**INVESTIGATING WHETHER THE JOHNS HOPKINS ACG CASE-MIX SYSTEM**

**EXPLAINS VARIATION IN UK GENERAL PRACTICE**

by

Caoimhe O Sullivan

*Thesis submitted for the degree of Doctor of Philosophy of University College London*

# University College London

November 2010

I, Caoimhe O Sullivan, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

_____

Caoimhe O Sullivan

# Abstract

This thesis describes the first large-scale studies in the United Kingdom to adjust for diagnostic-based morbidity when examining variation in home visits, specialist referrals and prescribing patterns in general practice. The Johns Hopkins ACG Case-Mix System was used since each patient's overall morbidity is a better predictor of health service resource use than individual diseases.

A literature review showed large variations in resource use measures such as consultations, referrals and prescribing practice patterns in general practice both in the UK and elsewhere and highlighted inappropriate use of statistical methodology that has the potential to produce misleading and erroneous conclusions. The review presents a strong argument for adjusting for diagnostic based morbidity when comparing variation in general practice outcomes in the UK.

Multilevel models were used to take account of clustering within general practices and partition variation in general practice outcomes into between and within practice variation. Statistical measures for appropriately dealing with the challenging methodological issues were explored with the aim of producing results that could be more easily communicated to policy makers, clinicians, and other healthcare professionals.

The datasets used contained detailed patient demographic, social class and diagnostic information from the Morbidity Statistics in General Practice Survey and the General Practice Research Database.

This research shows that a combination of measures is required to quantify the effect of model covariates on variability between practices. Morbidity explains a small proportion of total variation between general practices for the home visit and referral outcomes but substantially more for the prescribing outcome compared to age and sex. Most of the variation was within rather than between practices.

# Acknowledgements

This thesis would not have been possible without the inspiration and tremendous support of my husband, Alan. I couldn't have done this without him and can't thank him enough. Our beautiful girls became joyful additions to our lives in the midst of this work and they are a wonderful distraction.

A very special thanks to my fantastic parents, sister, brothers and friends for much needed practical help, encouragement and love.

Huge thanks to Dr. Rumana Omar, my chief supervisor, and Prof Azeem Majeed, my second supervisor, for their insightful feedback, expertise and patience. I have learned so much from working with both of them. In particular, thanks to Rumana for encouraging and persisting with me during the final difficult period.

Thankyou to my friends and former colleagues at what was previously known as the Research and Development Directorate, University College London Hospitals NHS Trust, particularly to Prof Allyson Pollock for her perceptive advice.

Thanks also to colleagues at University College London Department of Statistical Science, for their support while I was there as an Honorary Research Fellow. I'd particularly like to acknowledge Dr. Gareth Ambler, for interesting and stimulating debates over related topics.

Thanks to Dr. Irene Petersen and Amir Islam at the Dept of Primary Care and Population Sciences, University College London, for their help extracting GPRD data. I also gratefully acknowledge the help of Dr. Kevin Carroll for sharing his clinical coding work.

I am thankful to the team from the Johns Hopkins Bloomberg School of Public Health for supplying the ACG software and always being available to answer my queries. Particular thanks to Prof. Christopher Forrest for his contribution to one of my papers, and also Prof. Jonathan Weiner, Prof. Andrew Bindman and Dr. Karen Kinder.

My colleagues in the Public Health Directorate of NHS Richmond have been a great support in the final months of writing up and preparing for my viva and I am very grateful to Anna Raleigh, Prof. Jose Ortega, Dr. Usman Khan, Houda Al-Sharifi, Dr. Dagmar Zeuner and Oliver McKinley.

I am very grateful to my examiners, Prof. Irwin Nazareth and Dr. Sonia Saxena. Their great understanding and knowledge of my work helped strengthen the thesis and made the experience both challenging and enjoyable.

This PhD project was funded with a Department of Health National Primary Care Researcher Development Award and I am very grateful to the team there for the wonderful opportunities that this has presented.

Caoimhe

For my family

# Table of contents

# List of tables

# List of figures

# Chapter 1    Introduction

Healthcare resources are limited, and it is important that they are used efficiently and effectively.  Like other developed countries, people's expectations of what they can obtain from health services in the United Kingdom are rising (NHS Plan 2000). At the same time, health care costs have been rising more rapidly than the general rate of inflation, with Primary Care Trusts responsible for over 80% of the NHS Revenue Budget (£74.2bn of the NHS Revenue Settlement (£92.5bn) in 2007/8) (DH Departmental Report 2008). Hence, how health care resources are used, and in particular, whether they are being used efficiently, appropriately and effectively, is coming under increasing scrutiny in the United Kingdom and elsewhere.

The vast majority of the UK population is registered with a general practitioner (GP) and ninety percent of patient contacts with the National Health Service (NHS) occur in primary care (DH Departmental Report 2008).  GPs have autonomy in making decisions as to how their patients are managed, such as whether to prescribe them drugs, or refer them on to specialist care.  Hence, how GPs manage their patients has a direct influence on NHS service use both in primary and secondary care.  This potential of primary care to act as the gatekeeper to the care offered by the National Health Service (NHS) has long been recognised (The NHS Plan (2000); Mant, D. (1997)).  England's public health white paper, *Saving Lives: Our Healthier Nation*, states that within the restructured NHS: "setting standards and measuring progress is now an integral part of the planning and delivery of services to patients in primary care" (DH: Saving Lives: Our Healthier Nation).

Monitoring the decisions made in general practices in areas such as home visits, referrals to hospitals and drug prescribing rates, is one way of keeping track of how healthcare resources are being utilised (Majeed A et al, 2001a).  The implementation of performance monitoring procedures in UK primary care was a key goal set out in The NHS Plan, 2000.  Primary Care Trusts (PCTs) are required by the Department of Health to submit regular audit reports on performance in specific areas such as referral rates for

outpatient care. The introduction of regulatory bodies such as the Commission for Health Improvement (CHI) (set up to improve the quality of patient care in the NHS in England and Wales through monitoring of primary care organisations) and subsequently the Commission for Healthcare Audit and Inspection (CHAI) has meant that general practices' performance has come under much greater scrutiny, building on a trend that began in the 1990s (Majeed FA, Voss S. (1995)). Such monitoring allows apparent extremes to be detected and investigated further to see if they are reasonable given the specific characteristics of a practice.

In the UK, general practice resource use outcomes have been shown to vary widely between general practices (Aylin P, 1996; Majeed FA et al, 1996; Majeed A et al, 2001; Hippisley-Cox J et al, 1997). Comparisons of practice performance, workload and resource utilisation are often presented in terms of crude rates or proportions. These sometimes take into account differences in age, sex, and ecological measures of health and socio-economic status of the patient populations and practice factors such as size of practice population (Carr-Hill RA et al, 1996; NHS Executive, 1999). The use of crude rates or proportions are useful for understanding how many events occur in which groups of individuals, but, in comparisons between general practices these could lead to some practices being unfairly penalised. Case-mix classification is defined as the classification of people or treatment episodes into groups, using characteristics associated with the condition, treatment or outcome that can be used to predict need, resource use or outcome (Sanderson et al, 1998). Adjusting for the age and sex case-mix of practices may be an improvement, but it is possible that practices serving populations with higher morbidity may still be unfairly penalised (Salem-Schatz et al, 1994). For example, a practice serving a sicker population will have a higher workload, which in turn may lead to higher prescribing and referral rates. These adjustments may be sufficient for larger populations such as those of primary care trusts, but general practices are composed of much smaller populations, and so there are likely to be large differences among them in their clinical and socio-economic characteristics (Majeed A et al, 2001b; Salem-Schatz S et al, 1994; Reid R et al, 1999). It is important to identify factors that explain this variability and appropriately adjust for the case-mix of patients to compensate for such differences (Signorini et al, 1999; Majeed, A. et al, 2001b; Fowles et al, 1996). Attentions may be misdirected to problems that are less serious

than perceived, while ignoring the real problem areas. This leads to a waste of time, money and resources.

All general practices in the UK now record patients' clinical diagnoses onto computer, and so there is opportunity to investigate diagnostic based measures of case-mix. Of the several diagnostic based case-mix measurement systems available, an important feature of the Johns Hopkins Adjusted Clinical Groups (ACG) Case-Mix System is that, unlike other case-mix measurement systems, it measures each patients overall morbidity as this has been shown to be a better predictor of health services resource use than examining only specific diseases. The ACG system was developed specifically for use in primary care using primary care data, and is widely used and validated (www.acg.jhsph.edu) (Starfield B et al 1991; Weiner JP et al, 1991). It has been widely used and validated in primary care (Halling et al, 2006; Juncosa S et al, 1997 & 1999; Reid R et al, 1999&2001&2002; Carlsson et al, 2002). Other case-mix systems that measure primary care diagnoses were originally designed for hospital use only (Kahn K et al, 1990; Averill RF et al, 1999; Kronick RT et al, 1996). Most previous studies using case-mix adjustment have relatively homogeneous study groups such as members of a single plan or only the elderly (Fowles et al, 1996). The application of the ACG system in the UK is particularly interesting since most of the population is registered with a general practice (comparing Attribution Data Sets of GP registered populations and corresponding mid year population estimates from the Office for National Statistics).

This study aims to investigate whether variation between general practice outcomes may be explained by patient level diagnostic-based morbidity measures (See Section 2.7 for detailed aims and objectives). The work also aims to explore methods of appropriately dealing with the challenging methodological issues. This should contribute to raising awareness among primary care researchers and statisticians of the necessity for sound statistical input to the primary care research base.

The main objective is to apply the Johns Hopkins ACG Case-Mix System in comparisons of general practice process outcomes in populations in the UK. The system is used to assign case-mix measures to each patient based on a combination of their diagnoses, age and sex. Important general practice outcomes are selected which

have documented evidence of wide variations: home visits, referral and prescribing patterns. Variation between general practices for these outcomes and whether morbidity measures from the Johns Hopkins ACG Case-Mix System can explain some of this variation is examined. Large datasets containing detailed patient demographic and diagnostic information from the Morbidity Statistics in General Practice Survey (MSGP4) and the General Practice Research Database (GPRD) are used for the purpose of this research.

The statistical issues involved in this work are not straightforward due to certain features of the data. Firstly, patients within practices are likely to share more similarities than patients across practices since patients in the same practice will be exposed to the same practice policy and may share common neighbourhood and socio-economic characteristics. This inherent clustering of the data needs to be handled with appropriate statistical models; otherwise it may provide incorrect statistical inferences and lead to potentially misleading and erroneous conclusions (Omar et al. Stats in Med 2000; 19, 2675-2688). Secondly, measuring variation between practices for discrete health outcomes is not straightforward. Thirdly, the datasets used for the analyses are large as they include all age, sex, diagnoses, practice indicators and patient outcomes for each patient in a large number of practices. Running the models is therefore computationally intensive. A final objective is to explore methods for appropriately dealing with the challenging methodological issues while producing results that can be communicated easily to policy makers, clinicians, and other healthcare professionals.

Chapter 2 presents a literature review covering case-mix measurement systems and their applications, comparisons of general practice resource use in the UK, and a critique of development and applications of the ACG system. In chapter 3 the statistical methods used in applications of the ACG software in this area are critiqued; statistical methods previously used for case-mix adjustment in primary care are reviewed; and statistical issues arising in the course of this research are discussed. The process of converting the Oxmis and Read codes to ICD9 codes and constructing clinical case-mix measures (ADGs, ACGs and RUBs) is explained. Chapters 4, 5 and 6 examine how the ACG Case-Mix System was used to explain variation in home visit, referral and prescribing patterns in general practice and, in the case of home visits, social class is also examined. These chapters investigate whether more variability in these outcomes between general

practices can be explained by using the Johns Hopkins ACG Case-Mix System than the traditional age and sex methods using detailed general practice data from the MSGP4 survey and the GPRD. The predictive ability of the models is also investigated. Chapter 7 summarises work done, offers conclusions that are far reaching and provides recommendations for further work.

# Chapter 2    Literature Review

## 2.1   Introduction

The following section describes the relevant literature on variation in general practice in the UK. The focus is largely on research carried out prior to 2005, since the bulk of this research was done prior to that year. An in-depth review of the literature was conducted and goes far beyond what is recorded here. However, for the purpose of this work, a subset covering the most relevant literature is summarised here. Much of the background literature focuses on cost-related outcomes and hence has been excluded since the outcomes in this study are related to service activity.

The motivation for investigating how well diagnostic based case-mix can explain some of these variations is explained. The main systems for measuring diagnostic based case-mix are introduced together with the rationale for using the Johns Hopkins ACG Case-Mix System. Examples of how this system has been used in the UK and internationally for examining variations in general practice patterns are illustrated.

## 2.2 Variation in general practice patterns in the UK: Motivation for using diagnostic based case-mix system

Many studies highlight examples of variations in patterns of activity and resource use in general practice the UK (Carr-Hill et al (1996) & (2002); Aylin et al (1996); Reid, F et al (1999); O'Donnell (2000); Hippisley-Cox (1997); Carlisle (1998); Parry (1998)).

Davis P et al have an extensive body of research into variation in practice patterns in New Zealand. Several of their papers are based on a survey representing a 1% sample of GP visits (about 10,000 visits) at two points in time. Patient, diagnostic and doctor variables are controlled for in a study investigating prescribing patterns and the

conclusion reached is that these improve the predictive power of the model, but do not reduce the extent of variability between doctors in prescribing (Davis P et al, 1995). Further research by Davis et al (2000) explores economic vs health services research theories on variation in medical practice where health economists stress the influence of income incentives while health services research emphasise clinical ambiguity in doctor's decisions. The "supply hypothesis" incorporates both theories by positing both doctor and practice attributes as influencing clinical decisions. Income incentives, doctor agency and clinical ambiguity (measured as local doctor density, practitioner encounter initiation and diagnostic uncertainty respectively) were examined in relation to prescribing, test ordering and doctor request for follow-up. They found no relationship between competition and decision making; that doctor initiated follow up consultations were associated with lower rates of intervention, and that diagnostic uncertainty is associated with higher investigations and follow-up. They concluded that, for the variables studied, a clinical, rather than economic, model of doctor decision-making provided a more plausible interpretation of variation in rates of clinical activity in general practice. In contrast, in applying similar multilevel statistical techniques, Scott and Shiell (1997) found that GPs in areas of high competition were more likely to recommend a follow-up consultation than those in low competition areas for one out of the four medical conditions they analysed.

Davis et al's 2002 paper extends the above study to investigate the variability between doctors in their clinical activity, again measured as prescribing, ordering of investigations and doctor-initiated follow-up (Davis P et al, 2002). They found large variation between doctors in each of these measures, even after adjusting for case-mix, patient and practitioner factors. These factors explained from 15% to 29% of the total variance in the three outcomes, however, investigation of the components of variance concluded that only from 4% to 11% of the remaining unexplained variability was at the doctor level. The work then focussed on one diagnosis only: upper respiratory tract infection. For this diagnosis, they found that the proportion of total variance explained by the model decreased, although the doctor level residual variance increased. This paper has important parallels with the work of this thesis, as explained in relation to the work on referrals in Section 5.5.

Such activity was originally presented as crude measures when reporting variation in general practice patterns. The raw numbers are useful for understanding the overall burden of activity, for example, how many events occur in which groups of individuals (Sevcik AE et al, 2004). However for comparisons between practices these raw numbers are not always a fair representation because of the unequal distribution of patient characteristics across general practice populations. Age and sex are examples of patient characteristics that have long been recognised as confounding factors, for example, a practice with a higher proportion of older patients is likely to have higher than average referrals to specialist care. Similarly, females tend to be referred more than males. Since the 1980s, many studies comparing general practice populations have adjusted for age and sex to allow for a fairer comparison between practices (Reid F. et al (1999); Shenkman et al (2001)) and they remain a commonly used method of adjustment when benchmarking general practices (www.nhscomparators.nhs.uk).

Age, sex and survey based measures have generally been found to explain only a small proportion of the variation between practices. For example, Aylin (1996) compared age-sex standardised rates of home visits among practices and found an almost eight-fold variation. O'Donnell's (2000) literature review on variation in GP referral rates found that UK studies generally reported three to four fold variation in referral rates between practices (Crombie DL and Fleming D (1988); Noone A et al (1989); Wilkin D et al (1992)). Reid (1999) found a crude variation in overall hospital admission rates of 10 to 30 per 100 patients per annum and the findings were similar after indirect standardisation for age and sex.

Survey and census measures such as individuals' health perception, functional status/disability, self-reported clinical diagnoses and chronic disease risk are often used as measures of case-mix (Fowles et al, 1996; Dunn et al, 1996; McCormick A et al, 1995; ONS website neighbourhood statistics: www.neighbourhood.statistics.gov.uk). An important limitation of these measures is that they are subjective, depending on the individual. It is time-consuming and expensive to collect such measures for large populations.

Carr-Hill RA et al, 1996 examined GP consultation rates in general practice using census small area statistics to investigate associations with socioeconomic characteristics and health status (the latter in the form of whether or not a patient was registered as permanently sick). Carr-Hill concluded that demographic and socioeconomic factors can be powerful predictors of consultation patterns and advocated using these results in developing a resource allocation formula for general practice.

Hippisley-Cox et al (1997) reported a significant association between deprivation (Jarman score) and referral rates, and that deprivation explained 23% of the overall variation in referral rates among GP practices. Hull et al (1998) studied 63,000 adult attendances at A&E to investigate their association with practice characteristics and factors relating to deprivation. Results suggested that deprivation accounted for almost half of variation in attendance rates between practices. Attendance rates by patients from two apparently similar practices (both underprivileged and similar distances from nearest hospital) serving populations from the same ward were significantly different, even though the proportions of patients admitted and referred on to outpatients were similar. These results suggest that case-mix and severity vary between apparently similar practice populations. Similarly, Carlisle R et al (1998) found more than three-fold variation between electoral wards in UK out of hour's attendance rates for both general practice and A&E where both served populations from the same wards. Deprivation (Jarman index) accounted for 58% of this variation. Scores developed to assess deprivation, such as the Jarman score, have been criticised as they were originally constructed to measure workload rather than deprivation (O'Donnell (2000)).

Overall, the findings in the medical literature are that wide variation has been shown to exist for outcomes such as home visits, hospital admissions, referrals, A&E attendance and prescribing rates, even after adjustment for various measures of case-mix such as demographics or census and survey based measures of health and socio-economic status (Carr-Hill and Sheldon (1992); Majeed et al (2001)).

