\hat{\mu}_j$	$\hat{\mu}_j$	$\hat{\mu}_j$	$\hat{\mu}_j$
public-only	2.109***	0.382	0.1620	0.2913	-5.029***	-4.199***	-3.293***	-3.515***
cvol	2.75***				-3.362**			
countconst		0.0017***				-0.007***		
countc			0.0015***				-0.0069***	
vol90c				0.016***				-0.057**
Constant	-2.147***	-0.4601*	-0.2186	-0.402**	2.812**	2.154**	1.1604*	1.568*
Observations	1,160	1,160	1,160	1,160	864	864	864	864
R-squared	0.187	0.012	0.0057	0.0118	0.03	0.025	0.016	0.018

*** p<0.01, ** p<0.05, * p<0.1.
The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.

A negative and statistically significant coefficient on the job plan contract type variable is obtained in columns (5) to (8) of Table 4.5. Consultants on *public-only contracts* have lower levels of practice variation. The contract type of the consultant explains a proportion of the variation observed in the number of investigations performed on the patient. This supports the preliminary evidence found in the data whereby patients under the care of consultants on exclusive public sector contracts have fewer investigations in the post 2 days of their hospital stay. The fact that patients are being seen by consultants on *dual practice* contracts serves to positively explain part of the variation in practice patterns for the post 2 day hospital stay period.

A negative coefficient is observed for the different volume indicators. This suggests that lower consultant volume levels lead to higher practice variation when dealing with investigations carried out in the post 2 day hospital stay period. This is in contrast to the positive relationship observed between volume and practice variation in the first two days of hospital stay and could be explained by the fact that consultants with lower volume levels might decide to use such lower overall activity levels to increase the number of investigations ordered in the post 2 day hospital stay period. Both the consultant job plan type variable and each of the volume variables are important to explain practice variation.

4.6.4 Second stage regression – all investigations with interaction term and marginal effects

Table 4.6 shows the estimation results when the interaction term between the volume variable and the consultant job plan type variable is applied to all the investigations carried out on the patient during the full hospital stay. The average marginal effects arising from these results are provided in Table 4.7.

The estimation result identifies the compounding effect of the job plan type variable on the relationship between volume and practice variation. A number of different volume indicators were tested to assess the robustness of the volume indicator. An interaction term for *cvol* was not included given that there were no consultants on *public-only* contracts who performed more than 100 cases over the three year period.

Table 4.6: Second stage regression results – all investigations – with interactions

	(1)	(2)	(3)
Dependent variable	$\hat{\mu}_j$	$\hat{\mu}_j$	$\hat{\mu}_j$
public-only	-22.8746***	-4.9670***	-5.3864***
countconst	-0.0050**		
public-only#countconst	0.8987***		
countc		-0.0047**	
public-only#countc		0.3330***	
vol90c			-0.0450**
public-only#vol90c			0.9323***
Constant	1.4212*	0.6948	1.1255
Observations	1,162	1,162	1,162
R-squared	0.0616	0.0125	0.0349

*** p<0.01, ** p<0.05, * p<0.1.
The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.

The results in columns (1) to (3) show a negative value for the *public-only* coefficient which reflects a lower level of practice variation for those patients being seen by consultants on *public-only* contracts compared to consultants on *dual practice* contracts when volume is set to zero. The three measures of volume are negative and significant. This represents the slope coefficient for patients being seen by consultants on *dual practice* contracts. The value of the *public-only* by volume interaction term is positive and significant in all cases representing the difference in the responsiveness of practice variation to volume changes when consultants on *public-only* contracts are considered, compared to those consultants who are on *dual practice* contracts.

Table 4.7: Marginal effects (all investigations)

Volume measure	Type of Contract	(1) dy/dx	(2) dy/dx	(3) dy/dx
countconst	*dual practice*	-0.005**		
	public-only	0.894***		
countc	*dual practice*		-0.005**	
	public-only		0.328***	
vol90c	*dual practice*			-0.045**
	public-only			0.887***
Observations		1,162	1,162	1,162

*** p<0.01, ** p<0.05, * p<0.1.

The average marginal effects provided in Table 4.7 confirm that there is a significant difference in the slope of the volume indicators for different job plan contracts. A fall in volume leads to an increase (though minimal) in consultant practice variation for those on *dual practice* contracts. This relationship is positive (and stronger) for those on *public-only* contracts implying that increases in volume levels for consultants on *public-only* contracts will have a greater impact on practice variation[15]. The results are consistent for all three measures of volume.

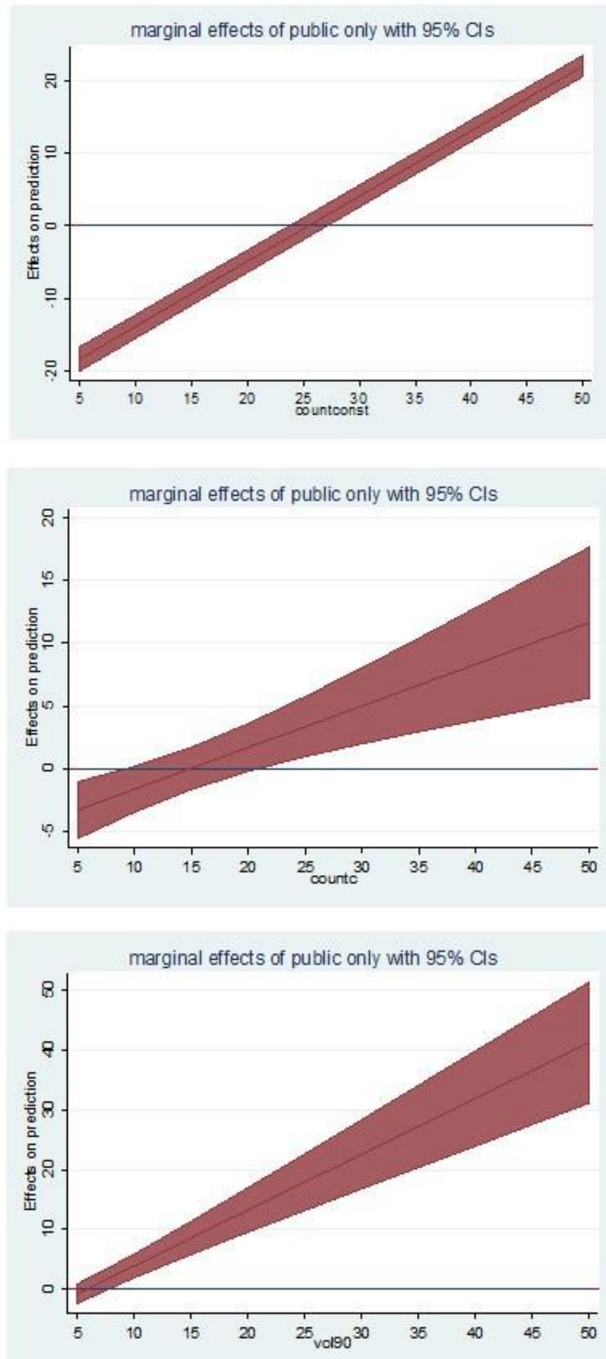
Furthermore, there is a significant difference in practice variation levels for patients seen by consultants on different contract types. Figure 4-4 represents the difference in practice variation levels for patients being seen by consultants on different contract types as volume[16] levels change. Different measures of volume: *countconst, countc* and *vol90c* are represented in each of the plots. The plots of the confidence interval for the difference between the two groups is also shown and whenever the 95% confidence

[15] The value of 0.894 in column 1 of Table 4.7 shows that when the volume for consultants on *public-only* contracts (represented by 'countconst') changes by one unit (and all other variables are kept fixed) there is an increase in practice variation of 0.894.

[16] The chosen value for the volume indicators represents the range of possible volume levels in the dataset with values taken between 5 and 50 cases, using multiples of 5 cases.

interval for the difference does not include zero, then the difference can be considered to be statistically significant.

Figure 4-4: Differences in practice variation levels (*inv*) between *public-only* and *dual practice* contracts for different volume indicators at different volume values



Data for the discrete change from the base level for each of the volume indicators is included in the Appendix[17] to this chapter. At very low levels of volume there is a negative difference between investigations ordered by *public-only* and *dual practice* consultants. The number of investigations ordered by *public-only* consultants are less than those ordered by *dual practice* consultants at low volume levels. As levels of volume for consultants on *public-only* contracts rise investigations increase by more than investigations ordered by *dual practice* consultants and thus the difference becomes positive. A possible explanation for this could be the fact that as volume increases the possibility of ordering investigations by *dual practice* consultants is reduced as *public-only* consultants increase their number of investigations ordered. The differences in practice levels between *public-only* and *dual practice* contract consultants increased as the value of volume increased. This difference was found to be statistically significant[18] for most of the different values of volume and for each of the volume indicators applied.

4.6.5 Second stage regression - with dated investigation categories – with interactions and marginal effects

Table 4.8 shows the results for the compounding effect of the consultant job plan type on the relationship between volume and practice variation for different dated investigation periods. The results, based on investigations undertaken in the first 2

[17] Refer to Table 4.10.

[18] The differences in effect on practice variation between consultants on different contract conditions are illustrated for the different volume levels with confidence interval bands to indicate the statistical significance of such differences.

days following admittance to hospital, are presented in columns (1) to (3), whilst columns (4) to (6) show the results for investigations occurring after the first 2 days of hospital stay. A number of different volume indicators were applied to test the robustness of the volume indicator. An interaction term on the *cvol* variable was not applied.[19]

Table 4.8: 2nd stage regression results – investigations with interactions (<2 and >2 days)

	(1) <2 days	(2) <2 days	(3) <2 days	(4) >2 days	(5) >2 days	(6) >2days
Dependent variable	$\hat{\mu}_j$	$\hat{\mu}_j$	$\hat{\mu}_j$	$\hat{\mu}_j$	$\hat{\mu}_j$	$\hat{\mu}_j$
public-only	-9.045***	-1.6291**	-1.346***	-21.139***	-6.324***	-7.161***
countconst	0.002***			-0.0072***		
public-only#countconst	0.399***			0.7062***		
countc		0.0015**			-0.0069***	
public-only#countc		0.1503***			0.2499***	
vol90c			0.014**			-0.0638***
public-only#vol90c			0.362***			0.7489***
Constant	-0.4453*	-0.2132	-0.3398*	2.1835**	1.1715*	1.7246*
Observations	1,160	1,160	1,160	864	864	864
R-squared	0.1047	0.0190	0.0511	0.0548	0.0196	0.0374

*** p<0.01, ** p<0.05, * p<0.1.

The baseline, *public-only=0,* indicates consultant is on a *dual practice* contract.