Although it is believed that social class might explain a large proportion of this variation, such a measure is not widely available in the US (Krieger N et al (1994)). Since healthcare providers in the US routinely record diagnostic codes on insurance claims forms (mainly to avoid refusal or delay of payment (Wrightson CW (2002)), researchers have been able to use this information in developing diagnostic-based case-mix measurement systems.  These systems were originally designed for adjustment of capitated payments to health plans (Kronick et al, 2000).  Many studies in the US and Canada have suggested a markedly reduced variation in resource use between general practices after adjusting for diagnostic-based case-mix compared with adjustment with age and sex (Starfield et al (1991); Weiner et al (1991); Salem-Schatz, S. et al (1994); Reid, R. et al (1999); Averill, R.F. et al (1999)).  The value of using a diagnostic based case-mix system to explore variability between general practices in the UK is not known.  The results of this review present a strong argument for adjusting for diagnostic based case-mix when comparing variation in general practice outcomes in the UK (Hull (1998); Carlisle(1998)).

## 2.3 Case-Mix systems

Diagnostic based case-mix adjustment systems, also known as risk adjustment systems, have many different applications.  Some of the main uses are:

- explaining variation between practices

- profiling health service use and practice patterns

- fairer allocation of funding to providers

- quality assurance and outcomes management

- identification of the need for case management

- identifying opportunities for disease management

Some of the main diagnostic case-mix systems publicly available for use in primary care are outlined below.

### 2.3.1   Diagnosis Related Group System

The Diagnosis Related Group (DRG) (Kahn et al, (1990)) system classifies hospital cases into one of about 500 groups expected to have similar hospital resource use.  It was developed for patients of the Medicare Inpatient Payment System by researchers at Yale University in the late 1960s. The aim was to create a tool to help monitor quality of care and service use in hospitals in the US.  They are now used mainly for costing and resource allocation and payment (Sanderson, 1998).  DRGs are assigned based on ICD diagnoses, procedures, age, sex, and the presence of complications or comorbidities.  Patient episodes are allocated, based on the primary diagnosis, to a major diagnostic category (MDC), which corresponds to the body systems.  Within the MDC the episode is allocated to either a surgical DRG or a medical DRG and can be further divided into high or low cost group using age (usually above or below 70 years) or presence of more complicating or co-morbid secondary diagnoses.  The DRG system was implemented in a prospective budget control system by the New Jersey State Department of Health in the US.  DRGs have been used since 1983 to determine how much Medicare pays the hospital, since patients within each category are similar clinically and are expected to use the same level of hospital resources.  The DRG system is modified annually to respond to changing patterns of care and diseases. Earlier versions of DRGs related hospital case-mix with costs arising from resource use and demand, not accounting for important factors such as severity of illness, greater treatment difficulty and poorer prognoses.  Refinements have led to several distinct DRG systems (e.g. HCFA-DRGs, AP-DRGs) to allow for different applications of the system.

England began testing the use of DRGs in hospital settings in 1982, and these were more systematically applied in the Resource Management Programme from 1988.  The DRGs were modified to make them more clinically meaningful for English hospital practice.   The modified versions are known as Healthcare Resource Groups (HRGs) and these were first released in 1992.

### 2.3.2   Johns Hopkins Adjusted Clinical Groups (ACG) Case-Mix System

The Johns Hopkins ACG Case-Mix System is a tool used to characterise the degree of overall morbidity in patients and populations http://www.acg.jhsph.edu.  In the 1980s, Barbara Starfield and colleagues produced a body of research evidence to suggest that clustering of morbidity is a better predictor of health service resource use than the presence of specific diseases (Starfield et al (1985)).  The ACG system was subsequently developed during the 1980s at the Johns Hopkins University in order to incorporate each patient's cluster of diagnoses into a measure of case-mix that could be used in the study of primary care populations (Starfield at al, 1991).  The original ACG system was released in 1990 (Weiner et al (1991)).  It was initially developed specifically for primary care use, hence the original name of Ambulatory Care Groups (Weiner et al (1991)).  More recently, it has been expanded to include hospital inpatient information and renamed as Adjusted Clinical Groups.

### 2.3.3   Chronic Illness & Disability Payment System

The Disability Payment System (DPS) was developed by Richard Kronick and colleagues at the University of California, San Diego in 1996 (Kronick et al (1996)). The aim was to fairly compensate health plans that serve people with disabilities or residents of low-income areas.  The DPS is currently used in the US in predicting expenditures for disabled Medicaid beneficiaries.  Some US state Medicaid programs also use the DPS to provide financial incentives for health programs to provide appropriate services for those with disabilities.  The Chronic Illness & Disability Payment System (Kronick R et al, 2000) http://cdps.ucsd.edu/ was developed in 1999 as a revision of the Disability Payment System to make the system more complete and more effective in its adjustment of payments for the 'Temporary Assistance to Needy Families' population.  Most of the diagnoses are not disabilities but diagnoses of disease – some very serious and many others, e.g., migraines or uncomplicated adult-onset diabetes, that are unlikely to be disabling conditions.  The name was changed to include chronic illness as the previous name gave a mistaken impression that the system could only be used for disabled patients.  It has been further adapted to produce CDPS-

Medicare, a model for use in adjusting capitated Medicare payments to health plans (Kronick R et al, 2002).

### 2.3.4   Clinical Risk Groups

Clinical Risk Groups (CRGs) were previously known as the Classification of Congenital and Chronic Health Conditions. CRGs were developed by the National Association of Children's Hospitals and Related Institutions (NACHRI) and 3M Health Information Systems (Salt Lake City, Utah) to describe the health status of those enrolled in Managed Care Organisations and to predict future use of services. The development of CRGs (Muldoon et al (1997); Averill et al (1999)) was influenced by the use of DRGs described previously. People with chronic illness are likely to have a high dependence on resource use and so the CRG system was designed to provide a classification system for these individuals. The system was released for public use in 2000. Each individual with a chronic health condition is assigned to a single mutually exclusive risk category based on a combination of their most significant chronic disease for each organ system being treated and the severity of illness of their most significant chronic disease. All medical services for an individual are classified over an extended period of time. Each grouping is intended to be clinically meaningful and to provide the basis for the prediction of future health care utilisation and cost. CRGs have been evaluated and validated with historical data (Muldoon et al, 1997; Averill et al, 1999)

### 2.3.5   Diagnostic Cost Groups

Diagnostic Cost Groups (DCGs) were developed by Arlene Ash and colleagues at Boston University (Pope GC et al (2000)). Original research began in 1984 and was based only on inpatient hospitalisation information. Ten years on, they were expanded to also include practice information. DCGs classify individuals into groups based on the diagnosis with the highest cost for each patient. The Washington State Health Care Authority applies the DCG system for prospectively risk adjusting its payments (Iezzoni LI et al, 1998). There are several developments of DCGs for specific purposes. Two of these are the Principal Inpatient DCG and the All-Diagnoses DCG. Principal Inpatient DCG classifies people by their single highest cost principal inpatient diagnosis. All-Diagnoses DCG adds secondary inpatient, hospital outpatient, and doctor diagnoses to

the principal inpatient diagnosis, and classifies people by their single highest predicted cost diagnosis.

## 2.4 Motivation for using Johns Hopkins ACG Case-Mix System

Each of the systems for measuring case-mix based on patient diagnoses was developed using different patient populations and each with a somewhat different emphasis. The result is that there are many differences between them (Hornbrook et al (1996); Shenkman et al (2001); Cumming et al (2002)).

The choice of case-mix system depends to a certain extent on the situation in which the case-mix measure will be applied. For example, the Clinical Risk Groups only classify patients with chronic illness and the DPS only classify patients with chronic illness or disability, the CDPS focuses primarily on Medicaid populations, especially 'Temporary Assistance to Needy Families' and disabled Medicaid beneficiaries, while the ACGs and DRGs classify all patients. Most of the systems, for example, the Diagnostic Risk Groups, work best for investigating past resource use, as they are assigned retrospectively. In contrast to this, the Clinical Risk Groups were designed specifically to predict future use and so should be used in these situations.

Other important considerations when choosing a case-mix system are how well the system can predict resource use, how simple the system is to implement and administrate, and how resistant the system is to manipulation. Several studies have compared these and other considerations for various case-mix adjustment methods (Dunn et al (1996); Fowles et al (1996)). Dunn et al (1996) compared the age-sex, ACG and DCG adjustment methods for various criteria and found that all of the diagnostic-based methods were a substantial improvement on the age-sex model for predictive accuracy.

The case-mix systems were initially developed as a way of coping with rising healthcare costs (Sanderson et al (1998)) by adjusting capitated payments to health plans (Kronick et al, 2000). Most of the systems were first developed and validated for use in hospital

settings (Shenkman et al (2001)) only.  An important feature of the Johns Hopkins ACG Case-Mix System is that it was developed specifically for the primary care setting using primary care data (although has since been expanded to include hospital inpatient data). The other systems were later adapted for use in primary care (DRGs (Kahn et al, 1990); CDPS (Kronick et al, 1996); CRGs: (Muldoon et al, 1997; Averill et al, 1999); DCGs: (Pope GC et al, 2000).

Fowles compared three different health status measures with standard demographic adjustment (Fowles et al (1996).  The adjustment factors considered were self-reported functional health status, self-reported chronic diseases and the ACG groupings.  Her findings suggested that ACGs performed best of all, while self-reported health status predicted expenditures twice as well as demographic measures.  Fowles concluded by recommending the use of case-mix adjustment methods based on diagnostic information where possible when selection bias is suspected.  In the absence of diagnostic information, she recommended employing a system using simple self-reported measures, such as the presence or absence of chronic conditions, rather than complex functional status measures or standard demographic adjustment.  One main advantage of using ACG measures derived from patient diagnoses over self-reported measures is that the former are not subject to the response bias and recall bias that is often present with self-reported measures for various reasons such as illiteracy, illness, language barriers and memory failure.

The grouping mechanisms of the various clinical case-mix adjustment systems differ. For example, Diagnosis Related Groups classify a single encounter at one point in time (e.g. hospitalisation); Clinical Risk Groups only classify individuals with congenital and chronic health conditions or significant acute conditions; DCGs classify individuals based on the diagnosis with the highest cost; while the ACGs classify all diagnoses for each individual over an extended period of time.   The ACG grouping mechanism is described in detail in Chapter 3.  It has some similarity to that used to assign hospital patients to Diagnostic Related Groups in the USA and Healthcare Resource Groups in the UK.  However, the unique feature of the ACG groupings is that ACGs make use of all the diagnoses in the patient's medical history during a specified period of time,

usually a one year period, and not just the diagnoses recorded from a single episode of hospital care.  As a result of this, an individual might be placed in a higher risk group if classified with an ACG than if classified by a DRG or DCG because all ambulatory care diagnoses are taken into account in assigning the grouping.

The fundamental difference between ACGs and other case-mix systems is that ACGs measure every patients overall morbidity as this has been shown to be a better predictor of health services resource use than examining only specific diseases (Starfield et al, 1985).  Fleming's 1991 paper stated that the 'analysis and interpretation of data from general practice should preferably be based on the person as the unit of analysis'.  This is one of the most compelling features of the ACG system and the main reason why this system was chosen over others for this research.  The transparency of the ACG grouping mechanism means that it can be adapted to suit the needs of the UK health care system.

## 2.5 UK use of ACG system

The ACG system has been used in the US, Canada and other countries such as Sweden, Spain, Australia and New Zealand for various applications such as provider profiling. Application of ACGs in the UK has been fairly limited to date.  The first published study applying ACGs in the UK was a feasibility study (Majeed et al, 2001).  The ACG System was applied to data from the Morbidity Statistics in General Practice (MSGP4), a 1% sample of the population of England and Wales.  Results were compared with populations from two large insurance plans in the US.  Distribution of ADGs was found to be similar to the US plans, although the US populations had a higher percentage of those with higher recorded levels of morbidity (>=5 ADGs and >=3 major ADGs).  The authors suggested that this might reflect differences in medical practice, information lost in translation of Read to ICD-9 codes, or more complete recording of diagnostic data in the US. This study demonstrated that the ACG system may work reasonably well in the UK and that further research was necessary.  A second study used the ACG system to control for case-mix in a comparison of variation in US and UK referral rates (Forrest CB et al, 2002). Patients were assigned to morbidity groups, with higher scores indicating higher morbidity and greater need for referral.  The percentage of patients

with one or more referral per year was 13.9% in the UK compared to 31.6% in the US. This research showed that UK referral rates were lower than the US regardless of morbidity burden, and the authors concluded that the large difference in primary care referral patterns between the two countries is most likely due to the large difference in supply of specialists.

Three papers based on Chapters 4, 5 and 6 of this thesis were then published in peer reviewed journals and this and other related work carried out during the course of this thesis (some that is outside the scope of this final document) has been presented at conferences and seminars in the UK and abroad.  Section 7.5 includes a list of relevant publications and selected presentations.  UK general practice Read codes have been integrated into later versions of the ACG System, although were not available within the tool when this research was undertaken.

Kinder-Siemens et al (2007) presented their findings at the 23rd Patient Classification Systems International Conference.  They used the ACG system to investigate population risk profiling, provider performance profiling and patient identification.  A strong relationship was found between risk and resource use with differences in risk distribution across geographical areas.  For performance profiling and allocation of budgets, they compared actual and expected resource use.  In identifying people at risk for care planning, they found that, of the outcomes they examined, total secondary care costs were best explained by ACG measures.  They found that for primary care outcomes, pharmacy use and lab tests had high explanatory power, the former a similar result to the findings in this thesis (Omar RZ and O'Sullivan C (joint authors) et al (2005)).

Since the work from this thesis was published, the ACG system is being piloted in several Primary Care Trusts for risk stratification and risk adjustment (www.acg.jhsph.org).  Just prior to publishing this thesis, some of the latest work involving the ACG system was presented at the 4th Johns Hopkins University's London Symposium on Case-Mix and highlight the growing use of this tool in the UK.  The

findings presented at this conference are to be made available on the ACG website (www.acg.jhsph.org).

## 2.6 International use of ACG system

The ACG system is widely used internationally, particularly in the United States and Canada.  One of the main applications of ACGs in the US and Canada is on pricing and risk-adjusting capitation rates, as these countries have comprehensive cost data available to them at primary care level.  The ACG website includes an extensive bibliography covering risk adjustment, performance profiling and other applications (http://www.acg.jhsph.org/public-docs/AcgBibliography.pdf).

ACGs are used in implementing risk adjustment payments made by the Minneapolis Buyers Health Care Action Group (Knutson D, 1998) and the Maryland and Minnesota State Medicaid programs (Wrightson CW, 2002).  British Columbia has been using the ACG System since 2000, primarily for practitioner profiling as part of a larger program of keeping the doctors accountable for fee for service (Reid RJ et al, 2001, 2002). The ACG system is used to adjust for different expected amount of costs for doctors' medical care based on the burden of illness they have in their patient population.  Reid RJ evaluated the use of ACGs for measuring morbidity in populations in Manitoba, Canada (Reid R et al, 1999; 2002) and found a strong relationship with ACG morbidity and subsequent rate of premature death.  The ACGs were found to explain most of the relationships between premature mortality and both socioeconomic status and doctor use.

Spain has been researching applying the ACG System since the 1990s (Bolanos-Carmona V et al, 2002; Juncosa S et al, 1996, 1997, 1999; Orueta JF et al, 1999). Original research included examining the performance of ACGs in various settings. Orueta JF et al, (1999) examined the performance of ACGs compared to a US health plan in a cross sectional study from primary health care centres in the Basque Health Service.  Orueta found ADGs and ACGs were a considerable improvement over age and sex for estimating doctor's workload.  Juncosa (1999) applied ACGs in an observational

study of a group of 13 primary care doctor and nurse teams in a region in Spain, following a random sample of about 2500 patients for a mean of six months. Performance of the ACG groups was found to be acceptable and results were not very different from the results obtained during original validation work of the ACG groups by the authors of the ACG System on the Columbia Medical Plan population (Weiner JP et al, 1991). Spain's plans for wider distribution of the ACG System are detailed on the ACG website (www.acg.jhsph.org).

The ACG system has been applied in parts of Sweden for some years now (Carlsson L et al (2002, 2004, 2005); Halling A et al (2006); Zielinski A et al (2009)). Carlsson's research reinforces the original ACG research ((Starfield et al (1985); Starfield et al (1991); Weiner et al (1991)), challenging traditional epidemiological approaches based on statistics on conditions, for example, diabetes, rather than the complex health status of certain patients in terms of their comorbidities (Carlsson L et al (2002, 2004, 2005).

As Swedish health care generally has no information on individual patient costs, other approaches have been used. Much of the variation in polypharmacy, as a proxy for health care costs, in an elderly population was shown to be explained by the system (Halling A et al (2006)). Several studies based on costs data (Carlsson L (2004), Engstrom SG (2006) and Zielinski et al (2009)) have found the ACG case-mix system explains much of the variance in Swedish primary health care costs in centres/regions of Sweden such as Blenkinge (Zielinski et al (2009) and is therefore one factor that can enable equitable health care in Swedish primary health care.

Franks et al (1999) examined variations in doctor referrals and adjusted for age, sex and Ambulatory Diagnostic Groups (ADGs). This study is based on a large sample size from a large Managed Care Organisation in Rochester, New York. As in this research, multilevel statistical techniques were used and the referral outcome was defined as at least one visit to a specialist. They used general linear mixed models rather than logistic mixed models, because of computer hardware limitations and also the difficulty in interpretation of the logistic results (They overcame this and were able to use logistic regression in a Generalised Estimating Equation (GEE) approach to adjust for clustering

of observations within doctor in a later study which also adjusted for further doctor practice and psychological factors (Franks et al, 2000).)  The 1999 study showed that ACG case-mix adjustment produced relatively little change on the range of referral rates.  The observed and case-mix-adjusted referral rates were moderately correlated with a number of the doctor level variables (older doctor, internists, solo practitioners, doctors practicing longer, and longer in their current practice, those with more sessions per week, and doctors with higher Herfindahl indices (a measure of doctors' experience in specific diagnostic areas) all referred more). Of the psychological variables, only risk aversion was found to be associated with a higher referral rate.  Franks work concluded that the variation in referrals was only minimally affected by adjustment for diagnostic-based patient case-mix (ADGs).  They found that the doctor component accounted for 93 percent of the variation in referrals, while the patient component only account for 6 percent.  However, Franks also adjusted for ACGs in a study of total patient expenditures and found that this explained over 60% of the variance in expenditures. The reverse was found in a Spanish study (Bolanus-Carmona et al 2002) where the R-squared was 19 percent, and inclusion of doctor factors increased this by 15 percent. This study's findings imply that referral rates are largely a doctor-driven behaviour that is relatively stable over time and can be generalised across different diagnostic categories.

Salem-Schatz et al (1996) investigated the influence of patient factors when comparing referral rates in a cohort of about 38,000 patients from 52 practices in a large Health Maintenance Organisation over one year.  Comparisons were made on the impact of adjusting for age and sex as opposed to adjusting for diagnostic based case-mix and her findings advocate the use of case-mix adjustment over age-sex when profiling practices.

A German sickness fund that reimburses healthcare providers has been piloting outpatient management of patients with psychosocial disease clusters in order to avoid expensive inpatient care and is exploring similar applications for conditions such as diabetes and hypertension (www.acg.jhspu.org).

Israel's largest health fund has recently begun to apply ACGs in several studies.  One of these was to examine the differences in healthcare use between those with high and low socioeconomic status, controlling for age, sex, and morbidity using both ACGs and the Charlson Co-morbidity Index. Age and sex adjustment showed a positive association with low socioeconomic class and diagnostic tests as well as specialty care use but this was shown to be inaccurate when morbidity was examined.  They compared ACGs with the Charlson Index and found that ACGs performed better.  The above demonstrates that ACGs are becoming more widely used for a variety of applications in healthcare, including variation in general practice resource use.

## 2.7 Thesis aims and objectives

This thesis presents the first large-scale studies in the UK to adjust for diagnostic-based morbidity when examining variation in general practice.  The literature review revealed evidence of large variations in resource use measures such as consultations, referrals and prescribing practice patterns in general practice in the UK and elsewhere.  A brief overview of case-mix and the reasons why case-mix systems based on patient diagnoses have been developed were presented.  The Johns Hopkins ACG Case-Mix System is selected since each patient's overall morbidity has been shown to be a better predictor of health service resource use than other measures, for example, measures based on specific diseases.  Also, it was originally designed specifically for use in primary care settings.  This thesis attempts to quantify the relative contribution of diagnostic based patient morbidity in explaining variation in important general practice outcomes.  Three outcomes with widely documented evidence of large variations between general practices were selected for the purpose of this work: home visits, outpatient referrals and prescribing patterns.

Large and complex datasets containing detailed patient demographic and diagnostic information from the Morbidity Statistics in General Practice Survey (MSGP4) and the General Practice Research Database (GPRD) will be used for the purpose of this research.  The methods that others have used for examining variation in general practice outcomes and for identifying factors that might explain the variation will be examined.

Methods for appropriately dealing with the challenging methodological issues (E.g. the clustered nature of general practice data; measuring variation in binary outcomes) will be explored with the aim of producing results that can be communicated easily to policy makers, clinicians and other healthcare professionals.

The aims and objectives forming the basis of the work from this thesis are set out below.

### 2.7.1   Aims

The overall aim of the work presented in this thesis is to investigate whether variation in home visit, referral and prescribing between UK general practices may be explained by patient level diagnostic-based morbidity measures.

### 2.7.2   Objectives

Objective 1: To review the literature on variation in general practice workload outcomes and on measures that may explain variability in these outcomes.

Objective 2: To investigate whether patient level measures of morbidity, assigned using the Johns Hopkins ACG Case-Mix System, explain more of the variation in home visits, referrals and prescribing patterns between general practices than age/sex and social class measures.

Objective 3: To explore methods for appropriately dealing with methodological issues arising from the clustered structure of general practice data and produce results that can be communicated easily to policy makers, clinicians and other healthcare professionals

## 2.8 Conclusions

In this chapter, the concepts of case-mix and adjusting for case-mix have been presented.  The various systems that have been developed for measuring case-mix and the main differences between the systems for measuring case-mix have been outlined.

Examples of the use of the ACG system in the UK and internationally were given. The main aims, objectives and research questions of the thesis are set out. Chapter 3 will critique the statistical methods used in the development and applications of the ACG software. This chapter will also review statistical methods previously used for case-mix adjustment in primary care settings and discuss the statistical issues that arise in the course of this research.

# Chapter 3    Statistical Methods

## 3.1 Introduction

Chapter 3 summarises how the grouping mechanism for the Johns Hopkins Adjusted Clinical Groups (ACG) System works, and describes the methods used in development and validation of the components of the ACG system (i.e. ADGs, ACGs and RUBs). The features of the general practice data used are discussed and the most common statistical methods used in the literature in studies examining variation between general practices are examined. The motivation for using multilevel modelling techniques is presented together with the measures explored in order to quantify the variability explained.

## 3.2 Johns Hopkins ACG Case-Mix System

The motivation for using the Johns Hopkins ACG System was explained in Section **2.4**. In this section the grouping of diagnoses into ADGs and ACGs is described.

### 3.2.1   The ACG Grouping mechanism

Figure 1 is a pictorial overview of how the ACG system assigns diagnosis codes first to Aggregated Diagnostic Groups (ADGs), then to Adjusted Clinical Groups (ACGs) and finally to Resource Utilisation Bands (RUBs).

**Figure 1  Illustrated summary of the ACG assignment process**

Diagnosis codes

32 morbidity groupings called ADGs

Each ADG is composed of ICD9 codes of similar expected healthcare resource use

Age, Sex

Adjusted Clinical Groups (ACGs)

Each individual is assigned to an ACG based on his/her combination of ADGs, age & sex

Resource Utilisation Bands (RUBs)
Groupings of ACGs of similar expected resource utilisation
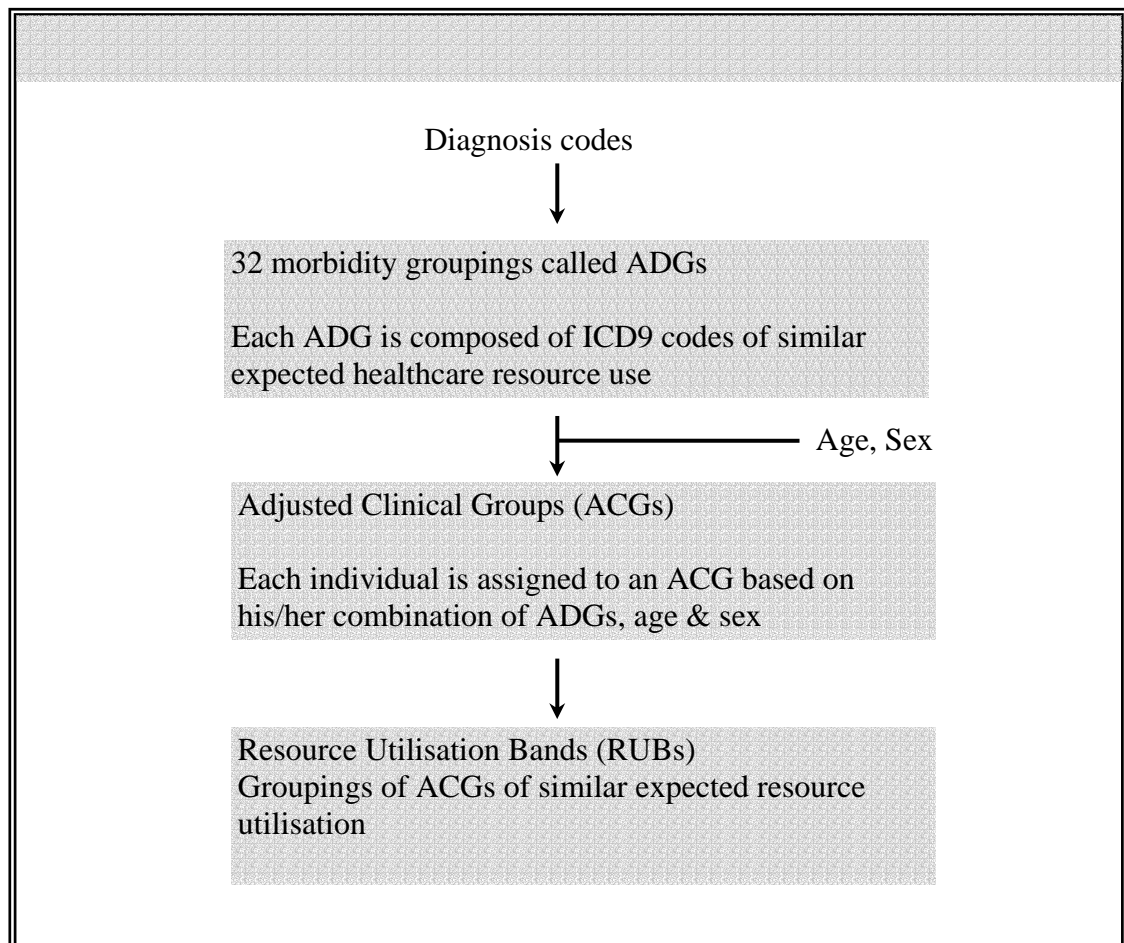
Source: www.acg.jhsph.edu

### 3.2.2   Diagnosis groups (ADGs)

Firstly, every diagnosis of every individual is classified from an ICD-9 code into one of

32 diagnosis clusters known as Aggregated Diagnosis Groups or ADGs.  Conditions are

clustered together based on their expected impact on health service resource

consumption (Starfield et al, 1991).  Conditions are assigned to one of the 32 diagnostic

groups (ADGs) according to several clinical criteria.  The clinical criteria are as follows

(www.acg.jhsph.edu):

- Duration of the condition (acute, recurrent, or chronic): *How long will healthcare resources be required for the management of this condition?*

- Severity of the condition (e.g. minor/stable vs major/unstable): *How intensely must healthcare resources be applied to manage the condition?*

- Diagnostic certainty (symptoms vs diseases): *Will a diagnostic evaluation be needed (symptoms) or will services for treatment be the primary focus (diseases/diagnoses)?*

- Aetiology of the condition (infectious, injury, or other): *What types of healthcare services will likely be used?*

- Specialty care involvement (medical, surgical, obstetric, haematology etc.): *To what degree will specialty care services be required?*

It is possible to categorise any clinical condition into an ADG grouping using the above criteria. Thus, many of the ADGs include conditions that appear unrelated but are thought to have similar future resource use. For example, cerebral thrombosis and acute pancreatitis are both in the same ADG grouping for progressive conditions likely to recur. Discrete conditions that are likely to recur, unstable chronic medical conditions, and time-limited minor psychosocial conditions are three separate ADG groupings. Several of the ADGs are more specific, for example, there are different groups for asthma, dermatologic conditions, malignancy, and pregnancy. Some groups are defined based on a combination of resource expectation and type of condition, for example, stable orthopaedic conditions, unstable eye conditions, unstable recurrent, and persistent psychosocial conditions.

Patients are assigned to an ADG if they have one of more of the diagnoses that make up that ADG. A patient can have any number of ADGs, from no ADGs up to 32 different ADGs. For example, a patient with both Obstructive Chronic Bronchitis and Congestive Heart Failure will be grouped into only ADG 11 (Chronic Medical: Unstable), whereas a patient with Candidiasis and Acute Upper Respiratory Infections will have two ADGs, ADG 8 (Likely to Recur: Discrete) and ADG 2 (Time-Limited: Minor-Primary Infections) respectively.

### 3.2.3   Adjusted Clinical Groups (ACGs)

In order to further assign each individual to a mutually exclusive ACG, the ADGs described above are collapsed into 12 categories known as Collapsed ADGs (CADGs) according to the following three clinical criteria: Similarity of likelihood of persistence or recurrence of diagnoses within the ADG, i.e., time-limited, likely to recur, or chronic groupings. Severity of the condition is used as a basis for additional categories, i.e., minor versus major and stable versus unstable. Some further CADGs were formed because of the types of healthcare services required for patient management--medical versus specialty, eye/dental, psycho-social, prevention/administrative, and pregnancy.

The 23 most frequently occurring combinations of CADGs among patients form MAC groups (originally known as Major Ambulatory Categories).  These are the main branches of the ACG decision tree, with a last branch (MAC 24) for those with multiple co-morbidities that cannot be classified elsewhere.  MAC-25 is for patients who either have not consulted, or have invalid diagnosis data. The final MAC-26 includes all infants (age<12 months) (ignoring their pattern of CADGs).

The final step in the process of assigning ACGs involved a combination of statistical methodology and clinical judgement.  AUTOGRP software from Yale University was used to sub-divide the MACs into one of about 90 mutually exclusive ACG categories with similar needs for healthcare resources based on their overall expenditures. The variables used were: age, sex, presence of specific ADGs, number of major ADGs, and total number of ADGs (excluding ADG 31 since it only contains preventative and administrative codes and therefore does not reflect morbidity).  Individuals of the same age with diagnoses in several ADGs may be classified into different ACGs depending on how many of their ADGs are considered major.  For example, a major ADG for adults is for progressive, likely-to-recur conditions, while the ADG for allergies is minor.

**Table 1  Examples of aggregated diagnosis groups (ADGs) and adjusted clinical groups (ACGs)**

|  |  |
|---|---|
| | MINOR AGGREGATED DIAGNOSIS GROUPS |
| ADG 2 | Time limited: infection |
| ADG 26 | Symptoms and signs: minor |
| ADG 31 | Preventive and administrative |
| | |
| | MAJOR AGGREGATED DIAGNOSIS GROUPS |
| ADG 11 | Chronic medical: unstable |
| ADG 16 | Chronic specialty unstable: orthopaedic |
| ADG 22 | Injuries and adverse effects: major |
| | |
| | ADJUSTED CLINICAL GROUPS |
| ACG 0300 | Acute minor, age < 16 years |
| ACG 2200 | Acute minor and likely to recur, age > 5 years, no allergy |
| ACG 3800 | 2-3 diagnostic group combinations, age > 34 years |

Source: Majeed et al *BMJ* 2001;323:607–10

### 3.2.4   Resource Utilisation Bands (RUBs)

A further collapsing of ACG groups into a smaller number of groups (usually six to eight) known as Resource Utilisation Bands (RUBs) is also possible, and is useful for analysis purposes. These cover morbidity groups ranging from healthiest through to sickest patients, for example: non-users, healthy users, low morbidity, moderate morbidity, high morbidity, and very high morbidity.  The user can either create his own RUB groupings or can use the RUB grouping algorithm provided by the ACG software (that are based on a US nationally representative database of 2 million <65 years who were enrolled in several US commercial health insurance plans).  The measure of choice of ADG/ACG/RUB depends on how the health measures are to be used.

The following example of two different patients each with a different degree of diabetes mellitus, from the ACG website, illustrates how the corresponding ADGs, ACGs, and RUBs assigned according to morbidity burden.   The patients are similar in age but patient B has a much higher number of diagnosis codes than patient A (see conditions and ICD9 codes).  Patient B is therefore assigned to 10 ADG groups compared to 2 ADG groups for patient A.  Patient B is assigned to an ACG and a RUB group of higher morbidity burden than patient A.

**Table 2  Example of diagnoses and corresponding ACG groups assigned to two patients**

|  | Patient A | Patient B |
|---|---|---|
| Age | 57 | 54 |
| Conditions | Diabetes | Diabetes mellitus, general medical exam, congestive heart failure, thrombophlebitis, contusions and abrasions, non-fungal infections of skin, disease of nail, chest pain, vertiginous syndromes, fibrositismyalgia, respiratory signs/symptoms, and cough |
| ICD9 Codes | 250.02 – diabetes mellitus, type II, w/o complications, uncontrolled IV70 – general medical exam | V70 – general medical exam; 250.02 – diabetes mellitus, type II, w/o complications, uncontrolled; 386.2 - central origin vertigo; 428.0 - congestive heart failure; 453.9 - venous thrombosis NOS; 681.10 - cellulitis, toe NOS; 703.9 - disease of nail NOS; 729.1 - myalgia and myositis NOS; 786.1 – stridor; 786.2 – cough; 786.50 - chest pain NOS; 924.20 - contusion of foot |
| ADGs | ADG10 - Chronic medical: stable; ADG31- Prevention/administrative | ADG0 1 - Time Limited: Minor; ADG0 4 - Time Limited: Major-Primary Infections; ADG0 9 - Likely to Recur: Progressive; ADG 10 - Chronic Medical: Stable; ADG 11 - Chronic Medical: Unstable; ADG 21 - Injuries/Adverse Effects: Minor; ADG 26 - Signs/Symptoms: Minor; ADG 27 - Signs/Symptoms: uncertain; ADG 28 - Signs/Symptoms: Major; ADG 31 - Prevention/Administrative |
| ACG | 0900: chronic medical, stable | 4930(6-9 other ADG combinations, age >34, 3 major ADGs) |
| RUB | 2 | 5 |

Source: www.acg.jhsph.org

## 3.3 Coding and transferability of codes from US to UK

The ACG case-mix system requires one year's worth of high quality demographic and diagnostic data in order to be used.  The system relies on a large degree of compliance from those recording the diagnostic information (e.g. GPs and other healthcare professionals).  Those recording information must comply with the procedures for data recording.  Diagnostic coding inaccuracies and inconsistencies may introduce bias into

the groupings, for example, different doctors might code the same illness differently. However, the impact of such inconsistencies in coding is reduced somewhat as every ICD9 code is categorised into one of 32 ADGs, and therefore usually only a completely different diagnosis will categorise a patient to a different morbidity group.  The version of the ACG Case-Mix System used in this thesis is also subject to the limitations of the International Classification of Disease (ICD) codes themselves (although it has since been updated to also be used with ICD10 codes).  The conversion of READ and Oxmis codes to ICD9 codes is described in detail in the appendix.  Despite the fact that the ICD9 codes are generally not recognised as adequately addressing the coding needs of primary care, they were adopted as the basis of ACGs as it was the only coding scheme universally applied across the US (Weiner et al, 1991).  (More recently the ACG system has been converted to also accept ICD-10 codes.)

As with all case-mix adjustment systems, there is room for the possibility of deliberate 'upcoding' of the data so that patients are deliberately recorded as being sicker than they actually are.  However, the ACG system is less likely to be prone to this for the reasons cited in the previous paragraph (Weiner et al (1991)).  Also, it would be more difficult for a GP to know how such strategies would affect their perceived use of resources after adjustment for their general practice morbidity burden.

## 3.4 Development and validation of the ACG system components

The components of the ACG system were developed by incorporating both clinical criteria and statistical methodology. The construction of the 32 diagnostic groups (ADGs) was based on several clinical criteria (such as expected persistance/recurrence over time, and likelihood that a specialty referral would be required) (Starfield et al, 1991). Five research sites and over 160,000 patient records were used in the initial development and validation of ACGs. The grouping of diagnoses was developed using data from one site (Columbia (Maryland) Medical Plan (CMP), a health maintenance organisation in the US). Data from the CMP site and from four other sites were used to validate the groupings.