A positive and significant volume coefficient for the interaction term is obtained in all specifications, indicating a positive difference in the responsiveness of practice variation when patients are under the care of *public-only* contract consultants in comparison to those on *dual practice* contracts. The responsiveness of practice variation to changes in volume for patients seen by consultants on *dual practice* contracts varies between the two periods of investigation; it is positive and significant for investigations in the first two days of hospital stay but negative for investigations in the remaining hospital stay period. We found that as volume levels of consultants on

[19] There are no consultants on exclusive public contracts who perform more than 100 cases and thus the *cvol* volume variable is not used as an interaction term.

dual practice contracts decreased, higher practice variation was observed in the post 2 day hospital stay period.

The average marginal effects provided in Table 4.9 show a significant difference in the slope of volume indicators for different consultant job plan contracts. Results are presented for the different investigation periods. Columns (1) to (3) show that there is an increase in practice variation as volume increases for consultants on *dual practice* contracts. A similar relationship (although stronger) is obtained for those on *public-only* contracts with increases in practice variation expected as volume increases. The results are consistent for all three measures of volume used in the analysis.

Table 4.9: Marginal effects – investigations (<2 and >2 days)

Volume measure	Type of contract	(1) dy/dx	(2) dy/dx	(3) dy/dx	(4) dy/dx	(5) dy/dx	(6) dy/dx
		<2 days	<2 days	<2 days	>2 days	>2 days	>2 days
countconst	*dual practice*	0.002***			-0.007***		
	public-only	0.401***			0.699***		
countc	*dual practice*		0.002**			-0.007***	
	public-only		0.152***			0.243***	
vol90c	*dual practice*			0.014**			-0.064***
	public-only			0.376***			0.685***
Observations		1,160	1,160	1,160	864	864	864

*** p<0.01, ** p<0.05, * p<0.1

The specifications in columns (4) to (6) show that practice variation increases with a fall in volume for patients seen by consultants on *dual practice* contracts. For consultants on *public-only* contracts practice variation is expected to increase as volume increases. The results are consistent for all three measures of volume used in the analysis. The coefficients obtained for consultants on *public-only* contracts are greater in magnitude when compared to those obtained for *dual practice* consultants.

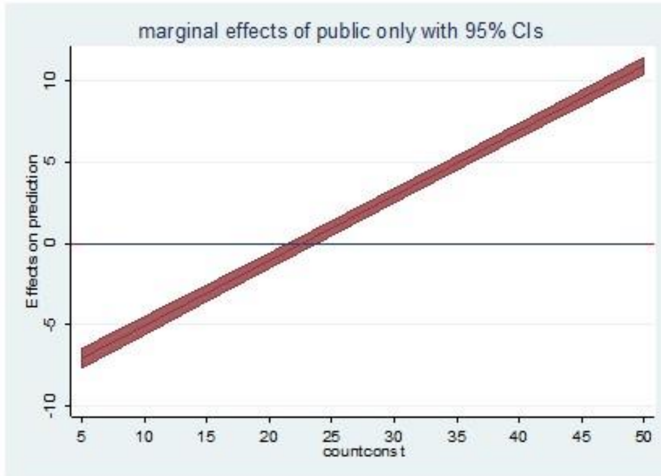
There is a difference in the impact on practice variation when volume changes for consultants on *dual practice* contracts if investigations over the two different periods are compared.

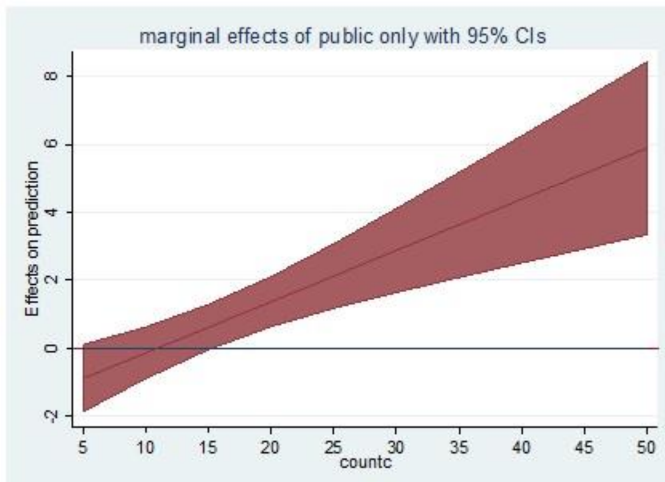
The results in Table 4.8 also show that for both investigation periods, a negative and significant coefficient value for the consultant job plan variable was obtained. This indicates a lower constant level of practice variation for patients seen by consultants on *public-only* contracts compared to consultants on *dual practice* contracts when the volume is set to zero. A greater negative level effect of *public-only* contracts on practice variation was found for the post 2 day hospital stay period.

The difference in practice variation levels between consultants with different job plan contracts varied for different values of volume. The lines in Figures 4-5 and 4-6 represent the difference in practice variation (both for *invf2d* and *inva2d*) for patients being seen by consultants on different job contract types as volume[20] levels change. Different measures of volume: *countconst, countc* and *vol90c* are represented in each of the figures. The plots of the confidence interval for the difference between the two groups of consultant job contract types is also shown and whenever the 95% confidence interval for the difference does not include zero, then the difference can be considered to be statistically significant. Figures 4-5 and 4-6 indicate that there is indeed difference in practice variation (both in *invf2d* and *inva2d*) between patients under the care of consultants on different job plan contracts as volume levels change.

[20] The chosen value for the volume indicators represents the range of possible volume levels in the dataset with values taken between 5 and 50 cases, using multiples of 5 cases.

Figure 4-5: Differences in practice variation levels (*invf2d*) between *public-only* and *dual practice* contracts for different volume indicators at different volume values





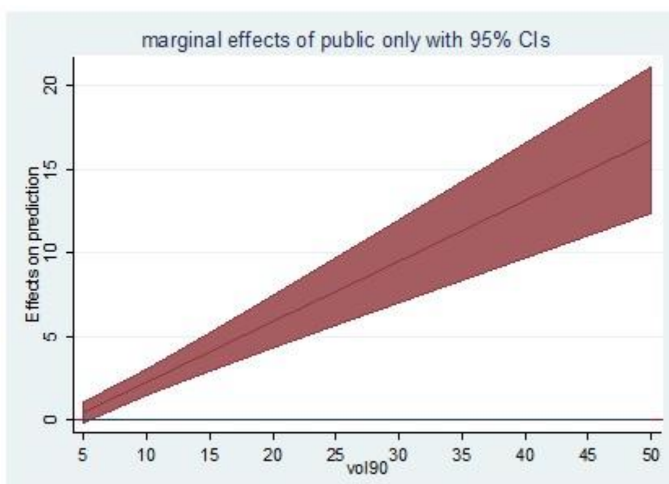
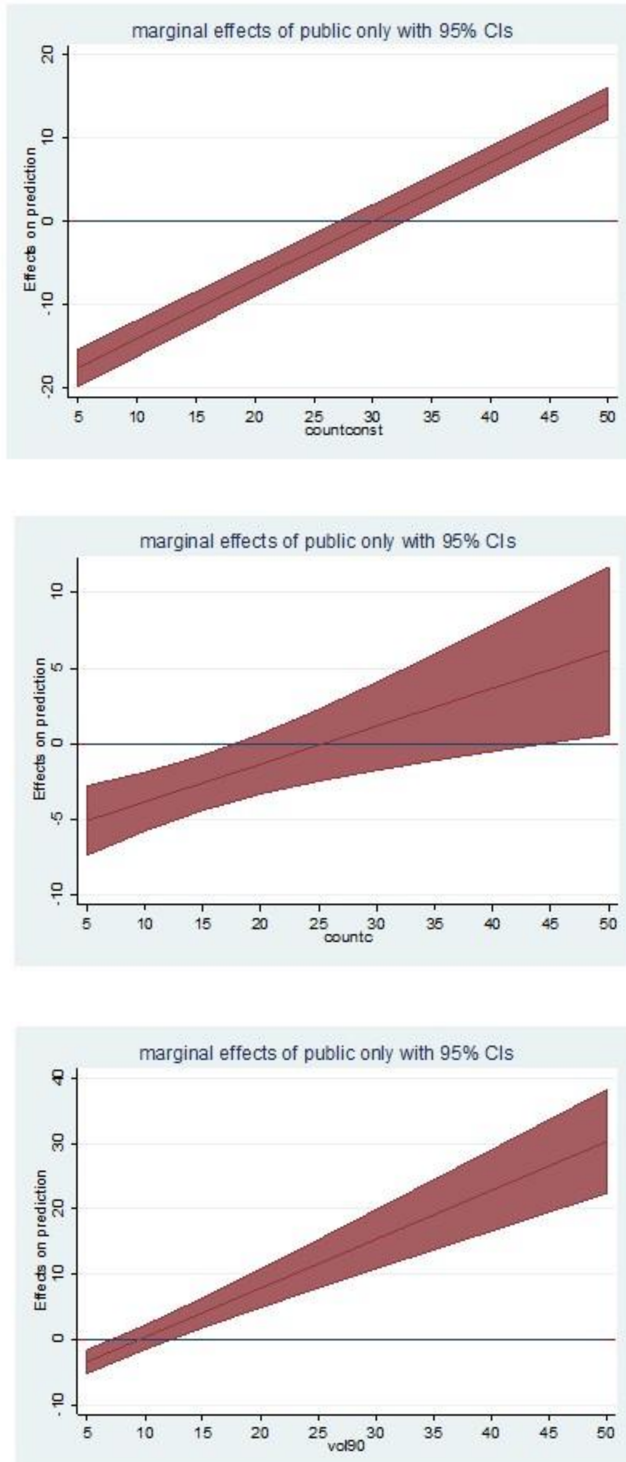


Figure 4-6: Differences in practice variation levels (*inva2d*) between *public-only* and *dual practice* contracts for different volume indicators at different volume values



The results presented in the Appendix[21] show the discrete change in practice variation from the base level considering different volume measures for investigations performed in the pre- and post 2 day hospital stay period. The difference in practice levels for patients seen by consultants on *public-only* contracts compared to *dual practice* contracts was found to be negative for very low levels of volume but increased with higher levels of volume. The differences in practice levels between *public-only* and *dual practice* contract consultants increased as the value of volume increased. This difference is statistically significant[22] for the majority of the different values of volume tested. Some statistically insignificant differences are recorded at very low volume levels.

4.7 Conclusion

This paper studies the behaviour of consultants working under very specific incentives and work practices created within the context of an acute general public hospital, the only acute general hospital on the Islands of Malta. Variations in practice patterns have important consequences for the quality and cost of health care treatment. Differences in practice variation, represented by the number of investigations performed on the patient, have been attributed to individual patient characteristics, consultant characteristics and their integrated effects.

[21] Refer to Table 4.11 and Table 4.12.

[22] The differences in effect on practice variation between consultants on different contract conditions are illustrated for the different volume levels with the confidence interval bands to indicate the statistical significance of such differences.

The theoretical model suggests an empirical framework for investigating the role of the number of investigations in explaining practice variation for the treatment of PTCA. A multilevel modelling approach was used to test this hypothesis linking practice variation in the behaviour of the consultant to the consultant job contract type.