Firstly, at CMP, diagnostic codes (ICDA 8 codes) were converted to ICD-9-CM codes, resulting in a list of 4000 different diagnostic codes. The 1981 National Ambulatory Medical Care Survey identified around 1,000 codes as accounting for over 90% of patient visits to US doctors (Lawrence L et al (1981)). Any codes not already on the list were added so that as many codes as possible could be used in the development of the ACG components.

Two doctors independently assigned each ICD code to one of 20 ADGs using the clinical criteria previously mentioned (excluding severity), together with the help of a similar grouping system previously developed for children. Results were compared and instances where there was disagreement were discussed and resolved. For unusual diagnoses, advice was sought from specialists within the Johns Hopkins medical school.

The ADG groupings were validated by analysing about 8000 members of the CMP site who had been plan members for over six years. Four other research sites also participated in the initial validation exercise (three Health Maintainance Organisations and one Medicaid program of the state of Maryland).

Where persistence or resource patterns associated with a particular diagnosis were not similar to those found for other conditions in the same ADG, the diagnosis was

reassigned to a different ADG. Factor analysis was used to assess how independent the ADG groups were from one another and results suggested that they were relatively distinct classes. Four external consultants then critically reviewed the content of the groupings. The criteria for grouping the ADGs were altered to allow separation of more serious (e.g. meningitis, polyarthritis) from less serious conditions (e.g. upper respiratory tract infection, torticollis) and the categories were thus expanded from 20 to 34 (currently reduced to 32).

## 3.5 How ACGs were developed from ADGs

The process of converting ADGs into ACGs was similar to that used in the DRG development process (Kahn et al, 1990). However, the key dependent measure was the number of ambulatory visits a person made during an extended period of time, usually one year (rather than the length of hospital stay measure used in DRG development). Clinical validity of groupings, as with the ADG development, remained an important consideration. The Yale AUTOGRP program was used to identify subgroups of patients with the lowest possible within-group variation in consultation rate. Independent covariates used were age, sex, number of unique ADGs and particular ADGs (Mills R et al, 1976). These were later expanded to adjust for severity of condition.

The ability of the ACG measures to explain variation in practice visits and costs (see dependent variables listed below) for all five sites and the ability to predict future visits and costs for the developmental CMP site were examined using linear regression.

The following dependent variables were assessed:

- Total number of ambulatory visits made in one year
- Number of ambulatory visits to specialists in one year
- Ancillary (e.g. lab and x-ray) charges associated with ambulatory visits
- Total ambulatory charges (including professional fees and ancillary services)
- Overall charges (including both ambulatory and in-patient)

Independent variables included in the models were:

- Age group and sex only

- Age group, sex and various combinations of the case-mix measures

Models were assessed using measures of R-squared and results ranged from 32% to 59% for the ADG and ACG models with visits or ambulatory charges as the dependent variable compared to 3% to 6% for age-sex models.  The R-squared values for predicting the following year were lower, ranging from 18% to 23% for the ADG and ACG models with visits or ambulatory charges as the dependent variable compared to 3% to 5% for age-sex models.

## 3.6 Clustering of patients within practices

Attempting to understand and explain variation in general practice outcomes is not straightforward and there are many examples in the literature where inappropriate or inefficient statistical techniques have been employed (Larsen K & Merlo J, 2005).  An important feature of general practice data is the clustering of patients within general practices. The characteristics of two randomly chosen patients from the same practice are in the long run more likely to bear similarities than those of two randomly chosen patients from different practices since patients in the same practice will be exposed to the same practice policy and may share common neighbourhood and socio-economic characteristics. For such data, observations are no longer independent, an assumption required by the standard statistical methods. A common error is to ignore the clustering and use standard statistical methods. This correlation of patients within general practices needs to be accounted for when modelling variation as it gives rise to variation at two levels, patient level and practice level.   Furthermore, this inherent clustering of the data needs to be handled with appropriate statistical models; otherwise it may provide incorrect statistical inferences and lead to potentially misleading and erroneous conclusions (Omar RZ & Thomson SG, 2000).

# 3.7 Statistical methods used in the literature to explain variation

In Section **2.4**, the motivation for adjusting for case-mix when investigating factors that might explain variation between practices was explained. Researchers have used a variety of different statistical methods to identify factors that explain variability in outcome measures (Goldstein et al (1995); Larsen K & Merlo J (1982)). Some of the most popular methods used are described below:

## 3.7.1 Ratio of observed to expected

In a paper on the use of multilevel models in health research Duncan C et al (1998) demonstrates how relationships that are shown to hold at an aggregate level cannot necessarily be extended to the individual level e.g. high levels of illness are associated with high levels of unemployment at regional level, but it may be that people who are ill in regions with high unemployment are those who are in employment. In order to overcome this, some researchers have disaggregated higher levels to the individual level before analysis. For example, when comparing a set of practices, indirect standardisation is a common method of calculating expected values after adjustment for, say, age and sex (Reid, FDA et al (1999); Starfield et al (1991)). These expected values may then be compared with observed performance, using, for example, a ratio of observed to expected ((O/E) sometimes referred to as an 'efficiency ratio'). Such a measure illustrates how populations such as general practices are performing relative to how they are expected to perform given their age and sex distributions.

## 3.7.2 Coefficient of Variation

The Coefficient of Variation (CV), a ratio of the standard deviation to the mean (Bland M, 1995), is often used in assessing which groups of patient characteristics explain more of the variability between practices (Salem-Schatz S et al, 1994). It has the advantage of being dimensionless and so is useful for comparing between datasets with different units or different means. Estimates of outcome proportions or rates are calculated separately for each practice after adjustment for patient characteristics. The mean and standard deviation of the estimated practice proportions are calculated and used to compute the CV. Adjustments for case-mix resulting in the smallest CV can

usually be interpreted as explaining the most variation.  See Section 4.4 for example calculation of CV.

### 3.7.3   Ordinary Least Squares Regression

Ordinary least squares (OLS) regression is another approach often used and there are examples in the literature where OLS regression is used to explain variability across practices without any adjustment for the clustering of patients within practices (Carlisle R et al, 1998; Reid FDA et al,1999).  When the between variance partition coefficients are small, the multilevel and OLS estimates can be expected to agree reasonably well (Goldstein (2003)).

### 3.7.4   Limitations

Both the ratio of observed to expected values and the coefficient of variation are summary measures summarised by practices and do not utilise the individual patient information fully.  The calculation of these measures can be cumbersome when adjusting for a large number of patient level covariates.

Ordinary least squares regression is a single level analysis.  Applying this to multilevel data such as general practices may give misleading results, particularly with larger sample sizes.  As the sample size increases, parameter estimates of explanatory variables will be biased, standard errors will be underestimated so confidence intervals will be too narrow and associations may be reported as significant when they are not (Hox J (1998); Goldstein (2003)).  As a result of using OLS, even if the fixed coefficients in the model were similar, one would not be able to study any multilevel structures with enough precision (Goldstein (2003)).  Even if practice clustering is small, this does not have a non negligible effect.  Moreover, it is impossible to discern whether variation is due to individual differences between patients or between practices. In turn, this can lead to inappropriate, unfair and misleading ranking of general practices (Merlo, 2001, O Sullivan C et al, 2005).  OLS can also be performed at practice level but the disadvantage of this is that the individual patient level information is again summarised at the practice level. This is not efficient use of information and may lead to finding associations at the population level that may not hold at the individual level, known as the ecological fallacy (Robinson, WS, 1950).

In a letter to the BMJ (1994), Ken MacRae illustrates the ecological fallacy (see **Table 3**
below). When area is the unit of observation there is an association between exposure
and disease. Area 2 shows greater exposure (300/1000) and greater incidence
(300/1000) of the disease than area 1 (100/1000 and 100/1000). However, this
association is not apparent when individuals within the areas are the units of
observation. The same proportions of exposed and non-exposed individuals have the
disease (10% in area 1 and 30% in area 2) in each area.

.

**Table 3 Example illustrating ecological fallacy. Relation between exposure and
disease in two areas***

|             | Exposed | Not exposed | Total |
|-------------|---------|-------------|-------|
| Area 1:     |         |             |       |
| Disease     | 10      | 90          | 100   |
| No disease  | 90      | 810         | 900   |
| Total       | 100     | 900         | 1000  |
|             |         |             |       |
| Area 2:     |         |             |       |
| Disease     | 90      | 210         | 300   |
| No disease  | 210     | 490         | 700   |
| Total       | 300     | 700         | 1000  |

*Values are numbers of people
Source: Table from MacRae K, *BMJ* 1994;309:1478-1479 (3 December)

## 3.8 Measures based on multilevel models

Section 3.6 explained that, in studies of variation in general practices, there will always
be a certain amount of variation between practices arising from the clustering of patients
within practices. The Ordinary Least Squares Regression method described in Section
3.7.3 does not take into account this clustering. Since this thesis aims to explain the
variation *between* practices, this must be separated from the variation within practices,
or there may be confounding of results. Multilevel modelling techniques allow a
separation and quantification of the contribution of both of these relative to the total
variation (Goldstein et al (1995)). In this thesis, 2 level random intercepts models are

used to enable partitioning of the total variation into that due to differences between

practices (level 2) and that due to differences within practices (level 1). Random

intercepts models are a further development of standard regression models in that they

take account of the between practice variability by allowing the model intercept to vary

across practices. Rather than a single line fit to all the data at once as in

Figure 2, the multilevel model with random intercepts allows for multiple lines, one for

each general practice. Figure 3 illustrates how the intercepts might vary.

**Figure 2  A fixed intercept model**
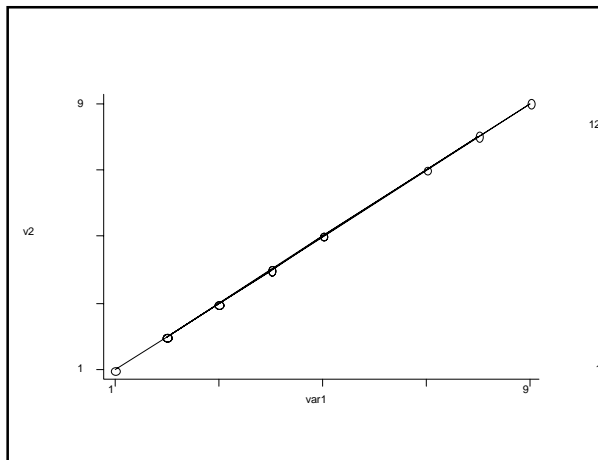


**Figure 3  A random intercepts model (intercepts varying across practices)**
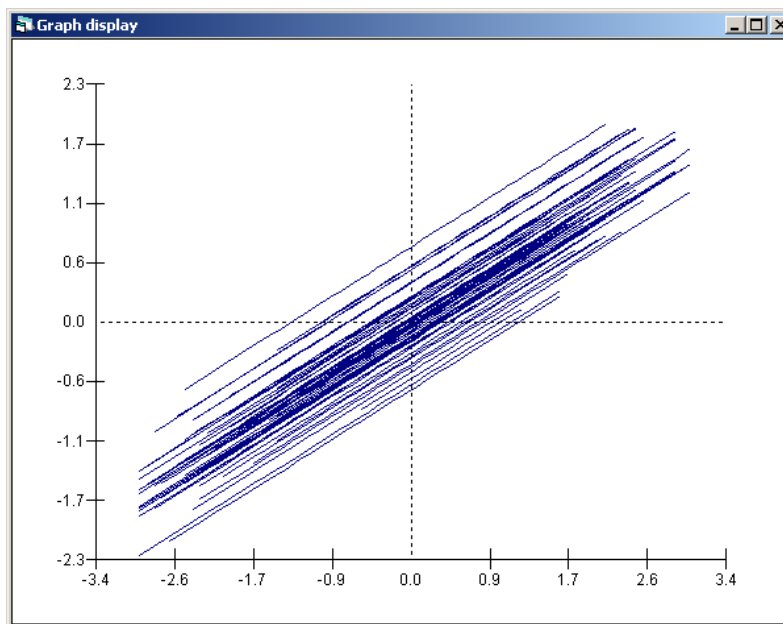


The resulting model residuals are assumed to have a normal distribution. Multilevel modelling has already been used widely in the UK for research in the field of education and human studies of inheritance and more recently in health service studies (Leyland AH, Goldstein H. (Eds.), 2001).  For example, in the UK it was used to take into account individual patient data and small area statistics when investigating the relationship of factors such as socio-economic and health status with measures of general practice workload such as consultation rates (Carr-Hill, RA et al, 1994; 1996). The multilevel modelling allowed them to compare the individual patient characteristics

with area level and to conclude that methods of resource allocation only based on area of residence will always be inferior to those taking into account individual patient characteristics. They found the variation between areas was highly significant, although for a typical patient the overall unexplained variation was relatively small (up to 8%) relative to the total unexplained variation. Multilevel modelling is increasingly used in studies of general practice variation and such studies are described in more detail in Chapter 2 (Bolanis-Carmona et al, 2002; Franks et al 1999; 2000; Davis et al, 2000; 2002; Scott and Shiell, 1997).

Duncan C et al (1998) used multilevel statistical methods in establishing whether regional variations in psychiatric morbidity in Britain remained after controlling for individual and area level factors. Their results contradicted earlier work suggesting a clear north-south divide in psychiatric morbidity. They found that individual characteristics were associated with mental wellbeing, but found that a large degree of individual variation remained unexplained.

Outcomes that are discrete require more complex multilevel analysis in comparison with continuous outcomes. Measures of general practice resource use are often discrete measures of activity such as the number of home visits, referrals, prescriptions and consultations. Despite complexities of multilevel modelling, results from using such techniques can be communicated to general practitioners, policy makers and other healthcare professionals using simple measures like odds ratios, the intracluster correlation coefficients and R-squared measures (Turner RM et al (2001); Goldstein H (1995), Snijders and Boskers (1999)).

Multilevel regression is superior to the ratio of observed to expected (O/E) and CV methods mentioned above because, unlike these methods, it is not a summary method and allows both patient and practice level covariates to be taken into account (allowing variation to occur at each level in the hierarchy structure). Such models can easily be extended to adjust for a number of covariates. A strong focus in this research was to ascertain how much variability is explained after adjusting for factors such as patients' morbidity, age and sex and to understand the degree of unexplained variation at the practice level.

A number of measures can be derived from results of fitting a multilevel model in order to estimate this variation. In this research, the ICC and R-squared measures are used since the proportion of variation at the practice level can be quantified using these measures (Snijders and Boskers, 1999; Goldstein et al, 2002, 2003; Turner RM et al, 2001).

## 3.9 Multilevel models: Total variation

For the 2 level random intercepts models, the total variation in the outcome (Equation 1 Total variation) can be expressed as the sum of the variation explained and of the unexplained variation both at practice and patient levels.


Equation 1  Total variation (2 level random intercepts model)

| |
|---|
| Total variation = variation explained by model |
| + variation unexplained by model at practice level |
| + variation unexplained by model at patient level |



## 3.10 Intracluster Correlation Coefficient (ICC)

The ICC estimated from a 2 level multilevel model fitted to general practice data can be interpreted as the proportion of the unexplained variability that is due to practices. (Commenges, D. and Jacqmin, H. (1994); Donald, A. and Donner, A. (1987)).  For a model with no covariates, it can be interpreted as measuring the proportion of the total variability that is accounted for by the clustering of patients within practices (Donner, A. and Klar, N. (2000)).  The ICC for a model with covariates is often called the residual ICC, as it denotes the proportion of the residual unexplained variation that is due to practices (S&B).  The ICC can also be interpreted as the correlation between any two patients from the same practice.  It can be conceptualised as a measure of the extent to which members of the same cluster (for example, patients within practices) are more similar to one another than to members of other categories (Cohen J et al (2003)).


It is not straightforward to estimate the ICC when the outcome of interest is binary for example whether a patient received a home visit or not or whether they had a referral Various methods have been proposed to estimate an ICC for binary outcomes (Snijders and Boskers, 1999; Goldstein et al, 2002, 2003; Turner RM et al, 2001)).  If an ICC is to be used in comparisons of models fitted to the same outcome as in this thesis, it is important that the method of estimating the ICC is chosen carefully.  Firstly, it must compare like with like, therefore a summary measure of overall ICC is most useful in this instance, rather than an ICC for each combination of covariates.  Secondly, when

comparing two multilevel models, each with two levels (practices and patients), either the numerator or the denominator of the ICC should be fixed to allow for comparison. The reason for this is as follows: The amount of variation explained by a model depends on the covariates included in the model. The denominator of the ICC is usually the unexplained variability (Turner's method is an exception). Each time a model is fitted to the data that is an improvement on the previous model in terms of explaining variability, the unexplained variability correspondingly decreases and therefore the denominator of the ICC decreases. Therefore, the ICCs from models with different sets of covariates fitted to the same data are not directly comparable if quoted without taking account of the variability explained by the model, because the denominator changes for each model. The exception to this is to use a measure of ICC that only has one unknown value for unexplained variation at practice level for example, the Turner, RM et al (2001) and the Snijder & Bosker (1999) methods (described below).

### 3.10.1 Estimating ICC from multilevel logistic regression models

The ICC estimated from a 2 level multilevel model is defined as:

Equation 2 Intracluster Correlation Coefficient – standard definition

ICC = between cluster variation / (between cluster variation + within cluster variation)

(Commenges, D. and Jacqmin, H (1994); Donald, A. and Donner, A. (1987)).

For discrete outcomes it is not straightforward to estimate the within cluster variance required for estimation. Several different methods of estimating ICC were used in this research. Only the Turner method and Snijder & Bosker method are presented in the body of this thesis (see below), the others are presented in the appendix. Two different methods of estimating ICC were used in this thesis, Turner, RM et al (2001) and Snijders & Bosker (1999) and results are compared.

### 3.10.2 ICC – Turner's method

Turner et al developed a simple method of estimating the ICC for clustered data with a binary outcome (Turner, RM et al (2001)):

Equation 3  Intracluster Correlation Coefficient – Turner's method

$$ICC \sim \sigma_u^2 \times \pi \times (1 - \pi)$$

where $\pi$ is the expected value of the cluster specific response proportion and $\sigma_u^2$ is the cluster level variance.  In the case of the 2 level multilevel models used in this research, the $\pi$ can be thought of as the expected practice outcome proportion and the $\sigma_u^2$ is the between practice variance.

This method assumes that response probabilities vary between practices with a probability $\pi_j$ and that individuals from the same practice respond independently. Turner's method is simple to implement and confidence intervals for the ICC estimates can be obtained.  An advantage of this method for the purpose of this research is that the denominator is constant, and therefore this measure can be used to directly compare the residual practice variation between different models.  This method is not appropriate when practice level covariates are available since practice specific response probabilities may vary (Commenges, D., Jacqmin, H. (1994)).  Confidence intervals for the ICCs computed using Turner's method were obtained with parametric bootstrapping (Goldstein, H (1995)).

### 3.10.3 ICC – Snijders & Bosker's method

An alternative method of estimating ICC proposed by Snijders and Bosker (1999) makes the assumption that there is a continuous unobservable latent variable underlying the binary response, for example, the propensity to issue a prescription.  This underlying distribution is then assumed to be logistic and therefore the within practice variance can be estimated by $\pi^2/3$ or 3.29 (variance of the logistic distribution).  Thus, only the between practice variance is needed in order to estimate the ICC and  comparisons of ICCs resulting from fitting different multilevel models will not be affected by different within-practice variance estimates.

## 3.11 R-squared - Snijders & Bosker's method

The R-squared measure proposed by Snijders and Bosker (1999) is very useful for investigating variation in models with different sets of covariates (applied to the same

data). The Snijders and Bosker R-squared measure can be computed from the 2-level random intercepts models in this thesis to calculate:

(1) The proportion of the total variation in outcome that is explained by the model

(2) The proportion of the total variation in outcome that is due to differences between practices,

(3) The proportion of the total variation in outcome that is due to differences between patients within practices

Whichever of the above proportions are calculated, the denominator remains the same (the total variation in the outcome) and sums to 1.

Equation 4 presents the R-squared calculation for the variation explained by the model. In simple terms, it is the unexplained variance that remains between practices after fitting the model divided by the total variance in outcome.

Equation 4  R-squared – Snijders & Bosker method

$$\mathrm{R}-squared = \frac{\sigma_F^2}{\sigma_F^2 + \sigma_u^2 + \sigma_e^2}$$

where $\sigma_F^2$ is the variance explained by the model, $\sigma_u^2$ is the between practice variance and $\sigma_e^2$ is the within practice variance.

Both the residual (after fitting the model) unexplained variation between practices and the residual unexplained variation between patients within practices can be computed in a similar manner to Equation 4 by substituting the numerator with $\sigma_u^2$ or $\sigma_e^2$ respectively.

In assessing which of two regression models, each with different sets of covariates, explain more of the variation in a performance outcome, one can compare the R-squared value resulting from each model. It is also possible to produce an R-squared value for both the unexplained variability at practice and patient levels respectively. The three R-squared measures will sum to 1. Where the difference in R-squared explained is large, say 20%, the conclusion as to which factors explain more variability is obvious, but when the difference in the values is small it may just be due to chance and there is no formal way of assessing whether this is the case. It is reasonable to compare R-

squared values from different models as long as the same population has been used for comparison.

A disadvantage of using R-squared is that it is a point estimate with no corresponding measure of precision such as a confidence interval

# 3.12 Median Odds Ratio

The median odds ratio (MOR) (Larsen K & Merlo J (2005)) allows us to quantify the variation between general practices in terms of an odds ratio. The MOR can be conceptualised in the following way: Consider two randomly chosen subjects with the same covariates but from two different general practices (e.g. two males, aged 65+) and conduct a hypothetical experiment calculating the odds ratio for the person with the higher propensity for, say, a referral versus the person with the lower propensity. Repeating this will lead to a series of odds ratios and the median of these is the MOR.

This may be mathematically expressed as

**Equation 5  Median Odds Ratio**

$$\exp\sqrt{2\sigma_u^2\phi^{-1}(0.75)}.\phi^{-1}(0.75)$$

where $\phi$ is the cumulative distribution function of the Normal distribution. The above formula calculates the median odds ratio between patients of higher propensity e.g. for, say, a specialist referral and patient of lower propensity from different practices.  A MOR of 1 implies that there is no variation between general practices.  As with the ICC, the MOR assumes an underlying latent distribution for the binary outcome (previously described in Section 3.10.3).  It has the advantage of being directly comparable with the odds ratios produced for the model predictors.  A limitation is that the MOR alone does not convey useful information regarding the model explained variation (a change in MOR is related to a change in the unexplained variability attributable to practices).

# 3.13 Graphs used to illustrate variability

Each general practice's observed outcome percentage can be plotted against their predicted percentage, together with 95% intervals. Practices lying outside the interval may either be performing better or worse than expected given their patient characteristics. These plots can be useful for informal comparison of different models and will show outlying practices. However it is cumbersome to present the results graphically in this manner when the number of general practices in the database is large.

# 3.14 Model predictive performance

### 3.14.1 Assessing predictive accuracy of models

A scatter plot of each general practice's observed percentage of patients experiencing the outcome of interest compared to the expected percentage after adjustment for case-mix can be produced for all models. The model with better predictive accuracy will have observations closer to the line of equality, while models with low accuracy will show a lot of variability around the line. The advantage of such plots is that they give insight into between practice variation within covariate groupings. The limitation is that the plots for different models applied to the same data are not directly comparable.

### 3.14.2 Receiver Operating Curve Area

The Receiver Operating Curve (ROC) area measures the discriminatory ability of the model. This is calculated by dividing patients into pairs and calculating the proportion of times a patient is correctly ranked based on their predicted probability from the model of experiencing the outcome (Hanley J et al (1982)) (In a pair of patients where one patient is observed to experience an outcome and the other does not, the former should have a higher probability predicted by the model). A ROC area of 1 indicates that the model discriminates perfectly between patients who have a greater chance of having the outcome and those who have less of a chance, whereas a ROC area of 0.5 indicates that the model discriminates no better than chance.

The above two methods provide information on the predictive ability of the model but do not provide direct information on variation between institutions.

## 3.15 Summary

The first part of this chapter covers the ACG system methodology, describing its components and giving an overview of how these were originally developed and validated. The application of the ACG system to general practice data from the UK is discussed. The coding system in general practices in the UK is different from the US, and the transferability of codes to the UK from the US is considered.

The second part of this chapter focuses on statistical methodology. Some of the most popular statistical methods previously used for explaining variation in general practice outcomes are outlined along with their advantages and disadvantages. The natural hierarchy of patients within general practices is described, and multilevel modelling techniques and associated derived measures for assessing the contribution of various factors in explaining variation in general practice outcomes are presented. The subsequent three chapters will demonstrate applications of these methods to general practice home visit data (Chapter 4), referral data (Chapter 5) and prescribing data (Chapter 6).

# Chapter 4     Home visits

## 4.1 Introduction

As a first step in applying the Johns Hopkins ACG system to general practice populations in the UK, GP home visits are investigated. Home visits remain an important aspect of general practitioner's workload, despite falling levels over the years, and home visiting patterns vary greatly from practice to practice (Aylin P et al, 1996). The overall practice age, sex, morbidity and social class composition may affect a practice's propensity to make home visits and thus should be taken into account in comparisons of home visits across practices. Age and sex are routinely recorded in general practice and relatively straightforward to extract, but information on morbidity and social class is more difficult to obtain. Of these two characteristics, a morbidity measure based on patient diagnoses is arguably the easier to collect, as practices are computerised and doctors are increasingly given incentives to record patient diagnoses on an ongoing basis. There are many different measures that have been developed as proxies for patient morbidity. Most of these tend to use the most important or most common morbidity of each patient (Ellis R et al, 1996; Averill RF et al, 1999), thereby potentially ignoring a host of other morbidity information. Unlike these, the Johns Hopkins ACG (Adjusted Clinical Groups) Case-Mix System is a method of measuring diagnostic-based case-mix based on GP records of diagnoses for an entire time period (usually one year) and the measures produced can be interpreted as a proxy for patient morbidity (http://www.acg.jhsph.edu; Weiner JP et al, 1991; Starfield et al, 1991). The ACG system has been described in more detail in Section 3.2.

This chapter illustrates the use of the Johns Hopkins ACG Case-Mix System in the UK primary care setting using data from the Fourth National Morbidity Survey (MSGP4), a one-year prospective cohort study of 500,000 patients in 60 general practices in England and Wales (RCGP, ONS & DH, 1995). The US has pioneered much of the development of diagnostic based morbidity measures, however, it is possible that social class, which is a measure not generally collected in the US, may provide as much

insight into practice patterns as diagnostic based morbidity. The aim of this chapter is to investigate whether patient level measures of morbidity, assigned using the Johns Hopkins ACG Case-Mix System, explain more of the variation in home visits between general practices from the Morbidity Survey in General Practice (MSGP4) than age/sex and social class measures. This study has the advantage that both a diagnostic based morbidity measure and a social class measure are investigated simultaneously.

In this chapter, the ACG system is used to assign each patient to a morbidity group in an investigation of GP home visits. Home visiting patterns are known to vary widely between practices. The odds of a home visit after adjusting for combinations of age, sex, morbidity and social class are examined. This should provide evidence as to the effect of these factors on home visits. For example, certain social groups might have lower odds of home visits than others, after taking into account their age, sex and morbidity distribution. In particular, the focus of this chapter is on whether morbidity and social class explain more of the variability in home visits between practices than age and sex alone.

Methods are used to try to quantify how much of the variation in home visits is at the practice level and how much is at the patient within practice level. This will contribute to the understanding of which factors explain variability at which levels. A simple summary measure often used in the primary care setting to explain variation in general practice outcomes is described (Salem-Schatz, S. et al (1994); Reid, R. et al (1999)). Methods of estimating variability based on multilevel models are also used. The results obtained from the methods based on the summary measure and the multilevel models are compared.

## 4.2 Methods

### 4.2.1    Morbidity Statistics in General Practice

The Fourth National Survey of Morbidity in General Practice (MSGP4) is a one-year prospective cohort study (October 1991 to September 1992) of over 500,000 patients registered with 60 general practices in England and Wales (RCGP, ONS & DH (1995)). The main objective of the survey was to examine the workload and pattern of disease in

general practice in relation to the age, sex, and socio-economic status of patients. The MSGP4 database includes every type of consultation, including home visits, recorded by the general practitioner and all diagnoses recorded by the GP for each patient. These data are linked to the socio-economic characteristics of each patient.

### 4.2.2    Data recording and validating

Doctors and nursing staff from each practice attended three two-day training sessions on the recording of morbidity data. Practices then collected data for two to four weeks before the start of the survey. The data collection software was designed so that all diagnoses were automatically coded using the Read classification system (Saint-Yves IF (1992)). The data was analysed and any errors and inconsistencies were reported to the practices and amended if necessary. Information on socio-economic status was obtained on 83% of patients in the survey by direct interview with specially trained interviewers. The interview method was successfully tested for feasibility and acceptability before the survey. The social class measure used was derived from occupation and employment status. Social class was grouped in a similar way to the UK census groupings (Table 4). An International Classification of Diseases, Ninth Revision (ICD-9) code was assigned at the Office of Population Censuses and Surveys. Data was well validated and the collection and validation process is described in detail in the MSGP4 publication (RCGP, ONS & DH (1995)).

### 4.2.3    Study population

The study population for the fourth National Morbidity Survey was a one per cent sample (502,493 patients) of the population from 60 general practices in England and Wales. All patients in these practices who were on the NHS age/sex register were included. The sample was representative of the population of England and Wales for characteristics such as age, sex, social class and housing tenure and under-representative of ethnic minority groups and people living alone (RCGP, ONS & DH (1995)).

### 4.2.4    Exclusions

The following patients were excluded: 137,273 (27%) did not consult in the study period; 15,682 (4%) did consult but were registered at a practice for less than six

months; 33 patients consulted but had no diagnosis information.  After applying the exclusion criteria, the analysis dataset comprised 349,505 patients.

### 4.2.5    Morbidity groups

The Johns Hopkins ACG Case-Mix System is a tool used to characterise the degree of overall morbidity in patients and populations http://www.acg.jhsph.edu (Section 3.2). In the example shown in Table 2, Section 3.2.4, two patients each have diabetes but are assigned to different ACG groups since one of the patients has a much greater overall morbidity than the other.  Each ICD-9 diagnosis code for each patient is mapped to one of 32 diagnosis groups known as ADGs (Aggregated Diagnostic Groups).  A small proportion (1%) of ICD-9 codes could not be assigned an ADG (Majeed A et al, 2001). Diagnoses are grouped within the same ADG based on similar severity and expected need for health care resources over time.

Each individual was also assigned a single mutually exclusive ACG (of which there are about 90 groupings), derived from a combination of age, sex, presence of specific ADGs, number of major ADGs and total number of ADGs.  The ACG groupings contain individuals with similar needs for health care resources based on overall expenditures.  Patients with similar predicted (or expected) overall utilisation may be assigned different ACGs if they have different epidemiologic patterns of morbidity.  For example, Section 3.2.4 illustrates how two different patients, each with diabetes mellitus, are assigned to different ADGs, ACGs, and RUBs according to their degree of morbidity burden.  For analysis, the ACGs were collapsed to one of eight mutually exclusive morbidity groups, known as Resource Utilisation Bands (RUBs), higher numbers indicating higher morbidity (Weiner et al (1991a)). The ACG groupings for the data on home visits were made using version 4.5.

## 4.3 Statistical methods

Home visit frequency distributions were calculated for each of the age, sex, morbidity and social class groupings. Multilevel logistic regression models (each with a random intercept) were used to investigate important predictors for home visits (Goldstein H (1995)). Five separate models with different sets of fixed predictors were examined, the outcome of interest being whether or not a patient had a home visit in the one year study period. The sets of predictors included were: (1) age group & sex; (2) morbidity; (3) social class; (4) morbidity & social class. Model diagnostics included plotting standardised residuals against their normal scores for each model to ensure that the assumption of normality was satisfied, and checking for overdispersion. Adjusted odds ratios were computed from the results of each of the models. The total variation in home visits is likely to be due to different characteristics influencing home visits both between patients within practices (within practice variance), for example, patients' expectations of receiving a home visit, and from practice to practice (between practice variance), for example, general practice's role perception. After fitting the model, the proportion of the total variation in home visits explained by each model is quantified with an R-squared measure proposed by Snijders and Bosker (1999). The proportion of unexplained variance in home visits that is attributable to differences between practices can be quantified with the ICC. An intracluster correlation coefficient (ICC) was estimated from each of the models and used to assess how the unexplained variation in home visits between practices altered when comparing models with different sets of predictors (Turner et al (2001)). The Snijders & Bosker formula was also applied for estimating ICC, and estimates the within practice variance to be 3.29. Thus, comparisons of ICCs resulting from the four models will not be affected by the different within practice variance estimates (Davis P et al (2000)).

After fitting the model with age and sex, the probability (and 95% interval) of having at least one home visit was estimated and plotted for each age group for both males and females. Similarly, after fitting the model with morbidity as a predictor, the probability of having at least one home visit (and 95% interval) was estimated and plotted for each morbidity group.

## 4.4 Summary measures of variability

The coefficient of variation (CV), defined as the ratio of the standard deviation to the mean (Bland M, 1995) (Section 3.7.2). The adjusted log odds of home visits was calculated separately for each of the 60 practices from the MSGP4 study. Adjustment for age and sex was made using the Woolf's method (Breslow, NE and Day, NE, 1980).

There are 8 age groups, thus providing a total of 16 age-sex groupings for each practice. A pooled estimate of log-odds is estimated across the 16 groups (Omar RZ, Thompson SG (2000)). The log odds adjusted for age and sex was then converted to the probability of a home visit for each practice. The average probability of home visit and the corresponding standard deviation for 60 practices were calculated and a CV estimated. A CV adjusting for the 8 morbidity groups was calculated in a similar manner. The adjustment factor with the lowest CV implies that more of the between practice variability in home visits is explained.

General practice data has a natural structure with practices at the higher level (level 2) and patients nested within practices at another level (level 1). Section 3.6 described the clustered nature of general practice data and the importance of using appropriate statistical methods for this in order to avoid potentially misleading and erroneous conclusions. Multilevel modelling allows a partitioning of the total variability present in the data into variation between practices and variation within practices. Some of the unexplained variation will be due to the differences between patients within practices. It is important to separate this out from the total variability so that the real variation between practices can be ascertained. Once the multilevel model has been fit to the data and results obtained, one can calculate the ICC (Section 3.10.1). The ICC($\rho$) estimated from a multilevel model defined as in Equation 6 below:

Equation 6  ICC($\rho$) estimated from a 2 level random intercepts model

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2}$$

where $\sigma_u^2$ is the between practice variance and $\sigma_e^2$ is the within practice variance

(Commenges, D and Jacqmin, H, 1994; Donald, A and Donner, A, 1987). Assuming

that $\rho$ cannot be negative, it can be interpreted as measuring the proportion of the

unexplained variability that is due to the real variability between practices (Donner, A

and Klar, N, 2000). The ICC value will be negative if the patients within a practice are

less similar to one another than they are to patients in another practice. This is an

unlikely scenario for the UK primary care setting and therefore it is appropriate to

assume $\rho \geq 0$ for these data. The ICC results from models with different adjustment

factors included were examined to assess the effects of different combinations of factors

on variation in home visits. The R-squared measure was also used as this allows us to

examine the proportion explained by the model and to see how much of the total

variability was explained at practice level and at patient level. The advantage of using

the multilevel modelling is that the inherent clustering of general practice data is

accounted for. Furthermore the models can be easily extended to include further

adjustment factors.

## 4.5 Estimating between-practice variation from multilevel logistic regression models

Several different methods for estimating the proportion of the unexplained variability

that is due to real variation between practices or between-practice variation from

multilevel models have been proposed. This proportion (or percentage) of unexplained

variability is known as the intracluster correlation coefficient (ICC). Turner's (2001)

method of calculating ICC is used and confidence intervals for the ICCs are obtained

using resampling methods (Section 3.10.2). Using the Turner formula for estimating

ICC the overall proportion of home visits can be related with an estimate of the

between-practice variance. An advantage of the Turner method is that it does not

require a direct estimate of the within-practice variance. Thus, ICCs from the four

models will not be affected by different estimates of within-practice variance.