The available data allowed the study of the variation at the individual patient level on a consultant by consultant basis. A number of data limitations have to be highlighted. Data on activity by consultants working in the private sector were not available and this prevented the study of the behaviour of the consultant when such activity levels changed. Furthermore, data is not available on how many of the patients who receive PTCA treatment at the public hospital, seek private sector intervention prior or post to their admittance to the public hospital. The number of investigations ordered during the hospital stay period could depend on patient private sector visits however, this is not controlled for within the regression analysis. Of particular interest to this study would be a measure of how much of the currently undertaken investigations within the public hospital could in fact be carried out within the private sector setting. The shift of activity from the current public hospital setting to the private setting will help to free up resources within the public hospital. No conclusions could thus be drawn on the merits of undertaking any of this activity in the private sector once the patient is discharged from hospital.

The possibility of distinguishing between patients who had visited a private consultant prior to the public hospital visit and those who visit the consultant post discharge would serve as an additional control variable to better understand the possible reasons of practice variation amongst PTCA patients. Moreover the available data allows only for the separation of consultants between those who can and those who cannot practice privately but does not provide information on the amount (hours per week) of private sector activity carried out by the *dual practice* consultants. There could indeed be an element of practice variation which depends on the size of the private sector practice of the consultant.

No particular weighting structure is applied to the different investigations ordered and carried out for the patients undertaking a PTCA within this study. There could indeed be a case in favour of weighting tests/investigations on the basis of some agreed rule. Data on costs associated with each of the tests could be suggested as a possible weighting rule, or the use of an indicator which measures the relative importance of the particular test for the overall treatment of the patient. A limitation of the study is that such information is not available and thus the weighting of tests is left as a point for further research.

In view of the very limited data available on surgeons working within the hospital, surgeon characteristics are not controlled for in this study. It is the consultant who is responsible for the patients who orders such investigations but information on surgeons (who work under the control of the consultant) could also be a source of explanation of practice variation. Further research in this area would incorporate the role of surgeon characteristics, primarily surgeon volume levels and the surgeon job plan contract type to help in the explanation of practice variation patterns.

In this study we have controlled for a range of patient and consultant characteristics which influence practice variation, particularly patient age, severity of illness, consultant volume levels and job contract type. One though recognises that there are a number of other factors which consultants put into the balance when making a choice between a *dual practice* contract and a *public-only* contract with the public hospital. The observed impact on practice variation may thus be due to factors other than the direct variation arising from job contract type. There may potentially be other explanations as to why variation in practice differs between consultants rather than the job contract type on which this study focuses.

It is indeed possible that consultants with different skills and work/leisure preferences have a tendency to opt for a particular type of contract type and thus their practice decisions are also influenced by such factors rather than just the particular chosen job contract type. Consultants might base their choice of job contract type also on what they perceive to be the mostly sought out form of contract type requested and desired by their patients. Patient preferences will thus have an impact on consultant contract type. Data on patient preferences for being treated by dual or *public-only* consultants is not available. Furthermore one could consider data in relation to years of experience at the job (to measure skills) and some form of indicator of work/leisure preferences to help disentangle the confounding impact of such factors on practice variation between consultants. Work in this area is a possible avenue for further research.

Issues related to the sample size of the available data are a source of concern. Indeed, the small sample size used in this work is the result of the particular characteristics of the single hospital and the procedure under study. The results obtained show that despite the limitations arising from the small number of consultants, significant results are found for the estimated coefficients. Indeed, the results derived from the two stage model are robust when compared to estimates derived from the one-stage multilevel approach. A possible suggestion for future work would be to carry out the analyses across a whole department or unit rather than basing the analysis on a single procedure. This would help to increase the sample size, both at the patient and at the consultant level. There is though a drawback related to the fact that differences across departments would be expected to be bigger given the diverse range of patients treated. In addition any possible policy recommendations under such conditions have to be made with caution given the variation in treatments being offered at each department. There is indeed an element of compromise which needs to be achieved between the accuracy of the estimates and the possible use of such estimates for policy recommendation purposes. Policy recommendations are thus to be devised with caution and in recognition of such characteristics.

The results indicate that a proportion of the practice variation occurs at the consultant level, after controlling for a number of individual patient characteristics. Indeed, this variation differs when the data for investigations is categorized into investigations occurring within or after the first 2 days of hospital stay. The results indicate that greater variation occurs at the consultant level when investigations in the post 2 day hospital stay period are considered. This is the period which is most likely to be used by the consultants to put into practice their particular style of treatment.

The results also suggest that when investigations in the post 2 day hospital stay period are analyzed, the job plan contract type of the consultant has an effect on practice variation. Indeed, lower practice variation was observed amongst patients under the care of consultants on *public-only* contracts. Changes in volume have contrasting impacts on practice variation, with a negative impact obtained when dealing with investigations occurring after the 2 day stay period at the hospital. A reduction in consultant volume levels is likely to cause an increase in practice variation. The relationship between consultant volume and practice variation is positive when investigations in the first two days of hospital stay are considered.

There is also a compounding effect of the consultant job plan on the relationship between volume and practice variation. A positive coefficient describes the responsiveness of practice variation to changes in volume when patients are under the care of *public-only* contract consultants in comparison to those on *dual practice* contracts. An increase in volume levels for consultants on *public-only* contracts will lead to a positive increase in practice variation. The responsiveness of practice variation to changes in volume levels for patients seen by consultants on *dual practice* contracts varies between the two periods of investigation. It can be concluded that as volume levels of consultants on *dual practice* contracts decrease, higher practice variation results in the post 2 day hospital stay period.

Furthermore, it was found that there was a difference in practice variation levels between *public-only* and *dual practice* contract consultants. This difference varied for different values of volume. The difference in practice levels for patients being seen by consultants on *public-only* contracts compared to *dual practice* contracts was found to be negative for very low levels of volume but increased and turned positive with higher levels of volume. The differences in practice levels between *public-only* and *dual practice* contract consultants increased as the value of volume increases.

Finally, we found that a greater proportion of variation in practice patterns, particularly in the post 2 day hospital stay period, could be explained by consultant characteristics. Patients seen by consultants who can practice privately were more likely to receive more investigations in their patients' last days of hospital stay compared to other patients. Furthermore, there is the impact of changing volume levels on explaining practice variation with differences arising for different periods of investigation and for different consultant job contract conditions. The results obtained lead to the conclusion that the job plan contract type of the consultant is important in explaining variations in the number of investigations undertaken especially in the post 2 day hospital stay period.

References

Allard, M., Jelovac, I., and Leger, P.T. (2011). 'Treatment and referral decisions under different physician payment mechanisms', *Journal of Health Economics,* vol.30(5), pp. 880-893.

Andersen, R.M. and Newman, J.F. (1973). 'Societal and Individual Determinants of Medical Care Utilization in the United States.', *Milbank Memorial Fund Quarterly,* vol.51, pp. 95-124.

Anderson, T.F. and Mooney, C. (1990). 'The Challenges of Medical Practice Variations'. Macmillan Press.

Banerjee, A. (1997). 'A theory of misgovernance'. *Quarterly Journal of Economics,* vol. 112(4), pp. 1289-1332.

Bell, A.A., Morgan, G.B., Kromrey, J.D. and Ferron, J.M. (2010). 'The impact of small cluster size on multilevel models: a monte carlo examination of two-level models with binary and continuous predictors.' *JSM Proceedings, Section on Survey Research Methods. American Statistical Association*. 1122-1129.

Biglaiser, G. and Ma, C. (2007). 'Moonlighting: Public Service and Private Practice', *The RAND Journal of Economics,* vol.38(4), pp. 1113-1133.

Bir, A. and Eggleston, K. (2003). 'Physician dual practice: access enhancement or demand inducement?'. Tufts University, Department of Economics Working Paper 11.

Blomqvist, A. (1991). 'The doctor as double agent: information asymmetry, health insurance, and medical care', *Journal of Health Economics,* vol.10(4), pp. 411-432.

Blomqvist, A. and Leger, P.T. (2005). 'Information asymmetry, insurance, and the decision to hospitalize', *Journal of Health Economics,* vol.24(4), pp.775-793.

Bloor, K., Maynard, A., and Freemantle, N. (2004). 'Variation in activity rates of consultant surgeons and the influence of reward structures in the English NHS', *Journal of Health Services Research and Policy,* vol.9(2), pp. 76-84.

Brekke, K.R. and Sorgard, L. (2007). 'Public versus private health care in a national health service', *Health Economics,* vol.16(6), pp. 579-601.

Carlsen, F., Grytten, J., and Skau, I. (2003). 'Financial incentives and the supply of laboratory tests', *Europeanournal of Health Economics.,* vol.4(4), pp. 279-285.

Clarke, P., and Wheaton, B. (2007). 'Addressing data sparseness in contextual population research using cluster analysis to create synthetic neighborhoods'. Sociological Methods and Research, vol. 35, pp.311-351.

De Jaegher, K. and Jegers, M. (2000). 'A model of physician behaviour with demand inducement', *Journal of Health Economics,* vol.19(2), pp. 231-258.

De Jong, J.D., Westert, G.P., Lagoe, R.F., and Groenewegen, P.P. (2006). 'Variation in hospital length of stay: do physicians adapt their length of stay decisions to what is usual in the hospital where they work?', *Health Services Research,* vol.41(2), pp. 374-394.

Dranove, D. (1988). 'Demand inducement and the physician/patient relationship', *Economic Inquiry,* vol.26(2), pp. 281-298.

Eisenberg, J.M. (2002). 'Physician utilization: the state of research about physicians' practice patterns', *Medical Care,* vol.40(11), pp. 1016-1035.

Ellis, R.P. and McGuire, T.G. (1986). 'Provider behaviour under prospective reimbursement: Cost sharing and supply', *Journal of Health Economics,* vol.5(2), pp. 129-151.

Ellis, R.P. and McGuire, T.G. (1990). 'Optimal payment systems for health services', *Journal of Health Economics,* vol.90(4), pp. 375-396.

Evans, R. (1974). 'Supply-Induced demand: some empirical evidence and implications', In: Perlman, M. ed. *The economics of health and medical care*, Macmillan, London, pp. 162-173.

Farley, P.J. (1986). 'Theories of the price and quantity of physician services. A synthesis and critique', *Journal of Health Economics,* vol.5(4), pp. 315-333.

Feldman, R. and Sloan, R. (1989). 'Reply from Feldman and Sloan', *Journal of Health Politics, Policy and Law,* vol.14(3), pp. 621-625.

Ferrinho, P., Van Lerberghe, W., Fronteira, I., Hipolito, F. and Biscaia, A. (2004). 'Dual practice in the health care sector: review of evidence', *Human Resources of Health*, vol.2(14).

Fuchs, V.R. (1978). 'The supply of surgeons and the demand for operations', The *Journal of Human Resources .,* vol.13 Suppl, pp. 35-56.