Another method of measuring ICC, proposed by Snijders and Bosker (1999), makes the assumption that the within practice variance can be approximated by $\frac{\pi^2}{3}$ or 3.29 so that only the between practice variance estimate ($\sigma_u^2$) is required:

Equation 7  Intracluster Correlation Coefficient - Snijders & Bosker

$$\rho \sim (\sigma_u^2)/((\sigma_u^2 + 3.29)$$

Several different methods of estimating ICC were used with the home visits data.  Only the Turner method and Snijder & Bosker method are presented in this chapter, the others are presented in the appendix.  The overall conclusions regarding the variability in home visits between general practices remained the same for each method.  The between practice variation in home visits was estimated for each of the four models.  Parameters were estimated with Penalised Quasi Likelihood (PQL) (2nd order).  Descriptive statistics were computed using Stata 7.0 (Statacorp (2001)); graphs were drawn in MS Excel; and multilevel modelling in MlwiN v1.10 (Rasbash et al (2000)).

# 4.6 Results

### 4.6.1    Demographics

The total number of patients included in the final analysis were the 349,505 (55% female, 45% male) who had consulted at least once in the previous year. Table 4 presents the number of patients by age, sex, morbidity and social class, the number of home visits and the odds ratios for each of these groups. Of all patients included in this study, 17% had at least one home visit over the study period. The crude percentage of patients requiring a home visit ranged from 7% to 31% across practices with a median of 18%.

Home visits showed a bimodal distribution and were lowest in the 16 to 44 age group, with peaks at 0 to 4 years and for those aged 65 years and over. The median (range) percentage receiving home visits according to age in the children (0 to 15 years), adult (16 to 64 years) and elderly (65 years onwards) populations respectively were: 20% (4% to 43%); 11% (4% to 21%); 38% (22% to 56%).   The odds of home visits are greatest

for those aged 85 years plus (OR=8.1(7.57 to 8.66)), while the odds of home visits are lowest for those aged between 16 and 64 years.

The number of patients in each ADG and the number and percentage with at least one home visit over the study period are presented by ADG in the Appendix (Table 17).

## 4.6.2   Results from models

The frequency distribution and odds ratios in Table 4 show that age group has a strong association with home visits. Visits are more common in the youngest and oldest age groups (30% for 0 to 4 year olds, and 50% upwards for the over 75s), and less common in adults (about 10% for 16 to 64 year olds). Females had higher home visits (19% vs 15%) than males. The percentage of patients having a home visit increases steadily with increasing morbidity. Home visits are highest for the lowest social classes.

The adjusted odds ratios from fitting the four models (Table 5) show a similar association, although the effect of sex is not as large (OR=0.87 (0.85 to 0.88) compared to OR of 0.75(0.74 to 0.79) in Table 4). All of the covariates fitted are highly significant, and have tight confidence intervals. In examining which of the factors explain most of the variability, the percent ICCs in Table 5 indicate that most of the unexplained variation in home visits is occurring between patients within the same practices rather than from practice to practice, even though the variation between practices is highly significant. Model 1 shows that 2.5% of the unexplained variation in home visits is attributable to practices after taking into account age and sex.

Morbidity was then included as a predictor for home visits to see if this could improve on the age-sex model in explaining variation in home visits across practices. There is a high statistically significant association between morbidity and home visits (p<0.001). Home visits increase steadily with increasing morbidity as indicated by the odds ratios in Table 5. After adjustment for morbidity, the odds of having a home visit are almost eleven times greater in the sickest patients compared to the healthiest patients (OR=10.8, 95% CI (10.3 to 11.2)). Estimates of the intercept log(odds), the between practice variance and the percentage ICCs and associated confidence intervals from

each of the models are included at the end of Table 5. The ICC was 1.6% for Model 2, which fitted only morbidity to the home visits (Note: A model with age group, sex and morbidity as explanatory variables gave similar results).  The ICC for Model 3 fitting only social class was the same as that from Model 2 (morbidity).  The decrease in ICC from a model including age and sex only is not likely to be statistically significant for Models 2 and 3 as the confidence intervals overlap considerably.  Model 4 with explanatory variables morbidity and social class resulted in the greatest reduction in ICC to 1.5% and again, the confidence intervals overlap.  Investigation of residual plots for each model suggests that the residuals were approximately normally distributed.

Table 6 presents the Snijders & Bosker (1999) R-squared measures for each of the models.  The age-sex model explained more of the total variability in home visits than the other models.  However, the morbidity and social class model explained more of the variability between practices, leaving only 3.3% between practices compared to 4.7% (age-sex model).  Similar to Turner's method, the ICCs calculated using Snijders and Bosker method suggest that the proportion of unexplained variability in home visits that is due to differences between practices after adjusting for morbidity is lower than for that adjusted for age and sex.

The probability of home visits is highest for children and elderly compared with adults and the 95% intervals demonstrate that the variability of these estimates for children and elderly patients is also higher (Figure 4).  Home visit probabilities were highest for the sickest patients and again, more variability is apparent in these estimates for the sickest patients (Figure 5).

Figure 6 and Figure 7 present the odds ratio and 95% confidence intervals for home visits by social class, firstly adjusting only for social class and secondly adjusting for both social class and morbidity. Interestingly, the relationship between home visits and social class is altered when morbidity is taken into account.  In Figure 3 the odds ratio of having a home visit are lowest at 0.92 (95% CI: 0.89 to 0.95)

For social class II, but increases steadily in the following order: groups I, IIIN, IIIM, IV & V when no other factors are taken into account.  In other words, the lower social classes have higher home visits, with social class V having an odds of home visit of 1.69 relative to social class I.  However, when morbidity is accounted for (Figure 4) the relationship between social class and home visits alters.  Home visits remain lowest among those in social class II, followed by IIIN.  Social class IIIM has a slightly higher odds ratio, but this is not significantly different from social class I.  The highest home visits were again among social classes IV and V, although difference between highest and lowest groups was attenuated (1.36 vs 1).

Although the social class variable is affected by adjusting for morbidity, the reverse is not true for morbidity.  The unadjusted odds of home visits for morbidity are similar to the odds of home visit for morbidity adjusted for social class.

## 4.7 Discussion

The lack of appropriate adjustment for case-mix is an important limitation of many previous studies because case-mix has been shown to have a major impact on general practitioners' workload and performance.  This chapter focused on the use of a case-mix adjustment system developed in the USA in an attempt to explain the variation in one important area of general practitioners' workload in the UK.  Findings showed that the Johns Hopkins ACG Case-Mix Adjustment System is a strong predictor of home visits for patients *within* practices in the UK.  Figure 4 and Figure 5  show that patterns of association are similar for both intermediate age groups and intermediate morbidity groups, while the elderly, children and sickest patients are more likely to have had a home visit and the variability in home visits for these groups is also highest.  Crude home visits varied from 7% to 31% across practices.  However, this crude variation is composed of variability arising both from different characteristics of home visits

between patients within practices, and from variability resulting from differences in home visits from practice to practice.

The age-sex model explained more of the total variability in home visits. Adjusting for morbidity in the model resulted in a small non-significant improvement in explaining variability in home visits between practices compared to the model adjusting for age and sex only. Adjusting for social class resulted in a similar non-significant improvement, while the model with both morbidity and social class explained slightly more of the between practice variation. The implications of these results are that care should be made when interpreting crude rates between practices and that appropriate consideration should be made of the sources of variation resulting from the clustered nature of general practice studies.

The results also suggest that, after adjusting for patient morbidity, the middle social classes received less home visits (Figure 6), and contrast with assertions of Julian le Grande that state that the middle classes receive the best level of care in the NHS (Le Grande, 2006).

A major strength of this study is that it is population-based as general practices in the UK register patients from all sections of society. Hence, unlike studies of US health maintenance organisations, no socio-economic group was excluded. The data was collected prospectively for one year as part of a morbidity survey and recorded to a high standard. Individual level social-economic information was collected during the survey and thus it was possible to compare the effect of adjustment for morbidity with adjustment for social class. Both of these had a similar effect in reducing variation between practices.

The limitation of the method based on CV is that it is based on a summary measure and does not utilise the full information contained in the data and therefore may be less sensitive. This may explain the lack of difference observed in the results obtained from the use of the two different sets of adjustment factors. Furthermore, this method is cumbersome when the number of adjustment factors needed to account for differences in patient characteristics is large. The CV adjusted for both morbidity and socio-economic class was not calculated for this reason.

The advantage of Turner's ICC method used with multilevel models is that it is simple to implement and confidence intervals for the ICC estimates can be obtained. This is, however, only considered as a rough measure of variability for the home visits outcome. In the statistical methods chapter, alternative methods of estimating ICCs from multilevel models, which incorporate patient level covariates, are described and are applied to the home visits data as part of a sensitivity analysis to examine whether these conclusions regarding variability in home visits still apply. The R-squared measure is also straightforward to implement and has the advantage that one can see how well each set of covariates performs with respect to the total variability in the home visits outcome as well as partitioning the total variability unexplained at between and within practices.

It is possible to extend the multilevel models considered to incorporate practice level predictors. The assumption of normality was satisfied by these data as checked by residual plots. The conclusions drawn in this paper about the various methods are based on data with a moderately large number of clusters and large cluster sizes. These conclusions may not necessarily extend to situations where the number of clusters or/and sizes are small.

Results suggests that, in studies of general practice workload where the social class measure is available, social class could be used in place of morbidity in adjusting for differences between practices. Adjusting for either morbidity or social class gives a small non-significant reduction in the variation in home visits between practices and there is far more variation within practices that remains unaccounted for. This implies that there are other factors occurring in practices that cannot be appropriately adjusted for by age, sex, morbidity and social class alone. Such factors might include varying patient demand for home visits or variation between home visits made by GPs within practices (Webb S et al, 1994; Britten N et al, 1997; Cockburn J et al, 1997). They also imply that care should be made when comparing home visit rates between practices and that appropriate consideration should be made of the sources of variation resulting from the clustered nature of general practice studies.

In the future, general practices' performance will come under greater scrutiny. Experience from the USA shows that it is important to take into account case-mix when

assessing practice performance and the use of resources to avoid 'good apples being labelled as bad' (Salem-Schatz S, et al, 1994). Morbidity measurement may have a role in the measurement of other outcomes in primary care.

## 4.8 Conclusions

The findings of this study suggest that the Johns Hopkins ACG System can be applied to general practice populations in the UK. Morbidity is a strong determinant of home visits within the UK. Adjusting for social class may be useful when comparing home visits between practices in situations where diagnostic information is not available. Morbidity and social class adjustment is a small improvement in explaining variability in home visits between practices compared with adjusting for age and sex. Most of the variation in home visits was attributed to differences between patients within practices rather than between practices and that age and sex explain more of this within practice variability. In addition to clinical case-mix and social class, there could also be other unmeasured factors, such as varying patient demand for home visits, disability or differences in GP home visiting practice style that influence this. Case-mix may also be an important factor in studies of other aspects of between-practice variation.

Even after adjusting for patient morbidity, the lower social classes received more home visits and, for this home visits outcome, they contrast with assertions of Julian le Grande that the middle classes receive the best level of care in the NHS (Le Grande, 2006).

In the next chapter GP referrals to specialists are examined and the ACG System is applied to general practice data from the General Practice Research Database. The extent to which patient level measures of case-mix explain the variability in referral activity and predict referrals made by general practices in the UK is examined.

**Table 4  All patients, percentage of patients with at least one home visit, and odds ratios (OR), by age, sex, morbidity and social class**

| Age(yrs) | Patients | Home visits | (%) | OR (95% CI) |
|---|---|---|---|---|
| 0-4 | 30,378 | 9,208 | (30) | 1 |
| 5-15 | 46,892 | 6,649 | (14) | 0.38 (0.37-0.39) |
| 16-24 | 41,695 | 4,283 | (10) | 0.26 (0.25-0.27) |
| 25-44 | 100,931 | 10,853 | (11) | 0.28 (0.27-0.29) |
| 45-64 | 73,335 | 8,270 | (11) | 0.29 (0.28-0.30) |
| 65-74 | 30,550 | 7,348 | (24) | 0.73 (0.70-0.75) |
| 75-84 | 20,113 | 10,067 | (50) | 2.30 (2.22-2.39) |
| 85+ | 5,611 | 4,370 | (78) | 8.10 (7.57-8.66) |
| Total | 349,505 | 61,048 | (17) | |
| | | | | |
| Sex | Patients | Home visits | (%) | OR (95% CI) |
| Female | 195,838 | 37,694 | (19) | 1 |
| Male | 153,667 | 23,354 | (15) | 0.75 (0.74-0.77) |
| Total | 349,505 | 61,048 | (17) | |
| | | | | |
| Morbidity | Patients | Home visits | (%) | OR (95% CI) |
| 1    Healthiest | 94,988 | 8,257 | (9) | 1 |
| 2 | 64,602 | 8,060 | (12) | 1.5 (1.45-1.55) |
| 3 | 67,511 | 10,030 | (15) | 1.83 (1.78-1.89) |
| 4 | 54,270 | 11,766 | (22) | 2.91 (2.82-3.00) |
| 5 | 25,074 | 5,996 | (24) | 3.30 (3.18-3.42) |
| 6 | 18,196 | 4,973 | (27) | 3.95 (3.80-4.11) |
| 7 | 13,443 | 6,053 | (45) | 8.60 (8.26-8.96) |
| 8    Sickest | 11,421 | 5,913 | (52) | 11.28 (10.80-11.77) |
| Total | 349,505 | 61,048 | (17) | |
| | | | | |
| *Social class | Patients | Home visits | (%) | OR (95% CI) |
| 1 | 22,160 | 2,654 | (12) | 1 |
| 2 | 73, 968 | 10,238 | (14) | 1.18 (1.13-1.24) |
| 3 | 42,799 | 6,985 | (16) | 1.43 (1.37-1.50) |
| 4 | 92,922 | 16,390 | (18) | 1.57 (1.51-1.64) |
| 5 | 48,236 | 9,925 | (21) | 1.90 (1.82-1.99) |
| 6 | 16,573 | 4,151 | (25) | 2.46 (2.33-2.59) |
| 7 | 2,910 | 527 | (18) | 1.63 (1.47-1.80) |
| 8 | 20,697 | 5,127 | (25) | 2.42 (2.30-2.55) |
| 9 | 29,240 | 5,051 | (17) | 1.53 (1.46-1.61) |
| Total | 349,505 | 61,048 | (17) | |

*Social class groupings

| | | |
|---|---|---|
| 1 | I | Professional etc occupations |
| 2 | II | Intermediate occupations |
| 3 | IIIN | Skilled occupations: non-manual |
| 4 | IIIM | Skilled occupations: manual |
| 5 | IV | Partly skilled occupations |
| 6 | V | Unskilled occupations |
| 7 | | Armed forces |
| 8 | | Unoccupied – includes students, housewives, persons of independent means, permanently sick or disabled, persons who have never worked & occupation not stated |
| 9 | | Inadequately described/Not available |

**Table 5  Odds ratios (OR) & 95% confidence intervals from multilevel logistic regression**

Intercept log odds, between practice variance estimates, intracluster correlation coefficient (ICC) estimates & their 95% confidence intervals are presented at the foot of the table.
Sets of predictors are:
1) Age, sex
2) Morbidity
3) Social class
4) Morbidity & social class

| | Age & Sex OR (95% CI) | Morbidity OR (95% CI) | Social Class OR (95% CI) | Morbidity & Social Class OR (95% CI) |
|---|---|---|---|---|
| **Age(yrs)** | | | | |
| 0-4 | 1 | | | |
| 5-15 | 0.27 (0.26, 0.28) | | | |
| 16-24 | 0.20 (0.19, 0.20) | | | |
| 25-44 | 0.22 (0.21, 0.22) | | | |
| 45-64 | 0.29 (0.28, 0.30) | | | |
| 65-74 | 0.72 (0.70, 0.75) | | | |
| 75-84 | 2.33 (2.25, 2.42) | | | |
| 85 on | 8.13 (7.59, 8.70) | | | |
| | | | | |
| **Sex** | | | | |
| Female | 1 | | | |
| Male | 0.87 (0.85, 0.88) | | | |
| | | | | |
| **Morbidity** | | | | |
| 1 | | 1 | | 1 |
| 2 | | 1.50 (1.45, 1.54) | | 1.50 (1.45, 1.55) |
| 3 | | 1.81 (1.76, 1.87) | | 1.81 (1.75, 1.87) |
| 4 | | 2.71 (2.63, 2.80) | | 2.88(2.79, 2.97) |
| 5 | | 2.69 (2.58, 2.79) | | 3.20 (3.08, 3.32) |
| 6 | | 3.43 (3.30, 3.58) | | 3.86 (3.71, 4.02) |
| 7 | | 8.15 (7.82, 8.49) | | 8.35(8.01, 8.70) |
| 8 | | 10.80 (10.3, 11.2) | | 10.70 (10.24, 11.16) |
| | | | | |
| **Social class** | | | | |
| 1 | | | 1 | 1 |
| 2 | | | 0.92 (0.89, 0.95) | 0.81 (0.78, 0.84) |
| 3 | | | 1.08 (1.05, 1.12) | 0.94 (0.90, 0.97) |
| 4 | | | 1.13 (1.10, 1.17) | 0.98 (0.95, 1.01) |
| 5 | | | 1.36 (1.31, 1.40) | 1.14 (1.10, 1.18) |
| 6 | | | 1.69 (1.62, 1.76) | 1.36 (1.30, 1.42) |
| 7 | | | 1.17 (1.05, 1.30) | 1.09 (0.98, 1.21) |
| 8 | | | 1.79 (1.72, 1.86) | 1.60 (1.53, 1.67) |
| 9 | | | 1.47 (1.15, 1.87) | 1.18 (0.92, 1.52) |
| log odds (intercept) | -0.708 | -2.338 | -1.777 | -2.352 |
| $\sigma_u^2$ | 0.18 | 0.13 | | 0.12 |
| ICC (%) | 2.5 (1.4–3.2) | 1.6 (1.1–2.4) | 1.6 (1.1–2.8) | 1.5 (1.1–2.2) |

*Morbidity grouping incorporates age & sex.  Reference groups:  Model 1: Females aged 0-4 yrs;  Model 2: Females & males in morbidity 1;  Model 3: Females & males in social class 1;  Model 4: Females & males in social class 1 & morbidity 1

**Table 6  Model unexplained variation in home visits at practice and patient level, & R - squared values**

| Variation | Empty | Age & Sex | Morbidity | Social class | Morbidity & Social class |
|---|---|---|---|---|---|
| between practices | 0.144 | 0.191 | 0.134 | 0.134 | 0.128 |
| within practices | 3.29 | 3.29 | 3.29 | 3.29 | 3.29 |
| predicted home visits | 0 | 0.57 | 0.393 | 0.04 | 0.436 |
|  |  |  |  |  |  |
| Unexplained practice level (%) (ICC (%)) | 4.2 (4.2) | 4.7 (5.5) | 3.5 (3.9) | 3.9 (3.9) | 3.3 (3.7) |
| Unexplained patient level (%) | 95.8 | 81.2 | 86.2 | 95 | 85.4 |
| R-squared (%) | 0 | 14.1 | 10.3 | 1.1 | 11.3 |

**Figure 4  Probability of home visit (95% interval) for males by age group (estimated from model including age group and sex)**
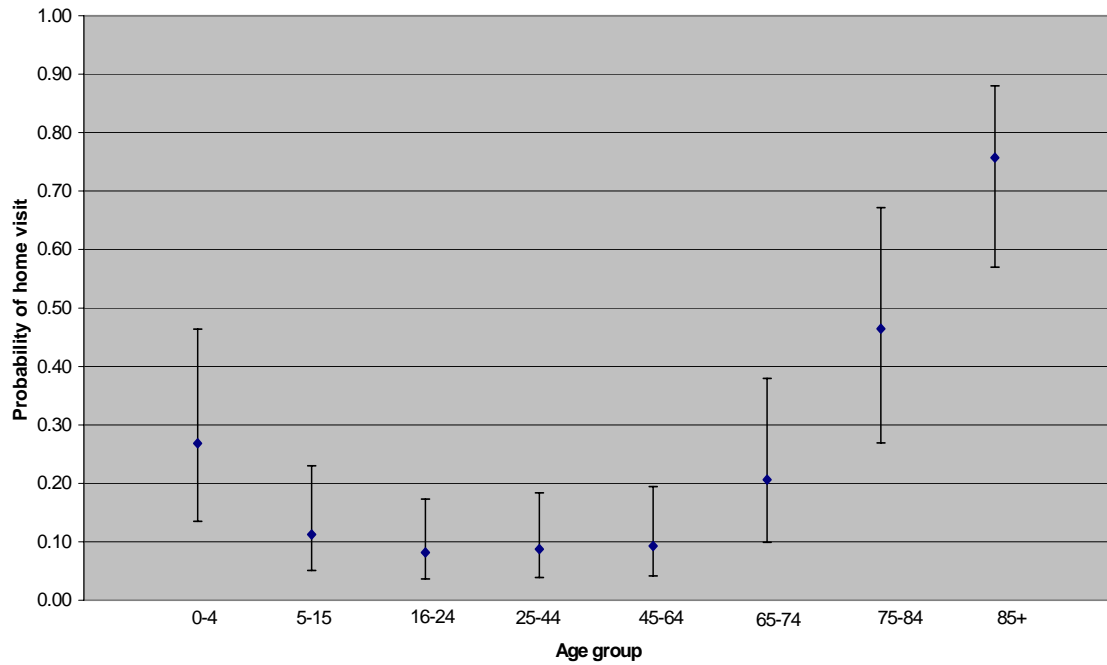
**Figure 5  Probability of home visit (95% interval) for males and females by morbidity group (estimated from model including morbidity)**
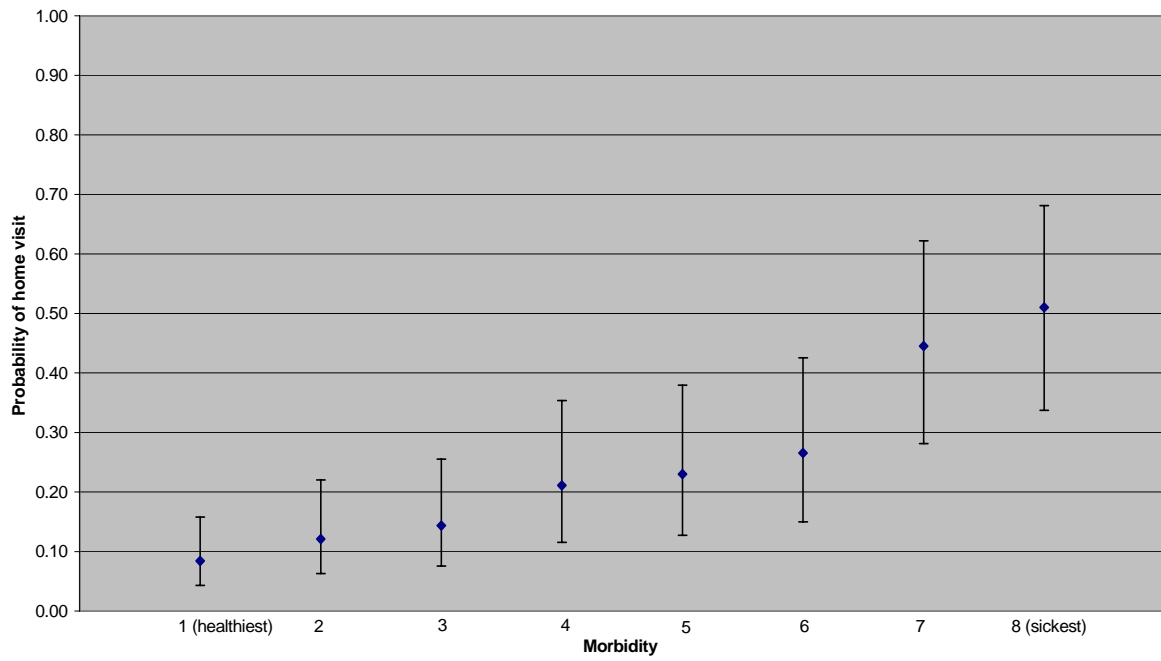
**Figure 6  Odds ratio of home visits, presented by social class**
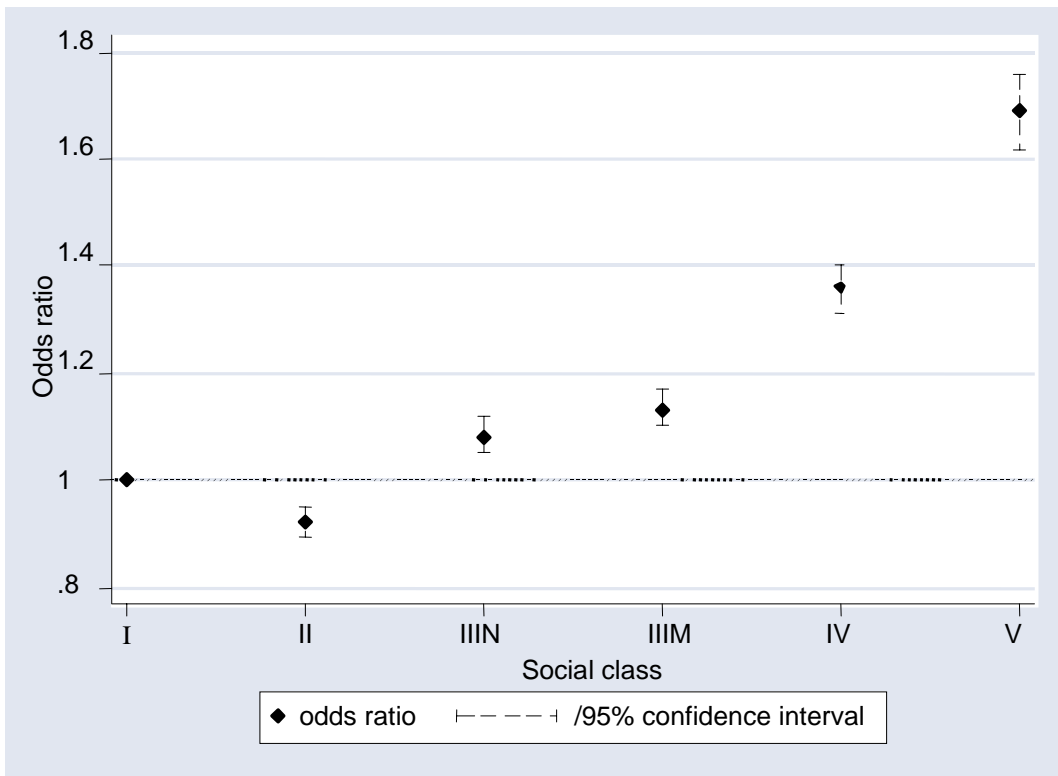
**Figure 7  Odds ratio of home visits adjusted for morbidity, presented by social class**
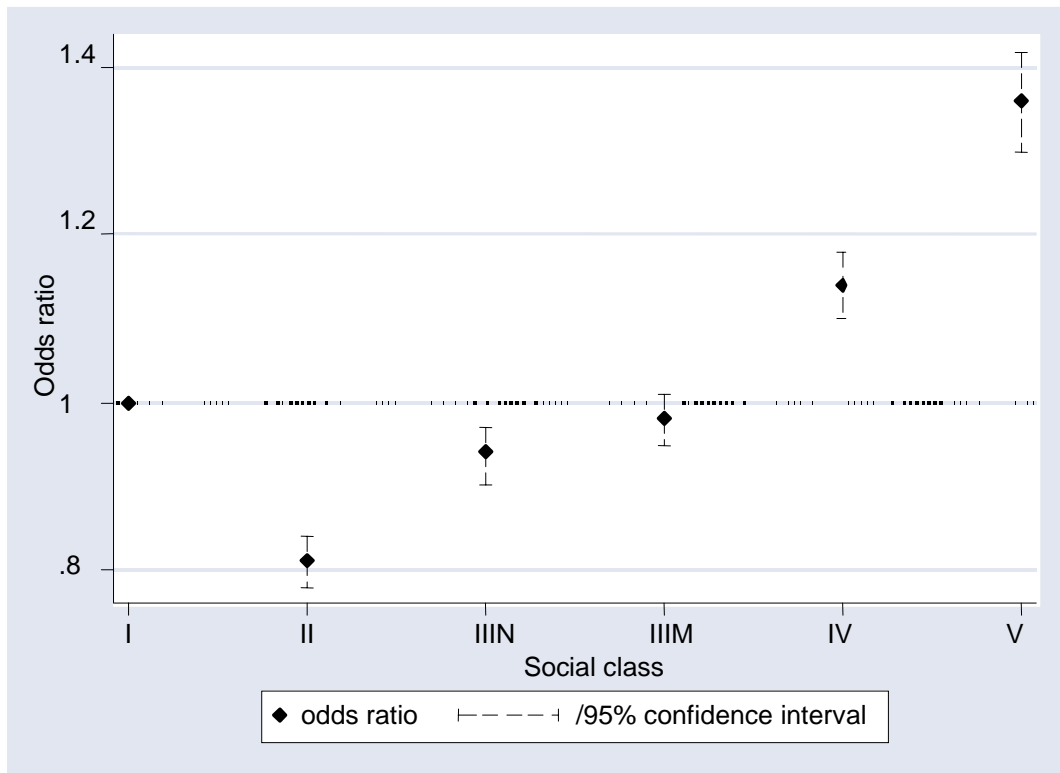
**Table 7  Coefficient of variation for models with home visits as outcome**

|  | Mean proportion of home visits (SD) | Coefficient of variation |
|---|---|---|
| Model |  |  |
| (1) age & sex | 0.191 (0.047) | 0.248 |
| (2) morbidity | 0.191 (0.050) | 0.263 |

# Chapter 5    Referrals

## 5.1 Introduction

In the United Kingdom, a non-emergency patient's first contact with the National Health Service is with a general practitioner (GP). The supply of specialists in the UK is limited (Forrest CB et al, 2003) and thus, a key role of general practitioners is to act as gatekeepers, with the responsibility of making decisions such as whether to refer a patient to specialist services or to continue managing the patient in primary care. The limited supply of specialists in the NHS means that there can be long delays before patients are seen after a routine outpatient referral (Blendon RJ et al, 2001). Specialist care is also typically much more expensive than treatments provided in general practice. Variations in referrals have important implications for NHS spending and patients' access to specialist health services. When patients are referred appropriately by general practitioners and managed appropriately by specialists, this ensures that they receive treatments that will improve their prognosis and quality of life. Hence, it is important to find the correct balance between the generally cheaper and more accessible services delivered by general practitioners and the more expensive services delivered by hospital specialists (Forrest CB et al, 2003).

Ideally, general practitioners would refer those patients who would benefit most from specialist care, whilst retaining the management of other patients within primary care. GPs attempt to make decisions about who to refer appropriately, yet there are many studies examining referral rates which have typically found large variations among general practices (O'Donnell CA, 2000). Results of these studies suggest that such variations are often poorly explained by factors such as practice demography or the socio-economic characteristics of the areas where practices are located (O'Donnell CA, 2000; Franks P et al, 1999).

The previous government placed increased emphasis on accountability in Primary Care Trusts, and, with the establishment of bodies such as the Commission for Healthcare Audit and Inspection, subsequently replaced by the Care Quality Commission, general practitioners have found that their referral patterns are under increasing scrutiny. Monitoring practices' referral rates, or other aspects of practice activity, runs the risk that practices that look after sicker populations may be unfairly scrutinised or penalised for having 'excessively' high rates. Hence, understanding the relationship between patients' morbidity and referrals is important. Fleming's 1991 paper on measuring morbidity in general practice stated that "In order to measure health care or to study the economics of the referral process, the requisite information must be related to morbidity". Despite this, few studies in the United Kingdom have examined this association (Forrest C et al, 2002). In contrast, studies in the USA and Canada have used case-mix systems to adjust for differences in patients' morbidity when investigating general practice referrals (Salem-Schatz et al, 1994).

In this chapter the Johns Hopkins ACG Case-Mix System (http://www.acg.jhsph.edu; Section 3.2) is used to investigate how well patient level measures of case-mix explain the variability in outpatient referral activity amongst general practices in the UK; to attempt to quantify the amount of variability in referrals that they explain; and to assess how well they can predict specialty referrals by general practice in the UK.

## 5.2 Methods

The dataset used for this application of the ACG system was from the General Practice Research Database (GPRD) on 1,323,611 patients registered in 211 general practices from England and Wales in 1997 (Lawson DH et al, 1998).

### 5.2.1   General Practice Research Database

The GPRD is the largest research database containing information on general practice morbidity and prescribing data in England and Wales. In 1996, the GPRD covered 5.6% of the population of England and Wales (Key Health Statistics in General Practice 1996). It was designed to record all prescriptions issued, the indication for all new prescriptions, and all "significant" events such as consultations resulting in a referral

and "events which the partner will require to be reminded of at a later date" (Mann RD et al, 1992) – for example, diagnoses such as cystic fibrosis and tuberculosis and information from hospital letters and coroners' reports.  The database records include information on demographics, disease incidence and referrals to a medical, surgical, or psychiatric specialist for patients in these practices.  GPs are not required to record minor conditions or follow up consultations for chronic conditions unless the consultation leads to a new treatment or referral.  The Office for National Statistics (ONS) evaluated the representativeness, quality and validity of the data by comparing the GPRD patient consulting rates for chicken pox, hay fever, asthma and diabetes with those from the 4th National Morbidity Survey from General Practice.  Results from this, and other studies, confirm the validity of the information recorded and indicate that the GPRD is a useful source of national morbidity data (Hollowell H, 1997; Jick H et al, 1991; Nazareth et al 1993; Herret et al 2009).  Comparisons of the age-sex distributions of patients in the GPRD database have been shown to be similar to national estimates (ONS, 1998).  The geographic distribution of practices participating in the GPRD is representative of the population of England and Wales, except for some under representation of inner city practices (ONS, 1998).

## 5.2.2   Morbidity groups

General practices contributing data to the GPRD followed guidelines for the recording of administrative, diagnostic, and referral data. Diagnoses were recorded using OXMIS or READ codes.

The OXMIS and READ codes were converted to ICD9 codes by clinicians and each patient was assigned an Adjusted Clinical Group (ACG) code, based on their age, sex, and diagnosis codes (Majeed A et al, 2001).  The ACGs were then grouped into 'resource utilisation bands' or RUBs (http://www.acg.jhsph.edu) following the same grouping mechanism used for the 'treated morbidity index' by Forrest C et al (2002). Higher RUB scores indicate sicker patients, a greater morbidity burden, and greater need for specialty referral; therefore RUB score is a proxy measure of patient morbidity. There were six morbidity groups, ranging from group 1 (healthiest) to group 6 (sickest).

Age was divided into four groups: children, young adults, older adults and elderly (0 to 15, 16 to 34, 35 to 64, and $\geq$ 65 years respectively).

### 5.2.3   Converting Read and Oxmis codes to ICD9 codes

The ACG coding system used in this research required diagnostic information to be input in the form of the International Classification of Diseases version 9 (ICD9) codes. General practices in England and Wales have traditionally used Read or Oxmis codes for recording patient diagnoses onto computer.  Therefore, in order to use the ACG system on UK data, the READ or OXMIS codes must first be converted to ICD9 codes. The ACG system software then assigns every patient's ICD9 code to one of 32 ADGs based on their expected health services resource consumption, and further assigns each patient to an ACG based on their age, sex and combination of ADGs over a one year period.  Ethics approval was obtained from the GPRD Scientific Ethics Advisory Group (SEAG, MHRA).

GPRD Medical and Patient datasets were obtained for 1997.  Each patients' date of birth, sex and any recorded diagnoses for all patients were included.  A table containing Read and Oxmis codes with their corresponding PCPS codes (i.e. codes assigned to unique Read and Oxmis codes in the GPRD database by UCL's department of Primary Care and Population Sciences) was merged with a table containing PCPS codes converted to ICD9 codes.  The conversion table was created by clinicians.  After mapping Oxmis and Read codes to ICD9, the coding table was checked manually to see which codes occur most frequently and whether there were any missing or odd looking codes.  Also, those Read codes resulting in no corresponding ICD9 code were checked to ensure that they were not diagnoses that had been missed.  The table containing the patient indicator and ICD9 codes was then merged with a table containing other patient information needed for conversion of codes to ACGs groupings (e.g. date of birth, sex). A program was written to allow the ACG software to read in the information and output it in the required format (patient id, date of birth, age, sex, ACG, ADG, morbidity grouping).  Output obtained from the ACG run was imported to Stata.  Any other relevant information (such as outcome of interest; and practice identifier) was merged with the ACG output.

### 5.2.4   Exclusions

Patients were excluded from analysis if they were registered with a practice for less than
180 days. Practices were excluded if less than 2% of their patients were recorded as
having had a referral to secondary care or if their deprivation code was missing.  After
these exclusions, the dataset consisted of 1,161,892 patients from 202 practices.