Gal-Or, E. (1999). 'Optimal reimbursement and malpractice sharing rules in health care markets', *Journal Regulatory Economics*, vol.16(3), pp. 237-265.

Gelman,A., (2005). 'Two-Stage Regression and Multilevel Modelling: A commentary', *Political Analysis*, vol.13, pp. 459-461.

Gonzalez, P. (2004). 'Should physicians' dual practice be limited? An incentive approach', *Health Economics,* vol.13(6), pp. 505-524.

Gonzalez, P. (2005). 'On a policy of transferring public patients to private practice', *Health Economics,* vol.14(5), pp. 513-527.

Gruen, R., Anwar,R., Begum, T., Killingsworth, JR., and Normand, C., (2002). 'Dual job holding practitioners in Bangladesh: an exploration', *Social Science & Medicine*, vol.(54), pp.267-79.

Gruber, J. and Owings, M. (1996). 'Physician financial incentives and cesarean section delivery', *RAND Journal of Economics*, vol.27(1), pp. 99-123.

Grytten, J. and Sorensen, R. (2003). 'Practice variation and physician-specific effects', *Journal of Health Economics,* vol.22(3), pp. 403-418.

Harper, J., Hauck, K., and Street, A. (2001). 'Analysis of costs and efficiency in general surgery specialties in the United Kingdom.', *European Journal of Health Economics,* vol.2, pp. 150-157.

Hayward, R.A., Manning, W.C., McMahon, L.F., and Bernard, A.M. (1994). 'Do attending or resident physician practice styles account for variations in hospital resource use?', *Medical Care,* vol.32(8), pp. 788-794.

Hennig-Schmidt, H., Selten, R., and Wiesen, D. (2011). 'How payment systems affect physicians' provision behaviour-an experimental investigation', *Journal of Health Economics,* vol.30(4), pp. 637-646.

Hox, J.J., Multilevel analysis, Techniques and Applications, Routledge (2010), Second edition.

Hvenegaard, A., Street, A., Sorensen, T.H., and Gyrd-Hansen, D. (2009). 'Comparing hospital costs: what is gained by accounting for more than a case-mix index?', *Social Science & Medicine,* vol.69(4), pp. 640-647.

Jacobs, R., Smith, P.C., and Street, A. (2006). 'Measuring efficiency in health care: Analytical techniques and health policy'. Cambridge, Cambridge University Press.

Jusko, L.K. and Shively, W.P. (2005). 'Applying a two-step strategy to the analysis of cross-national public opinion data', *Political Analysis,* vol.13, pp. 327-344.

Kreft, I.G.G. (1996). 'Are multilevel techniques necessary? An overview, including simulation studies'. Unpublished manuscript, California State University at Los Angeles.

Kumbhakar, SC. and Lovell, CAK., (2000). 'Stochastic Frontier Analysis.' Cambridge University Press: Cambridge.

Lang, K. (1994). 'Does the human-capital/educational-sorting debate matter for development policy?', American Economic Review, vol. 84(1), pp. 353-358.

Laudicella, M., Olsen, K.R., and Street, A. (2010). 'Estimating cost variation across hospital departments - a two stage multi-level approach using patient-level data', *Social Science & Medicine,* vol.71, pp. 1872-1881.

Lewis, J.B. and Linzer, D.A. (2005). 'Estimating regression models in which the dependent variable is based on estimates', *Political Analysis,* vol.13, pp. 345-364.

Maas, J.M., Hox, J.J.,(2004). 'Robustness issues in multilevel regression analysis', *Statistica Neerlandica*, vol.58, pp. 127-137.

Maas, J.M., Hox, J.J., (2005) 'Sufficient Sample Sizes for Multilevel Modeling', *Methodology*, vol. 1(3), pp.86–92.

McGuire, T.G. and Pauly, M.V. (1991). 'Physician response to fee changes with multiple payers', *Journal of Health Economics,* vol.10(4), pp. 385-410.

McGuire, T. (2002). 'Physician agency'. In A.J. Culyer, & J.P. Newhouse (Eds.), *Handbook of Health Economics,* vol. 1, pp. 461-536. Elsevier.

Mercuri, M., Natarajan M.K. , Norman, G.F., and Gafni, A. (2012). 'An even smaller area variation: differing practice patterns among interventional cardiologists within a single high volume tertiary cardiac centre', *Health Policy,* vol.104(2), pp. 179-185.

Mitchell, J.B., Hadley, J., and Gaskin, D. (2000). 'Physicians responses to Medical Fee Schedule Reductions', *Medical Care,* vol.38(10), pp. 1029-1039.

Mitchell, J.M. and Sass, T.R. (1995). 'Physician ownership of ancillary services: Indirect demand inducement or quality assurance?', *Journal of Health Economics,* vol.14(3), pp. 263-289.

O'Neill, L. and Kuder, J. (2005). 'Explaining variation in physician practice patterns and their propensities to recommend services', *Medical Care Research and Review,* vol.62(3), pp. 339-357.

Olsen, K.R., and Street, A. (2008). 'The analysis of efficiency among a small number of organisations: how inference can be improved by exploiting patient - level data'. *Health Economics*, vol. 17, pp. 671-681.

Pauly, M. and Redisch, M. (1973). 'The Not-For-Profit Hospital as a Physicians' Cooperative', *American Economic Review,* vol.63(1), pp. 87-99.

Pauly, M.V. (1980). 'Doctors and their Workshops: economic models of physician behaviour'. National Bureau of Economic Research (NBER), University of Chicago Press.

Pauly, M.V. and Satterthwaite, M.A. (1981). 'The pricing of primary care physicians' services: A test of the role of consumer infomation', *Bell Journal of Economics,* vol.12(2), pp. 485-506.

Rice, N. and Jones, A. (1997). 'Multilevel models and health economics', *Health Economics,* vol.6(6), pp. 561-575.

Rice, N. and Leyland, A. (1996). 'Multilevel models: applications to health data', *Journal Health Services Research Policy Law,* vol.1(3), pp. 154-164.

Rice, T. and Labelle, R.J. (1989). 'Do physicians induce demand for medical services?', *Journal of Health Politics Policy,* vol.14, pp. 587-600.

Rickman, N. and McGuire, A. (1999). 'Regulating providers' reimbursement in a mixed market for health care', *Scottish Journal of Political Economy,* vol.46(1), pp. 53-71.

Ro, K.K. (1969). 'Patient characteristics, hospital characteristics and hospital use', *Medical Care,* vol.7(4), pp. 295-312.

Robinson, W.S. (1950). 'Ecological correlations and the behaviour of individuals.' *American Social Review*, vol.15, pp. 351-357.

Selvin, H. (1958). 'Durkheim's suicide and problems of empirical research'. *American Journal of Sociology*, vol.63, pp. 601-619.

Schmidt, P. and Sickles, R.C. (1984). 'Production frontiers and panel data'. *Journal of Business and Economic Studies*, vol. 2. pp. 299-326.

Shwartz, M., Ash, A.S., Anderson, J., Iezzoni, I.I., Payne, S.M.C. and Restruccia, J.D. (1994). 'Small area variations in hospitalizations rates: how much you see depends on how much you look'. Medical Care, vol. 32, pp. 189-201.

Snijders, A.B. and Bosker, R.J. (2012). Multilevel analysis: an introduction to Basic and Advanced multilevel modeling, second edition, London Sage Publishers.

Snijders, A.B. (2005). 'Power and Sample size in multilevel and linear models'. *Encyclopedia of statistics in behavioral science*, vol. (3) pp. 1570-1573.

Socha, K.Z. (2010). 'Physician dual practice and the public health care provision'. *Health Economics Papers*, No4, University of Southern Denmark.

Socha, K.Z. and Bech, M. (2011). 'Physician dual practice: a review of literature', *Health Policy,* vol.102(1), pp. 1-7.

Van der Leeden, R., & Busing, F. (1994). 'First iteration versus IGLS RIGLS estimates in two-level models: A Monte Carlo study with ML3', Unpublished manuscript, Leiden University, the Netherlands.

Wennberg, J.E. (1985). 'Variations in medical practice and hospital costs', *Connecticut Medicine,* vol.49(7), pp. 444-453.

Wennberg, J.E., Barnes, B.A., and Zubkoff, M. (1982). 'Professional uncertainty and the problem of supplier-induced demand', *Social Science & Medicine,* vol.16(7), pp. 811-824.

Wennberg, J.E. and Gittlesohn, A. (1973). 'Small area variation in health care delivery', *Science,* vol.182(4117), pp. 1102-1108.

Westert, G.P., Nieboer, A.P., and Groenewegen, P.P. (1993). 'Variation in duration of hospital stay between hospitals and between doctors within hospitals', *Social Science & Medicine,* vol.37(6), pp. 833-839.

Wilensky, G.R. and Rossiter, L.F. (1983). 'The relative importance of physician-induced demand in the demand for medical care', *The Milbank Memorial Fund Quarterly, Health and Society,* vol.61(2).

Winand, E. (1997). 'Credence Goods and Fraudulent Experts', *RAND Journal of Economics,* vol.28(1), pp. 107-119.

Wolinsky, A. (1994). 'Small Deviations from Maximizing Behavior in a Simple Dynamic Model'. *Quarterly Journal of Economics*, vol.109(2), pp. 443-464.

World Health Organisation. (2000). 'The effects of economic and plicy incentives on provider practice', Geneva.

Appendix A

Table 4.10: Differences in practice variation levels – all investigations

Differences in practice variation levels (All investigations) between job plan contract types at different values for the volume measure (from 5 to 50 cases in multiples of 5 cases), *countconst*

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.cjobplan						
_at						
1	-18.38093	.8466531	-21.71	0.000	-20.04034	-16.72152
2	-13.88726	.8276226	-16.78	0.000	-15.50937	-12.26515
3	-9.393602	.810388	-11.59	0.000	-10.98193	-7.805271
4	-4.89994	.795066	-6.16	0.000	-6.458241	-3.34164
5	-.4062786	.7817693	-0.52	0.603	-1.938518	1.125961
6	4.087383	.7706026	5.30	0.000	2.57703	5.597737
7	8.581045	.7616597	11.27	0.000	7.08822	10.07387
8	13.07471	.7550195	17.32	0.000	11.5949	14.55452
9	17.56837	.7507431	23.40	0.000	16.09694	19.0398
10	22.06203	.7488711	29.46	0.000	20.59427	23.52979

Note: dy/dx for factor levels is the discrete change from the base level.

Differences in practice variation levels (All investigations) between job plan contract types at different values for the volume measure (from 5 to 50 cases in multiples of 5 cases), *countc*

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.cjobplan						
_at						
1	-3.302217	1.158481	-2.85	0.004	-5.572797	-1.031637
2	-1.637355	.9291144	-1.76	0.078	-3.458386	.1836754
3	.0275061	.8538719	0.03	0.974	-1.646052	1.701064
4	1.692368	.9693339	1.75	0.081	-.2074918	3.592227
5	3.357229	1.222615	2.75	0.006	.9609487	5.75351
6	5.022091	1.54746	3.25	0.001	1.989125	8.055056
7	6.686952	1.907654	3.51	0.000	2.948019	10.42589
8	8.351814	2.286553	3.65	0.000	3.870252	12.83338
9	10.01668	2.676223	3.74	0.000	4.771374	15.26198
10	11.68154	3.07257	3.80	0.000	5.65941	17.70366

Note: dy/dx for factor levels is the discrete change from the base level.

Differences in practice variation levels (All investigations) between job plan contract types at different values for the volume measure (from 5 to 50 cases in multiples of 5 cases), *vol90c*

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.cjobplan						
_at						
1	-.7246725	.8254847	-0.88	0.380	-2.342593	.8932478
2	3.937019	1.002063	3.93	0.000	1.97301	5.901027
3	8.59871	1.403557	6.13	0.000	5.847789	11.34963
4	13.2604	1.89181	7.01	0.000	9.552521	16.96828
5	17.92209	2.414758	7.42	0.000	13.18925	22.65493
6	22.58378	2.954031	7.65	0.000	16.79399	28.37358
7	27.24547	3.502096	7.78	0.000	20.38149	34.10946
8	31.90717	4.05539	7.87	0.000	23.95875	39.85558
9	36.56886	4.612031	7.93	0.000	27.52944	45.60827
10	41.23055	5.170939	7.97	0.000	31.09569	51.3654

Note: dy/dx for factor levels is the discrete change from the base level.

Table 4.11: Differences in practice variation levels – investigations <2 days

Differences in practice variation levels (<2 days investigations) between job plan contract types at different values for the volume measure (from 5 to 50 cases in multiples of 5), *countconst*

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.cjobplan						
_at						
1	-7.049544	.3103984	-22.71	0.000	-7.657913	-6.441174
2	-5.054146	.2837745	-17.81	0.000	-5.610333	-4.497958
3	-3.058748	.2605217	-11.74	0.000	-3.569361	-2.548134
4	-1.063349	.2416152	-4.40	0.000	-1.536907	-.5897923
5	.9320486	.2281382	4.09	0.000	.4849059	1.379191
6	2.927447	.2210858	13.24	0.000	2.494126	3.360767
7	4.922845	.2210738	22.27	0.000	4.489548	5.356141
8	6.918243	.2281032	30.33	0.000	6.471169	7.365317
9	8.913641	.2415601	36.90	0.000	8.440192	9.38709
10	10.90904	.2604502	41.89	0.000	10.39857	11.41951

Note: dy/dx for factor levels is the discrete change from the base level.

Differences in practice variation levels (<2 days investigations) between job plan contract types at different values for the volume measure (from 5 to 50 cases in multiples of 5), *countc*

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.cjobplan						
_at						
1	-.8776613	.5011983	-1.75	0.080	-1.859992	.1046694
2	-.1261681	.3868847	-0.33	0.744	-.8844481	.632112
3	.6253252	.3363899	1.86	0.063	-.033987	1.284637
4	1.376818	.3763443	3.66	0.000	.6391971	2.11444
5	2.128312	.4848818	4.39	0.000	1.177961	3.078662
6	2.879805	.6273627	4.59	0.000	1.650196	4.109413
7	3.631298	.7855292	4.62	0.000	2.091689	5.170907
8	4.382791	.9515917	4.61	0.000	2.517706	6.247877
9	5.134285	1.12205	4.58	0.000	2.935107	7.333462
10	5.885778	1.295169	4.54	0.000	3.347293	8.424263

Note: dy/dx for factor levels is the discrete change from the base level.

Differences in practice variation levels (<2 days investigations) between job plan contract types at different values for the volume measure (from 5 to 50 cases in multiples of 5), *vol90c*

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.cjobplan						
_at						
1	.4627069	.3170679	1.46	0.144	-.1587348	1.084149
2	2.271247	.3987673	5.70	0.000	1.489677	3.052816
3	4.079787	.5821494	7.01	0.000	2.938795	5.220779
4	5.888327	.8001124	7.36	0.000	4.320136	7.456519
5	7.696867	1.030952	7.47	0.000	5.676239	9.717495
6	9.505407	1.267652	7.50	0.000	7.020855	11.98996
7	11.31395	1.507455	7.51	0.000	8.359389	14.26851
8	13.12249	1.749086	7.50	0.000	9.694342	16.55063
9	14.93103	1.991878	7.50	0.000	11.02702	18.83504
10	16.73957	2.235455	7.49	0.000	12.35816	21.12098

Note: dy/dx for factor levels is the discrete change from the base level.

Table 4.12: Differences in practice variation levels – investigations >2 days

Differences in practice variation levels (>2 days investigations) between job plan contract types at different values for the volume measure (from 5 to 50 cases in multiples of 5), *countconst*

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.cjobplan						
_at						
1	-17.60882	1.146235	-15.36	0.000	-19.8554	-15.36224
2	-14.07782	1.100955	-12.79	0.000	-16.23565	-11.91999
3	-10.54681	1.061596	-9.93	0.000	-12.6275	-8.466121
4	-7.015805	1.02884	-6.82	0.000	-9.032295	-4.999316
5	-3.484799	1.003333	-3.47	0.001	-5.451296	-1.518302
6	.0462075	.9856382	0.05	0.963	-1.885608	1.978023
7	3.577214	.9761802	3.66	0.000	1.663936	5.490492
8	7.10822	.9751988	7.29	0.000	5.196866	9.019575
9	10.63923	.9827194	10.83	0.000	8.713132	12.56532
10	14.17023	.99855	14.19	0.000	12.21311	16.12736

Differences in practice variation levels (>2 days investigations) between job plan contract types at different values for the volume measure (from 5 to 50 cases in multiples of 5), c*ountc*

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.cjobplan						
_at						
1	-5.07439	1.176213	-4.31	0.000	-7.379725	-2.769054
2	-3.824727	.9876803	-3.87	0.000	-5.760544	-1.888909
3	-2.575064	.9236499	-2.79	0.005	-4.385384	-.7647431
4	-1.325401	1.008131	-1.31	0.189	-3.301301	.6504997
5	-.0757375	1.210417	-0.06	0.950	-2.448111	2.296636
6	1.173926	1.483062	0.79	0.429	-1.732822	4.080673
7	2.423589	1.794273	1.35	0.177	-1.093122	5.940299
8	3.673252	2.127191	1.73	0.084	-.4959668	7.84247
9	4.922915	2.473065	1.99	0.047	.0757959	9.770033
10	6.172578	2.827144	2.18	0.029	.6314776	11.71368

Note: dy/dx for factor levels is the discrete change from the base level.

Differences in practice variation levels (>2 days investigations) between job plan contract types at different values for the volume measure (from 5 to 50 cases in multiples of 5), *vol90c*

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
1.cjobplan						
_at						
1	-3.416529	.9204895	-3.71	0.000	-5.220655	-1.612402
2	.3281328	.9476628	0.35	0.729	-1.529252	2.185518
3	4.072794	1.169931	3.48	0.000	1.779771	6.365817
4	7.817456	1.503105	5.20	0.000	4.871425	10.76349
5	11.56212	1.889396	6.12	0.000	7.858969	15.26527
6	15.30678	2.302221	6.65	0.000	10.79451	19.81905
7	19.05144	2.729567	6.98	0.000	13.70159	24.40129
8	22.7961	3.165559	7.20	0.000	16.59172	29.00048
9	26.54076	3.607063	7.36	0.000	19.47105	33.61048
10	30.28542	4.052277	7.47	0.000	22.34311	38.22774

Note: dy/dx for factor levels is the discrete change from the base level.

Appendix B

Table 4.13: Main investigations ordered for patients undergoing PTCA treatment

Investigation/test name	Number of investigations carried out
Renal Profile (Serum)	6785
CK (Serum)	6783
Estimated GFR	5657
Full Blood Count	3465
APTT	3130
Troponin I	2796
Coagulation Screen	2419
Chest X-Ray	1035
Magnesium (Serum/Plasma)	1010
Lipid Profile (Serum)	956
Glucose - Random (Urgent) YELLOW CAP	925
Liver Profile (Serum)	894
Calcium and Phosphate (Serum)	699
CKMB Isoenzyme (Serum)	654
CKMB/CK Ratio (Serum)	652
Amylase (Serum/Plasma)	495
Glucose - Random (Plasma)	486
Type and Screen	445
INR	443
Osmolality Calculated	412
Thyroid Function Test	401
Calcium (Serum)	337
Alpha-Hydroxybutyric Acid (Serum)	334
TSH and FT4	302
Antibody Screen.	298
AST (Serum)	297
C-Reactive Protein (Serum)	279
Prothrombin Time (PT)	244
Protein and Albumin (Serum)	206

Source: Analysis of hospital episode data.

Appendix C

In view of the envisaged statistical concerns expected due to the relative small size of the dataset, a one-stage multilevel model was used to check for the robustness of the effect of job contract type and consultant volume levels on the number of investigations ordered by consultants within the public hospital. Tables 4.14-4.16 below represent the results obtained.[23] Table 4.14 presents the results when all the investigations carried out during the hospital stay period are taken into account. Table 4.15 presents the results when the investigations during the first two days of hospital stay are taken as the dependent variable. Table 4.16 presents the results based for the investigations carried out in the post 2 day hospital stay period.

The results in Table 4.14 (all investigations) show that age and the measure of severity (Charlsonl) are positive and statistically significant in explaining the number of investigations for PTCA patients. The results in columns 2-5 are roughly in line with those obtained from the two stage estimation process showing that the job contract type of the consultant is not statistically significant in explaining differences in the number of investigations when the full set of investigations carried out during the patient hospital stay is considered. The robustness of the results obtained for the

[23] To ease in the comparison of results the tables include a reference heading within each column to indicate the comparable column of results already presented in Section 4.6 of the study.

various volume indicators using the two stage regression is questioned when such results are compared to the values obtained from the one-stage multilevel model estimation. Whilst similar results in terms of magnitude of effect are obtained the volume coefficients are found to be statistically insignificant (columns 2-5).

The results in column 6-8 of Table 4.14 include the interaction term between the various volume indicators used and the job contract type of the consultant. These results are comparable to the results in Table 4.6. The robustness of the age and severity measures is confirmed. A negative and significant coefficient for patients being seen by *public-only* contracts is also obtained in column 6 using the one-stage multilevel modelling approach[24]. The results obtained for the different volume measures in columns 6-8 are similar to those obtained under the two stage modelling framework (although statistically insignificant) [25]. Similar results are recorded for the *public-only* by volume (countconst) interaction term between the two estimation methods. This represents the difference in the responsiveness of practice variation to volume changes when consultants on *public-only* contracts are considered compared to those consultants on *dual practice* contracts.

The results within the lower section of the table (the random effects parameters) show that the standard deviation of the error term at the consultant level is significant although very small (indicating very slight variation in the intercept due to the

---

[24] The coefficient for *public only* contracts in columns 7-8 are insignificant.

[25] This confirms some of the concerns related to the use of the two stage approach – of obtaining biased standard error values, thus leading to inaccurate 'p' and 't' values.

consultant effect). The Intraclass Correlation Coefficient (ICC)[26] measures the proportion of the variance explained by the grouping structure. Very low values for the ICC are obtained using the results obtained from Table 4.14. This indicates that a very low proportion of the variance is explained by the grouping structure.

[26] Estimated using the formulation in Hox (2010) 'Multilevel Analysis –Techniques and Applications' Second Edition, pg 15.

Table 4.14: One-stage multilevel model results with 'inv' as the dependant variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reference table in Chapter	Table 4.3 (1)	Table 4.4 (1)	Table 4.4 (2)	Table 4.4 (3)	Table 4.4 (4)	Table 4.6 (1)	Table 4.6 (2)	Table 4.6 (3)
	inv	inv	inv	inv	inv	inv	inv	inv
Fixed effect parameters								
age	0.159* (0.0568)	0.158* (0.0602)	0.155* (0.0634)	0.158* (0.0576)	0.157* (0.0599)	0.160* (0.0545)	0.161* (0.0541)	0.156* (0.0613)
CharlsonI	18.15*** (0.0000)	18.17*** (0.0000)	18.27*** (0.0000)	18.08*** (0.0000)	18.21*** (0.0000)	18.04*** (0.0000)	18.13*** (0.0000)	18.30*** (0.0000)
public-only		-0.617 (0.8768)	-1.690 (0.6582)	-0.126 (0.9728)	-0.890 (0.8115)			
cvol		-0.290 (0.8984)						
countconst			-0.00512 (0.3381)			-0.00514 (0.3353)		
countc				0.00209 (0.7742)			0.00222 (0.7596)	
vol90c					-0.0251 (0.6422)			-0.0212 (0.6959)
public-only						-22.68** (0.0500)	7.631 (0.2473)	1.862 (0.6940)
public-only#countconst						0.888* (0.0546)		
public-only #countc							-0.651 (0.1559)	
public-only #vol90c								-0.607 (0.3451)
Constant	15.32*** (0.0050)	15.64*** (0.0078)	16.94*** (0.0030)	15.11*** (0.0064)	16.03*** (0.0046)	16.75*** (0.0033)	14.93*** (0.0070)	15.94*** (0.0048)
Random effects parameters								
sd(_cons)	0.000000449*** (0.0001)	0.00000592*** (0.0026)	1.31e-09*** (0.0000)	3.90e-08*** (0.0000)	0.000000159*** (0.0005)	8.58e-08*** (0.0002)	0.000000143*** (0.0002)	0.000000294*** (0.0001)
sd(residual)	30.44*** (0.0000)	30.44*** (0.0000)	30.43*** (0.0000)	30.44*** (0.0000)	30.44*** (0.0000)	30.38*** (0.0000)	30.41*** (0.0000)	30.43*** (0.0000)

p-values in parentheses, * p<0.10, ** p<0.05, *** p<0.010
The baseline, *public-only*=0, indicates consultant is on a *dual practice* contract.
sd(_cons) refers to the standard deviation at the consultant level, sd(residual) refers to the standard deviation at the patient level.

The results in Table 4.15 deal with the investigations carried out over the first two days of hospital stay. The results obtained for the age and severity variables using the one-stage multilevel modelling approach are broadly in line with those obtained using the two stage approach. Age is negatively related to having investigations performed within the first 2 days of hospital stay. A positive and significant coefficient is obtained for the severity variable. The results in columns 2-5 are roughly in line with those obtained from the two stage fixed effect estimation process showing that the job contract type of the consultant is not statistically significant in explaining differences in the number of investigations during the first two days of hospital stay. The robustness of the results obtained for the volume indicators in the two stage estimation method (measured by *cvol* and *countc*) are confirmed from the results obtained from the one-stage multilevel model estimation (same sign and magnitude).

The results in column 6-8 of Table 4.15 include the interaction term between the various volume indicators used and the job contract type of the consultant. The robustness of the age and severity measures is confirmed. There is some variation in the results obtained for the coefficient value of the consultant job plan contract variable. The results in columns 6,7,8 compared to those obtained from the two stage modelling framework show that the original results lack some robustness. The significant and positive volume coefficient (measured by *countconst*) for the interaction term, indicates a positive difference in the responsiveness of practice variation when patients are under the care of *public-only* contract consultants in comparison to those on *dual practice* contracts.

The results within the lower section of the table (the random effects parameters) show that the standard deviation of the error term at the consultant level is not significant indicating no variation in the intercept due to the consultant. Very low (close to zero) ICC values are obtained when using the results from Table 4.15, thus indicating a low proportion of the variation being explained by the consultant grouping structure.

Table 4.15: One-stage multilevel model results with 'invf2d' as the dependant variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reference table in Chapter	Table 4.3 (2)	Table 4.5 (1)	Table 4.5 (2)	Table 4.5 (3)	Table 4.5 (4)	Table 4.8 (1)	Table 4.8 (2)	Table 4.8 (3)
	invf2d	invf2d	invf2d	invf2d	invf2d	invf2d	invf2d	invf2d
Fixed effect parameters								
age	-0.110*** (0.0013)	-0.106*** (0.0019)	-0.107*** (0.0016)	-0.111*** (0.0011)	-0.108*** (0.0016)	-0.107*** (0.0017)	-0.111*** (0.0011)	-0.107*** (0.0016)
CharlsonI	10.56*** (0.0000)	10.47*** (0.0000)	10.51*** (0.0000)	10.42*** (0.0000)	10.50*** (0.0000)	10.46*** (0.0000)	10.42*** (0.0000)	10.50*** (0.0000)
public-only		1.953 (0.2528)	0.999 (0.5843)	0.895 (0.6126)	0.901 (0.6123)			
cvol		2.856*** (0.0078)						
countconst			0.00428 (0.2309)			0.00401 (0.2296)		
countc				0.00705** (0.0455)			0.00708** (0.0451)	
vol90c					0.0496 (0.1066)			0.0517* (0.0995)
public-only						-8.158 (0.1209)	1.729 (0.5434)	1.514 (0.4871)
public-only#countconst						0.393* (0.0654)		
public-only #countc							-0.0710 (0.7079)	
public-only #vol90c								-0.143 (0.6190)
Constant	23.45*** (0.0000)	21.77*** (0.0000)	22.60*** (0.0000)	22.98*** (0.0000)	22.61*** (0.0000)	22.65*** (0.0000)	22.94*** (0.0000)	22.55*** (0.0000)
Random effects parameters								
sd(_cons)	1.696 (0.2029)	0.970 (0.9631)	1.632 (0.2037)	1.625 (0.2127)	1.620 (0.2060)	1.446 (0.3645)	1.656 (0.1955)	1.701 (0.1710)
sd(residual)	12.35*** (0.0000)	12.35*** (0.0000)	12.34*** (0.0000)	12.33*** (0.0000)	12.34*** (0.0000)	12.33*** (0.0000)	12.33*** (0.0000)	12.33*** (0.0000)

p-values in parentheses, * p<0.10, ** p<0.05, *** p<0.010
The baseline, *public-only*=0, indicates consultant is on a *dual practice* contract.
sd(_cons) refers to the standard deviation at the consultant level, sd(residual) refers to the standard deviation at the patient level.

The results in Table 4.16 deal with the investigations carried out during the post 2 day hospital stay for the patient. Results show that both age and the measure of severity are positive and statistically significant in explaining the number of investigations for PTCA patients carried out in the post 2 day hospital stay period. The results in column 2-5 for the coefficient of the *public-only* variable and the various volume indicators are roughly in line in terms of magnitude and sign when compared with those obtained when undertaking the two stage estimation procedure. The statistical robustness of the results obtained from the two stage method is though questioned given that the coefficient obtained on the *public-only* variable and the various volume indicators lack statistical significance.

The results in column 6-8 of Table 4.16 include the interaction term between various volume indicators and the job contract type of the consultant. The LR test carried out between these three specifications and the comparable specifications in columns 3-5 confirm an improvement in specification (LR values of 0.11, 0.0235, 0.0481 respectively). The robustness of the age and severity measures is confirmed. Results show that similar coefficients in terms of magnitude and sign (although insignificant) are generally obtained for each of the volume indicators using the one-stage multilevel modelling approach. This measures the responsiveness of the number of investigations to changes in volume of patients seen by consultants on *dual practice* contracts. The results obtained in column 6 show a significant and negative coefficient for the consultant job plan variable (-21.12), indicating a lower level of investigations for patients seen by consultant being under the care of consultants who have *public-only* contracts compared to *dual practice* contract consultants. This result confirms the conclusions obtained in column 4 of Table 4.8. A similar result (although of a smaller magnitude and statistically insignificant) for the coefficient of the job plan variable is obtained when the volume indicator being considered is *vol90c*. A positive although statistically insignificant coefficient for the interaction term between the volume measure *countconst* and job plan is obtained raising some robustness issues with the coefficients of the two stage model. There are some concerns with regards to the interaction terms obtained in columns 7 and 8 of Table 4.16 in comparison to the results obtained for similar estimations using the two stage approach. A negative and significant coefficient for the interaction term is obtained when the one-stage multilevel modelling approach is applied indicating a negative difference in the responsiveness of the number of investigations ordered as volume changes for patients who are under the care of *public only* contract consultants.

The results within the lower section of the table (the random effects parameters) show that the standard deviation of the error term at the consultant level is significant indicating some variation in the intercept due to the consultant effect. The Intraclass Correlation Coefficient (ICC) obtained for the results in Table 4.16 particularly for the estimations carried out in column 4, 7 and 8 show that some of the variance in the number of investigations carried out in the post 2 days hospital stay period is indeed at the consultant level.

Table 4.16: One-stage multilevel model results with 'inva2d' as the dependant variable

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Reference table in Chapter	Table 4.3 (3)	Table 4.5 (5)	Table 4.5 (6)	Table 4.5 (7)	Table 4.5 (8)	Table 4.8 (4)	Table 4.8 (5)	Table 4.8 (6)
	inva2d	inva2d	inva2d	inva2d	inva2d	inva2d	inva2d	inva2d
Fixed effect parameters								
age	0.363*** (0.0000)	0.344*** (0.0000)	0.355*** (0.0000)	0.364*** (0.0000)	0.357*** (0.0000)	0.362*** (0.0000)	0.373*** (0.0000)	0.376*** (0.0000)
CharlsonI	4.086** (0.0101)	4.576*** (0.0050)	4.391*** (0.0065)	4.253*** (0.0089)	4.358*** (0.0068)	4.210*** (0.0092)	4.287*** (0.0081)	4.452*** (0.0056)
public-only		-5.188 (0.1913)	-4.242 (0.2627)	-6.208 (0.2835)	-3.806 (0.3016)			
cvol		-3.582 (0.1297)						
countconst			-0.00726 (0.1867)			-0.00721 (0.1893)		
countc				0.00415 (0.6626)			0.00569 (0.5549)	
vol90c					-0.0713 (0.1909)			-0.0476 (0.6413)
public-only						-21.12* (0.0613)	4.173 (0.6234)	-0.515 (0.9523)
public-only#countconst						0.704 (0.1125)		
public-only #countc							-1.005** (0.0207)	
public-only #vol90c								-2.200*** (0.0044)
Constant	-8.417 (0.1224)	-4.629 (0.4352)	-5.953 (0.2969)	-5.686 (0.3385)	-6.344 (0.2598)	-6.307 (0.2688)	-5.333 (0.3816)	-4.089 (0.5171)
Random effects parameters								
sd(_cons)	0.000000241*** (0.0000)	8.43e-08*** (0.0002)	4.18e-08*** (0.0000)	8.065*** (0.0096)	0.00000233*** (0.0034)	8.70e-08*** (0.0000)	10.63*** (0.0000)	13.29*** (0.0000)
sd(residual)	26.39*** (0.0000)	26.35*** (0.0000)	26.36*** (0.0000)	26.02*** (0.0000)	26.36*** (0.0000)	26.32*** (0.0000)	25.80*** (0.0000)	25.63*** (0.0000)

p-values in parentheses, * p<0.10, ** p<0.05, *** p<0.010

The baseline, *public-only*=0, indicates consultant is on a *dual practice* contract.

sd(_cons) refers to the standard deviation at the consultant level, sd(residual) refers to the standard deviation at the patient level.

The results from the one-stage multilevel modelling approach provide some element of robustness check for the coefficients derived from the two stage modelling approach. The results show that the small sample sizes available and used within this study might be a constraint to the statistical significance of the obtained results. An increase in the sample size may indeed help to obtain more robust estimates. Policy recommendations based on the above results are thus to be articulated with caution and in recognition to this important limitation.

Chapter

5

Conclusions and Recommendations

This study was set out to assess the activity undertaken within the very specific and particular health care sector in Malta. To this effect, the measurement of hospital health care output, through the applicaton of a DRG system, serves as a tool to help describe, manage and measure resource use. The study also sought to evaluate and understand the relationship between health care outcomes and the behaviour of physicians working at the public hospital. Of specific interest to the study is the role of the specific consultant job contract type arrangement in explaining hospital outcomes. The relationship between the health care providers, the hospital authorities and the patients is expected to influence the current and future workings of the Maltese health care system. It is the scope of this study to contribute towards a better understanding of this relationship as this leads to improvements in the setting of policy within the health care sector.

This section presents the conclusions of the chapter specific analysis carried out within this thesis. The potential implications on the overall running of the health care system is discussed and the relevance of the conclusions within the context of the known literature is assessed. Whilst acknowledging that this study has encountered a number of limitations, a number of recommendations, both for further research and for policy implementation purposes, are proposed for consideration of the health care authorities.

In the context of the available data for the Maltese health care sector, the results of the research show that there is a firm basis for the implementation of a DRG casemix classification system, at least initially to help describe and manage resource use. Every medical case treated at a hospital carries with it an economic component and the use of DRGs would allow hospital management authorities to think along product lines. In the context of the ever escalating health care expenditure and the increasing pressure to control spending, DRGs provide policy makers with an additional tool to achieve more transparent and efficient resource use.

The application of the MS-DRG (Version 27) Grouper software, although not specifically configured for the Maltese health care system, proved to be a good starting point given that this was the first occasion for Maltese data to be applied to a casemix Grouper software. As noted in the literature, a health care Grouper software is considered useful if it partitions the hospital episode population in an informative way, both in terms of the clinical inputs required and also the resources that are expected to be used in the treatment of patients. The Coefficient of Variation (CV) and the Coefficient of Multiple Determination (R^2) applied in this analysis showed that the derived DRGs carried a certain degree of homogeneity within the set groups and also demonstrated an adequate degree of heterogeneity when compared to the other derived DRG categories. An R^2 obtained of approximately 0.3 compared well with studies reported for other countries. Furthermore, following a data trimming exercise, around 85% of the DRG cases had a CV of less than 1 implying that the DRGs obtained were homogeneous within the groups identified.

One recognizes that most LOS distributions used in this study are asymmetric and the presence of high outlier values has a significant impact on the R^2 values obtained. In view of the characteristics of the data under analysis and given the relative smallness of the dataset, outlier values are removed from the study and various trimming algoritms are applied. Nothwithstanding this process, the distributional properties of the data where found to still not fully satisfy the estimation requirements and this is considered as a limitation of this study. One though finds that the use of the MDC categories in such instances helps to achieve better distributional properties for the data in hand.

Homogeneous DRG categories provide the policy maker with an additional tool to gauge efficiency in the treatment of cases as it identifies those cases that use more hospital resources than the average case. However, the variation in resource use may also be due to other factors not specifically captured by the DRG classification. These include the individual and behavioural characteristics of the consultants and surgeons working within the hospital. The DRG system provides a solid basis for studying the performance of the consultants and surgeons working within the hospital. The variation in health care outcomes and practice patterns arising due to such factors was studied in Chapters 3 and 4 as part of this research.

In view of the fact that this study is the first application of a casemix system grouper to the Maltese health care system, one recognises the limitations imposed by the deficiencies in the current available dataset. The data used for this study were not collected specifically for the application of a DRG system and therefore, considerable improvements in the configuration of the DRG categories will be made if missing or miscoded data is kept to a minimum. It is positive to note that nothwithstanding such limitations, the results obtained are fairly in line with other literature focusing on similar countries. The experience from other health care systems shows that the introduction of the DRG system will itself lead towards an improvement in the data collection process. This is a positive implication of the introduction of the DRG system. The success of the DRG system depends primarily on the quality of the collected data and thus the collection of more accurate information will serve to provide an added instrument for the allocation of the public hospital budget. This provides a better picture of the current hospital outputs being derived for the amount of resources invested by the government in the health care system.

The research undertaken reveals that there is a need to adequately deal with outlier cases especially in view of the size of the Maltese health care sector. In fact, dealing with defined outliers would result in average savings of approximately 1 day in terms of length of stay for the hospital. Based on insights gained from this study, further work on the understanding of the factors which underpin the presence of outliers is recommended. Indeed, given the small size of the dataset, the applied trimming options have an impact on the results, apart from the fact that there could be other factors which need to be controlled for when trying to understand the underpinning differences in resource utilisation within DRGs. Furthermore, with DRGs in place, the Maltese hospital authorities would be in a better position to plan for cases which are likely to end up as outlier cases. Also of interest for further study is the fact that a number of the obtained DRG groups include only few episodes of care, primarily due to the given size of the health care sector. This entails a further challenge to the use of DRG's when applied to aid in the process of budget allocation for the different health care activities undertaken within the hospital.

In the author's view, although the introduction of such a system in Malta may be a long and complex process, which will require both the commitment of the hospital management team and hospital practicing professionals, a functioning casemix system would serve to benefit the entire health care provision system in Malta. In view of the current economic and political commitments, the use of the casemix system for financing purposes may be still far from materializing however having a DRG system in place would lead to more informed decisions and potentially better policy. On a practical level, the information gained from the setting of a DRG system could serve as an instrument for the allocation of public hospital budgets. Activity that can be measured can be managed more effectively and efficiently.

It is recommended that a different DRG Grouper software system, rather than the MS DRG (Version 27.0), should be applied to the data to obtain further evidence in relation to the gains of applying a casemix system. This would serve to further assess the quality of the available data. It is recommended that one should consider possible adaptations to the already available casemix Groupers used by other countries to make them more appropriate to the particular data and characteristics of the Maltese health care system. Most countries adopting a DRG system have chosen to go down this route. Furthermore, this study would benefit from the incorporation of additional years of data as this would help to assess the consistency of hospital activity measurement for the years under study.

The incentives and benefits resulting from the introduction of the DRG casemix system are expected to have a significant impact on the economic underpinnings of the health care system. The benefits expected to be achieved within the health care system following the introduction of DRG's are based on the experience of other countries and reflect the realities of the current state of the Maltese health system. The experience gained from other countries shows that such a development will have an impact on both resource use and the overall management of the available resources. The most recent developments within the Maltese health care system are increasingly pointing towards the need for the adoption of a DRG casemix system.

Chapter 3 of this thesis focused on competing risks events to assess the nature of the volume-outcome relationship for consultants and surgeons practicing at the hospital level. Whereas most of the literature has sought to analyse the volume-outcome relationship amongst different hospitals, this study specifically looked at understanding the volume-outcome relationship at the individual consultant and surgeon level operating within a single and sole public hospital setting. A number of different volume measures where adopted to assess the sensitivity of the results to changing volume definitions. The impact of different consultant job contract conditions on the relationship between the event of interest and volume levels was of particular interest.

Data for PTCA activity undertaken between 2009 and 2011 within the Maltese health care sector were applied to both non-parametric and semi-parametric survival analysis methods. Both methods confirmed differences in the treatment failure probability due to differences in consultant and surgeon volume levels and variations in the consultant job contract characteristics. The robustness of the results obtained using the survival modelling methods where confirmed though the use a multinomial logistic model. The research found that the treatment failure probability was slightly higher if the patient was under the care of consultants with high patient volume levels or if the procedure was performed by surgeons with low volume levels. The results obtained for patient outcomes as surgeon volume levels increase support the practice makes perfect hypothesis. The same result cannot be confirmed when consultant volume levels are considered.

The research also concluded that the rate of failure from the treatment procedure was slightly lower if patients were seen by consultants operating exclusively with the public sector. Consultants on *public-only* contracts had low patient volume levels compared to consultants on *dual practice* contracts and this could be one possible explanation for this result. Furthermore, this would imply that consultants on *public-only* contracts are in a better position to manage there public hospital workload given that they are fully focused on their practice within the public hospital. The performance of *dual practice* contract consultants within the public hospital, as viewed by the patient in general, is of concern to such consultants as this is expected to have an impact on their reputation.

Furthermore, the impact of changes in volume on the hazard rate of the events of interest was found to be affected by the consultant job plan contract type. The analysis showed that there was only a slight increase in the incidence of failure from the intervention for the patient under the control of a consultant on a *dual practice* contract, if volume was increased. If a patient was under the control of a consultant with a *public-only* contract, then the incidence of failing increases by more when volume changed compared to those patients under consultants with *dual practice* contracts. A unit change in volume was found to have a bigger impact on the hazard rate of failing if patients were under the care of consultants with *public-only* job contracts. The clear policy implication from this conclusion is that policies aimed at varying patient volume levels for consultants on *public-only* job contracts could result in an impact on outcomes and have to be treated with extreme caution. The research also showed that the likelihood of the event *death* is primarily affected, as expected, by the covariates for age and patient LOS at the hospital.

Whilst noting the conclusion from this study one acknowledges a number of limitations. The survival models used assume homogeneity where all individuals are subject to the same risks embodied in the hazard function. This study controls for a range of patient and consultant/surgeon characteristics however there could indeed still be unobserved sources of heterogeneity in the estimations. The empirical estimates obtained from these models are sensitive to the choice of the 'a priori' identifying assumptions and the distributional assumptions of the unobservables can have an impact on the results obtained. This could lead to estimates which poorly describe the true behavioural models generating survival analysis data.

There is also concern that potential confounding effects on the volume-outcome relationship can arise from the fact that the consultant choice of contract type depends on particular issues which would indirectly influence the outcome of the procedure. The ability to disentangle and single out the factors affecting the choice of contract type by the consultant would help to gauge their particular impact on outcomes which impact would have otherwise been incorporated within the consultant's chosen job contract type. Morover, the available data lacks information on the contract type conditions of the surgeons working within the public hospital. A possible extension for this study could be to incorporate additional information on the characteristics of the surgeons working within the hospital and thus to analyse their particular impact of patient outcomes and how surgeons and consultants interact when treating a patient.

The importance which consultants place on the role of their involvement within the private sector for their personal career development could also be a factor which affects the choice of job contract type. Further work in this regard should be considered. To this effect, data on the different skill levels of consultants, such as the number of years of practising as consultants and the university institution from which they obtained their training, would be possible proxies. Furthermore, the collection of information in terms of their work and leisure preferences and the number of hours spent working within the private sector could serve as indications of the importance which consultants place on their activity in the private sector.

We do not have data on patient preference for being treated by *dual* or *public-only* consultants. However, consultants might base their choice of job contract type also on what they perceive to be the mostly sought out form of contract type requested and desired by the patient. The patient's preference can have an influence on consultant practice as patients with certain preferences might seek out care from consultants on particular job contract types. If patients have the ability to somehow influence the job contract type decision of who is responsible for their treatment then this could have an impact on overall outcome differences between consultants.

Nevertheless one recognises the unavailability of data in this regard and this makes it difficult to single out the potential confounding effects. Work in this area is being suggested as one of the possible avenues for future research. Whereby one can study the identification of patient preferences which are expected to influence the decision of consultants to choose one type of contract over another and the impact of this on the volume-outcome relationship. Determining the causes of the effect of contract type on outcome is important especially for policy reasons.

The fact that this study focuses on a single hospital setting, whereby no other public sector hospitals exist would serve to reduce the possible confounding effects which arise from the choice available to consultants to practice in other public sector institutions. The identification of the volume-outcome relationship could in the context of this study be partly controlled by the fact that there is only one single public hospital on the island. Furthermore, the size of the sample used within this study is constrained by the fact that the sample includes all the existing PTCA activity carried out by consultants within the hospital over the three year period, thus serving to reduce potential biases in sample selection.

The above results provide an important contribution towards understanding the possible interlinkages which exist within the setting of the single hospital in relation to consultant and surgeon volume levels and to other consultant related characteristics. One would expect that even within a single hospital setting, differences in the behaviour of consultants and surgeons at an individual level exist and such differences could have an impact on outcome. Decisions by policy makers to implement measures which are likely to affect the volume levels of both surgeons and consultants working within the hospital should therefore also consider the specific job working conditions of the consultants. The widening of this study to other specific procedures within the hospital would help policy makers improve their assessment of the sensitivity of changes in volume and other consultant and surgeon characteristics on hospital outcomes. This contributes towards the setting up of effective policy initiatives within the health care sector.

Chapter 4 of this thesis studied the heterogeneity in practice patterns, analysed at the individual consultant level within the single hospital setting. The emphasis was on whether variation depends on the consultant job plan contract type after controlling for patient and consultant characteristics. In the developed theoretical model consultants are assumed to gain utility from the income earned and the reputation gained from the success achieved when exerting effort in the provision of health care services.

The empirical analysis used data drawn from investigations on patients who received PTCA treatment between 2009 and 2011. These investigations were further subdivided into those performed within the first two days of hospital stay and those performed during the rest of the patients' hospital stay period. The analysis implicated that after controlling for a number of patient characteristics, a proportion of the observed practice variation occurred at the consultant level. Furthermore, practice variation was found to differ by consultant job contract type and varied in magnitude in relation to the period when the investigations were undertaken.

The use of the available administrative (not sample) patient level data helps to deal with differences in patient characteristics whilst offering analytical advantages in terms of the analysis of consultant characteristics which are expected to affect practice variation. The two stage model used in this study is based on the individual based data thus allowing for the partitioning of the overall variation into that due to differences in patients and that arising from differences in the consultant propensity to prescribe investigations or tests. The results from the two-stage model are evaluated against those based on a one stage multilevel model which includes group level predictors as explanatory variables.

As expected, a proportion of the variation in investigations carried out in the post 2 day hospital stay period could be attributed to consultant related differences. The variation in practice patterns, during the post 2 days of hospital patient stay was found to be negatively related to the fact that consultants operate on exclusive public sector contracts. Patients under the care of consultants who also practiced privately were likely to register higher variation in practice patterns for investigations carried out in the post two day hospital stay period. From this observation it can be implied that consultants are more likely to use the post two day hospital stay period to implement their own style and treatment preferences when dealing with patients.

The relationship between consultant volume levels and practice variation also differed under the different periods of study. A positive relationship between volume levels and practice variation was found for investigations carried out in the first two days of patient hospital stay. On the other hand, greater variation was found in the number of investigations carried out in the post two day patient hospital stay period when consultant volume levels fell. A possible explanation for this is that consultants with lower patient volume levels would have more flexibility to order additional investigations for their patients during the patients' post two day hospital period.

The job plan contract type of the consultant was found to have a compounding effect on the relationship between volume and practice variation. An increase in the volume levels of consultants who can only practice in the public hospital will lead to an increase in practice variation. The research found that the response of practice variation to changes in volume levels for consultants on *dual practice* job contracts varied between the two different periods. In particular, an increase in volume would lead to an increase in variation during the first two days of hospital stay and a fall in variation during the rest of the hospital stay period. Therefore it can be concluded that consultants on *dual practice* contracts would increase variation in practice if they had lower volume levels in the post two day patient hospital stay period.

The difference in practice variation levels between consultants with different job plan contracts also varied at different values of volume. The difference in practice variation was found to be negative at very low levels of volume for patients being seen by consultants on *public-only* contracts compared to *dual practice* contracts. However, this difference turned positive and increased as values of volume levels increased.

A number of data limitations have to be highlighted. Data on activity by consultants working in the private sector was not available and this prevented the study of the behaviour of the consultant when such activity levels changed. There could indeed be an element of variation in practice patterns which depends on the size of the activity in the private sector for the consultants who also practice within the public hospital.

Furthermore, information is not available on how many of the patients who receive PTCA treatment at the public hospital, seek private sector intervention prior or post to their admittance to the public hospital. Of particular interest to this study would be a measure of how much of the currently undertaken investigations within the public hospital could in fact be carried out within the private sector setting. The shift of activity from the current public hospital setting to the private setting will help to free up resources within the public hospital having significant implications on the overall running of the hospital. Due to the lack of data, no conclusions could be drawn on the merits of undertaking any of this activity in the private sector once the patient is discharged from hospital.

A possible extension of this study is to focus on the relative importance of each of the investigations being carried out at the public hospital, No particular weighting structure is applied to the different investigations ordered and carried out for the treatment of PTCA within this study. Work on the weighting of investigations and its resulting impact on practice patterns is left as a point for further research.

In view of the very limited data available on surgeons working within the hospital, surgeon characteristics are not controlled for in this study. It is the consultant, who is responsible for the patient, who orders such investigations. Surgeons, who work under the control of the consultant, could also be a source of explanation of practice variation. Further research in this area would incorporate the role of surgeon characteristics and specifically the surgeon job plan contract conditions to help in the explanation of practice variation patterns.

Furthermore, whilst noting that in the study we have controlled for a range of patient and consultant characteristics one though recognises that there are a number of other factors which consultants put into the balance when making a choice between a *dual practice* contract and a *public-only* contract with the public hospital. The observed results may thus be due to factors other than the direct variation arising from job contract type. Data for consultant job contract type may be thus capturing factors which affect the actual decision of the consultant to choose the particular type of job plan contract. There may potentially be other explanations as to why variation in practice differs between consultants rather than the job contract type variable identified in this study. Work on the identification of factors which determine the consultants choice of contract condtions and the impact on practice variation is a possible area for further research.

Issues pertinent to what constitutes a sufficient sample size for accurate estimation and evaluation purposes have been reviewed. The size of the dataset in this study is in fact constrained by the characteristics of the Maltese health care sector. Better inference is expected to be achieved if patient level data is used in such instances. A possible suggestion for future work would be to carry out the analyses across a whole department or unit rather than basing the analysis on a single procedure. This would help to increase the sample size, both at the patient and at the consultant level. A possible drawback would be related to the fact that differences across departments would be expected to be bigger given the diverse range of patients treated. Any possible policy recommendations under such conditions have to be made with caution given the variation in treatments being offered at each department. There is indeed an element of compromise which needs to be achieved between the accuracy of the estimates and the possible use of such estimates for policy recommendation purposes.

In view of the envisaged statistical concerns expected due to the relative small size of the data, a one stage multilevel model was used to estimate the effect of job contract type and consultant volume levels on the number of investigations ordered by consultants within the public hospital. The results obtained from the one stage multilevel modelling approach provide some element of robustness check for the coefficients derived from the two stage modelling estimations. The results show that the small sample size available and used within this study might be a constraint to the statistical significance of the obtained results. The complete population of the PTCA activity undertaken at the hospital is considered in this study and thus the option of increasing the sample size under current conditions is not available. Policy recommendations are to be articulated with caution and in recognition to this important consideration.

The above findings provide interesting results in terms of understanding the role of consultant volume levels and job contract conditions to explain practice variation within this particular hospital setting. Practice variation has implications on efficiency levels within the hospital apart from the expected impact on patient outcomes. The study of the impact of practice variation amongst consultants on patient outcomes within this setting is an area for further research.

Further work, applied to other specific procedures within the hospital setting, would provide policy makers with a better view of the importance of specific consultant characteristics in explaining differences within other treatment practices offered at the hospital. This would give policy makers the opportunity to better appreciate the role of individual consultant characteristics in explaining variation in practice patterns within the whole hospital.

In summary, the research undertaken and described in this thesis recommends a number of important policy considerations that will benefit both the health care setting in Malta and health care practice in general. A hospital cannot be managed efficiently without an adequate measure of output. Furthermore, it is only by fully understanding the impact of certain characteristics of consultant and surgeon working patterns on patient and hospital outcomes, that effective policy options could be formulated. The findings of this thesis may contribute towards the design of policies which are based on the knowledge gained of the link between the variation in practice patterns and volume-outcome relationships to the different consultant job contract types. The conclusions of this thesis help in setting policy strategies based on evidence informed research within the health care sector.