## 5.3 Statistical methods

The coefficient of variation (CV) is often used as a measure of variability in primary
care research (chapter 2).  The (Woolf adjusted) proportion of referrals for each practice
was estimated without and then with adjustment for covariates and the CV calculated as
the ratio of the standard deviation (SD) of the referral proportions to the mean of the
proportions (See appendix for description of Woolf adjustment and calculation of CV).
The process was carried out three times: firstly, making no adjustment for covariates;
secondly, adjusting for age and sex; and thirdly, adjusting for morbidity.  The
adjustment factors resulting in the smallest CV may be interpreted as explaining most of
the variation in the referral outcome.  Since most of those referred during the study
period only had one referral, the referral outcome was treated as binary in all subsequent
analyses.

Multilevel logistic regression models with random intercepts were fitted to the data,
thus taking into account the clustered nature of the data (patients within practices).  Four
models were fit to the data, each with different covariates: model 1 had no covariates;
model 2 adjusted for age group and sex; model 3 adjusted for morbidity; model 4
adjusted for age, sex and morbidity.  The fixed part of multilevel models allow
estimation of odds ratios adjusted for covariates, and the random part allows estimation
of the proportion of the unexplained between practice variation in referrals.

R-squared values estimating the proportion of the total variability in referrals explained
by each of the models were calculated using a method for multilevel models (Snijders &
Bosker, 1999).  Graphical comparisons were made between the observed referrals by

practice and those predicted by the models. Receiver Operating Characteristic (ROC) areas were calculated to assess the discriminatory ability of the models (Hanley JA et al, 1982). A value of 1 for the ROC area indicates that the associated model discriminates perfectly between patients with high and low referrals, while an area of 0.5 indicates that the model discriminates no better than chance. MlwiN v1.10 software was used for multilevel modelling (Rasbash J et al, 2000); Stata 8.0 was used for all other analyses (StataCorp, 2003).

## 5.4 Results

The characteristics of the study participants are given in Table 8. The analysis dataset consisted of 1,161,892 records from 202 general practices for the year 1997. The median (range) practice size was 5,055 (1,364 to 14,587). There were a similar proportion of males (49.37%) and females in the study sample. The percentage of males ranged from 40.7% to 58.7% across practices. Over a third (38%) of all patients were in the 35 to 64 year age group, while the oldest category of 65 years and above contained the lowest (15.8%) proportion of patients. The mean age was about 39 years and ranged from 26 to 47 years across practices. There was a marked difference in the distribution of patients across morbidity class. The percentage of patients in each group decreases with increasing morbidity. The healthiest group has 32% of all patients while the sickest group has only 7%. About one in seven (14.7%) of all patients had at least one referral in the year period, only 2% of all patients had two referrals and only 0.4% had three or more referrals (Table 9). The referral response was therefore treated as binary.

The median percentage of patients referred by practice was 14.8% (range 2.4% to 24.4%). Table 10 summarises the distribution of patients and the percentage referred in each age, sex and morbidity group. The overall percentage of referrals increases steadily with increasing age group. 7.5% of patients aged 0-15 years were referred compared to 21.1% of patients in the 65 plus age group. There is an even stronger increase in overall referrals with increasing morbidity. Females had more referrals than males (17.1% vs 12.2%).

Figure 8 illustrates the percentage of patients in the ten most common ACGs. Note that, unlike ADGs, ACGs are mutually exclusive (each patient is only assigned to one ACG). Almost one in four patients does not consult in the study time period.  More than one in ten patients are in the next most popular ACG, which is composed of patients aged 6 and above with acute minor conditions.  The remaining eight of the top ten ACGs mainly consist of patients with acute major and acute minor conditions.   Those ACGs with the highest proportion of patients referred are mainly composed of either patients with a large number of ADGs and/or some major ADGs, or pregnant patients with a large number of ADGs.  The least referring ACGs include: patients with asthma and no other ADGs (ACG 700); acute minor ACGs (ACGs 100-300); Likely to recur, without allergies (ACG 600); eye/dental (ACG 1100); and Chronic medical, stable (900).  Many of these are acute conditions where patients are more likely to present at Accident and Emergency departments of hospital than attend general practice during an episode of illness.

Almost one in three of all patients (31%) are grouped among the 16 ACGs that are made up of one type of illness alone (ACGs 0100-1600).  The ACGs running from 1800 to 4100 (24 ACGs) comprise of a mix of patients with two to three different types of illnesses. Over one quarter (30.79%) of all patients were assigned to one of these 24 ACGs. Another 21% of all patients are in the four main groups comprising many mixed states (ACGs 3800-5070).  Just over three percent (3.07%) had pregnancy ACGs.  Less than a third of a percent (0.31%) of patients are of unstable chronic condition (ACG 0800), while under two percent (1.86%) are of stable chronic condition (ACG 0900). 2.36% have problems of a psychosocial nature (ACGs 1300-1500, 2500-2700, 3500 and 3700).

The results for models 1 to 3 are presented in Table 11.  The unadjusted CV (Model 1) was estimated at 34%, the age & sex adjusted CV (Model 2) was 33% and that adjusted for morbidity (Model 3) was 32%.  These findings indicate that the standard deviation of the referral proportions for practices was about a third of the mean practice

proportion of referrals.  The results imply that the age-sex and morbidity covariates explain only slightly more of the variation in referrals between practices.

Table 12 shows the model-based odds ratios and 95% confidence intervals for referral for Models 2 to 4. The odds of referral increase significantly with increasing age (Model 2) and they increase more strongly with increasing morbidity (Model 3).  The odds of referral for a patient in the sickest group relative to a healthy patient is 36.2 (95% CI 15.6 to 84.1). The results for morbidity in model 4 were similar to that in model 3 but the effect of age was no longer statistically significant and therefore the focus was on comparisons between Models 2 and 3 for the remainder of this study of general practice referrals.

The variation in referrals can be considered to have two components: variation between and variation within practices.  The R-squared values show that patient morbidity explains considerably more of the total variability in referrals than patients' age and sex alone (30.4% vs 5.3%, Table 12). The variation in referrals between practices is highly significant, although it is only responsible for 4.5% of the total variability for Model 2 (age-sex), and reduces to 3.6% for Model 3 (morbidity).  The variation in referrals within practices after adjustment for age and sex alone is 90.2%, and is considerably less (66.1%) after adjusting for morbidity.   The overall conclusion to be made from these results is that morbidity explains 25% more of the total variability in referrals than age and sex; however most of the variability that morbidity explains is at the level between patients within practices rather than between practices.

The model residuals were approximately normally distributed and there was no evidence of extra-binomial variation or homogeneity of variance.  All of the covariates tested in each of the models were highly significant. The results for model 2 in Table 12 illustrate the increasing odds of referral for with increasing age. The odds of referral in the 65 plus age group is more than three times that of the youngest age group (0 to 15 years).  The odds of referral for females is higher than for males but this is not statistically significant.  Figure 9 and Figure 10 both compare the predicted and observed percentages referred by practice.  If the model predictions are accurate, the

points will lie close to the line of equality. The range of predicted practice referrals is narrow in Figure 9 (11.6% to 15.2%), representing Model 2.  Model 3 (Figure 10) is a marked improvement on model 2, with predicted referrals ranging from 6.2% to 18.8%, closer to those of the observed range.  However, even after adjusting for morbidity, substantial departures from the line of equality remain.

The area under the ROC curve was 0.616 (95% CI 0.614, 0.617) for Model 2 and 0.768 (95% CI 0.766, 0.769) for Model 3. These show that the morbidity model discriminates significantly better between patients with and without referrals than the model with age and sex alone.

## 5.5 Discussion

Even after excluding practices with very low percentage of referrals (less than 2% of patients referred annually), there was still a 10-fold variation between the practices with the highest and lowest proportion of patients referred. Patient morbidity explains considerably more of the total variability in referrals than patient's age and sex.  The results also show that, after adjusting for patient case-mix, most of the unexplained variation in how patients are referred is occurring between patients within practices rather than between practices; and that patient morbidity explains substantially more of this within-practice variation (around 24% more) than age and sex.  These findings support previous research suggesting that the amount of variability between practices may be less than that implied by previous studies based on aggregate information (Davis P et al, 2002).

Many previous studies have reported wide variations in general practice referral rates. Studies that have adjusted for age and sex have typically found that the observed variation decreased by less than 10% (Davis P et al, 2002), in line with the findings of this research.  Adjusting for other factors, such as socio-economic deprivation or practice characteristics, may improve on this, but most of the variation in referral rates still remains unaccounted for (O'Donnell CA, 2000).  The role of diagnostic-based case-mix measures in explaining variation in general practice referrals has not

previously been investigated in the United Kingdom.  Studies from the USA, however, suggest that it is better than age and sex in explaining variation in referrals, again in line with these findings (Weiner JP et al, 1991).

This was a large study that used data from general practices contributing to the GPRD, an extensively validated high quality database (Hollowell J, 1997; Jick H et al, 1991). The average size of the practices is greater than the national average, but it is reasonably well representative of the UK population (ONS, 1998).  The database provided us with individual level data, and the clustering of patients within practices is incorporated in the statistical models, thus avoiding the 'ecological fallacy' (i.e. finding associations at the population level that may not hold at the individual level).  This was the first large-scale study in the United Kingdom to control for diagnostic-based morbidity groupings specifically designed for use in primary care when examining variation in specialist referrals in general practice. The ACG system was derived for use in the USA and is likely to need some further adaptation to maximise its utility in the United Kingdom (Majeed A et al, 2001).  Finally, the system depends on diagnostic codes recorded by general practitioners and is therefore a proxy measure of patient morbidity. Hence, differences in the way that general practitioners treat chronic diseases and code information on their practice computers could introduce bias.

The results of this study of GP referrals support similar research conducted by Peter Davis et al, 2002 in New Zealand.  Davis et al investigated three resource use outcomes of prescribing, ordering of investigations, and recommendation of a future follow-up visit.  They also used logistic regression multilevel techniques and applied a similar method to estimate the total variance explained (R-squared) and the variance between practices (ICC) to the one used in this study (Snijders & Bosker, 1999).  Their findings were that, even after adjusting for practice and patient measures (including diagnostic-based case-mix), less than a third of the total variability in the outcomes was explained by the models.  Less than around 10% of the unexplained variation was due to differences between doctors (they had access to doctor level data), while the remainder was at the patient level.

However, the similar findings of this work and Davis' work contradict research findings produced by Franks P et al (1999). Franks also used the ACG system in adjusting for patient case-mix when comparing referrals across practices. Similar to us, Franks found that the case-mix predicted referral rate was only minimally affected by case-mix. He found that most (93%) of the variation was due to differences between doctors, and hence concluded that referrals were a largely doctor driven practice. However, because of limited computer specifications, it was not possible to apply multilevel logistic regression models, and therefore he applied methods meant for continuous outcomes, despite the fact that his outcomes were binary. This is likely to give biased estimates since the probability of referral in the GPRD is considerably lower than 0.5. It is impossible to know how different their results might have been had they been in a position to apply the correct methodology.

Since access to data at doctor level (only at practice and patient levels) was not possible, one can only hypothesise as to how the results might be affected. It is likely that there would be some redistribution of the proportion of variance attributable to the patient and practice level (Moerbeek M (2004)). However, in practice such information is difficult to collect, since patients are often seen by more than one doctor or intern on different occasions. Even for the data in the Franks study above, which discusses doctor level variation, the (primary) doctor who has seen the patient more than 50% of the time is recorded as the doctor for the purpose of the study, and hence the decisions of other doctors is likely to be represented as being the decision of the primary doctor for a large proportion of visits.

Variation in general practice referrals is an important issue for clinicians, managers, patients and politicians. Age and sex seem to explain relatively little of this. Although morbidity explains substantially more variation, most remains unexplained. The unexplained variation seems to occur largely at the within practice level, so that the GP decision of whether to refer a patient varies, even for patients with similar age, sex and morbidity. This may be due to patient factors that cannot be easily measured, such as the patient's insistence on referral (Haste F, 2002) or their level of confidence in their general practitioner (Armstrong D et al, 1991; Greenhalgh T, 1998). The experience of

the doctor in treating the condition for which the patient was referred and the local provision of specialist and diagnostic services and outreach clinics may also be important (Davis P et al, 2000; Haste F et al, 2002).

The findings suggest that any investigation of how general practices refer to specialist services should be interpreted cautiously even after adjustment for age and sex. In contrast, measures such as diagnostic-based morbidity are likely to give more useful information, but caution should be exercised even after adjusting for these, as much of the variability in referrals is likely to remain unexplained. Similar caveats are likely to apply to other areas of general practice, such as prescribing costs and mortality rates. The association of these measures with case-mix should be explored to ensure that practices that seem to have a high intensity of resource use, that have high death rates, or that do not achieve quality targets are not inappropriately scrutinised or penalised for this.

## 5.6 Conclusions

Diagnostic-based morbidity measures, produced using the Johns Hopkins ACG Case-Mix System, appear to be useful in describing and explaining variation in specialist referrals in general practice. Hence, their application in primary care needs to be explored further, for example, in performance management and distribution of resources to general practices.  However, more research is needed to assess whether the above results hold across different general practice populations.  If this is truly the case, then the cost-effectiveness of implementing and administrating such a system should be evaluated.  For now, any investigation of specialist referrals from general practice should be interpreted cautiously, even after adjustment for age, sex and morbidity, as it appears from this research that much of the variability remains to be explained. Identifying and understanding the other reasons for variation in referrals will help to ensure that NHS resources are used efficiently, interventions put in place at an appropriate time and that patients gain appropriate access to specialist services. Other areas of primary care activity, such as performance and mortality measures, should also be interpreted cautiously for similar reasons. Even after adjustment for case-mix, there

are other important factors that remain unaccounted for, which, if they were possible to measure, could further alter performance and mortality measures.

**Table 8  Characteristics of General Practice Research Database study participants**

| Characteristic | Percent (*UNLESS OTHERWISE INDICATED*) |
|---|---|
| Number* | 1,161,892* |
| | |
| Male | 49.37 |
| | |
| Age group | |
| Mean (SD)** | 38.5 (22.9)** |
| Median (range)*** | 37 (0 to 113)*** |
| 1 (0 to 15 years) | 20.13 |
| 2  (16 to 34 years) | 26.03 |
| 3 (35 to 64 years) | 38.07 |
| 4 (65+ years) | 15.76 |
| | |
| Morbidity class | |
| Median (range)*** | 2 (1 to 6)*** |
| 1 (healthiest) | 32.13 |
| 2 | 25.62 |
| 3 | 18.65 |
| 4 | 12.08 |
| 5 | 4.50 |
| 6 (sickest) | 7.01 |
| | |
| Referrals | |
| Mean (SD)* | 17.6 (0.464)** |
| Median (range)*** | 0 (0 to 20)*** |
| 1 or more | 14.7 |

* Number
** Mean (SD) of the practice means
*** Median (range) of the practice mean

**Table 9  Count of number of referrals by patient**

| Number of referrals | Number of patients | Percent |
|---|---|---|
| 0 | 990,915 | 85.28 |
| 1 | 143,155 | 12.32 |
| 2 | 23,259 | 2.00 |
| 3 | 3,775 | 0.32 |
| 4 | 630 | 0.05 |
| 5 | 114 | 0.01 |
| 6 to 20 | 44 | <0.004 |

**Table 10  All patients and percent patients with at least one referral by age, sex and morbidity**

| Age(yrs) | Patients | Referrals (%) | |
|---|---|---|---|
| 0-15 | 233,984 | 17,599 | (7.5) |
| 16-34 | 302,498 | 40,509 | (13.4) |
| 35-64 | 442,364 | 74,179 | (16.8) |
| 65+ | 183,136 | 38,690 | (21.1) |
| Total | 1,161,892 | 170,977 | (14.7) |
| | | | |
| Sex | Patients | Referrals (%) | |
| Male | 573,677 | 70,225 | (12.2) |
| Female | 588,215 | 100,752 | (17.1) |
| Total | 1,161,892 | 170,977 | (14.7) |
| | | | |
| Morbidity class | Patients | Referrals (%) | |
| 1    Healthiest | 373,298 | 8,256 | (2.2) |
| 2 | 297,705 | 31,688 | (10.6) |
| 3 | 216,707 | 36,684 | (16.9) |
| 4 | 140,389 | 39,756 | (28.3) |
| 5 | 52,324 | 19,255 | (36.8) |
| 6    Sickest | 81,469 | 35,338 | (43.4) |
| Total | 1,161,892 | 170,977 | (14.7) |

**Table 11  Coefficient of variation for models with referrals as outcome**

| | Coefficient of variation (%) |
|---|---|
| Model 1 Unadjusted | 34 |
| Model 2 Age & Sex | 33 |
| Model 3 Morbidity | 32 |

**Table 12  Results of models and percentage of variation explained**

| Sets of predictors are:<br>1) No predictors (empty model)<br>2) Age & Sex<br>3) Morbidity class<br>4) Age, Sex and Morbidity class | | | | |
|---|---|---|---|---|
| | Model 1<br>Odds Ratio (95% interval) | Model 2<br>Odds Ratio (95% interval) | Model 3<br>Odds Ratio (95% interval) | Model 4<br>Odds Ratio (95% interval) |
| Male | | 1 | | 1 |
| Female | | 1.5 (0.7, 3.2) | | 1.1 (0.5, 2.6) |
| | | | | |
| Age 16 to 34 yrs | | 1.9 ( 0.9, 4.3 ) | | 1.4 (0.6, 3.3) |
| Age 35 to 64 yrs | | 2.5 (1.1, 5.6 ) | | 1.4 (0.6, 3.3) |
| Age 65 onwards | | 3.2 (1.4, 7.2 ) | | 1.2 (0.5, 2.7) |
| | | | | |
| Morbidity 1 | | | 1 | 1 |
| Morbidity 2 | | | 5.4 (2.3, 12.5) | 5.4 (2.3, 12.6) |
| Morbidity 3 | | | 9.3 (4, 21.6) | 9.2 (4, 21.4) |
| Morbidity 4 | | | 18.2 (7.8, 42.4) | 17.4 (7.5, 40.4) |
| Morbidity 5 | | | 27.4 (11.8, 63.8) | 26.3 (11.3, 61.2) |
| Morbidity 6 (sickest) | | | 36.2 (15.6, 84.1) | 35.3 (15.2, 82.0) |
| Variation | % | % | % | % |
| Unexplained at practice level | 4.6 | 4.5 | 3.6 | 3.6 |
| Unexplained at patient level | 95.4 | 90.2 | 66.1 | 65.6 |
| R-squared: Proportion of total variance explained | 0 | 5.3 | 30.4 | 30.8 |

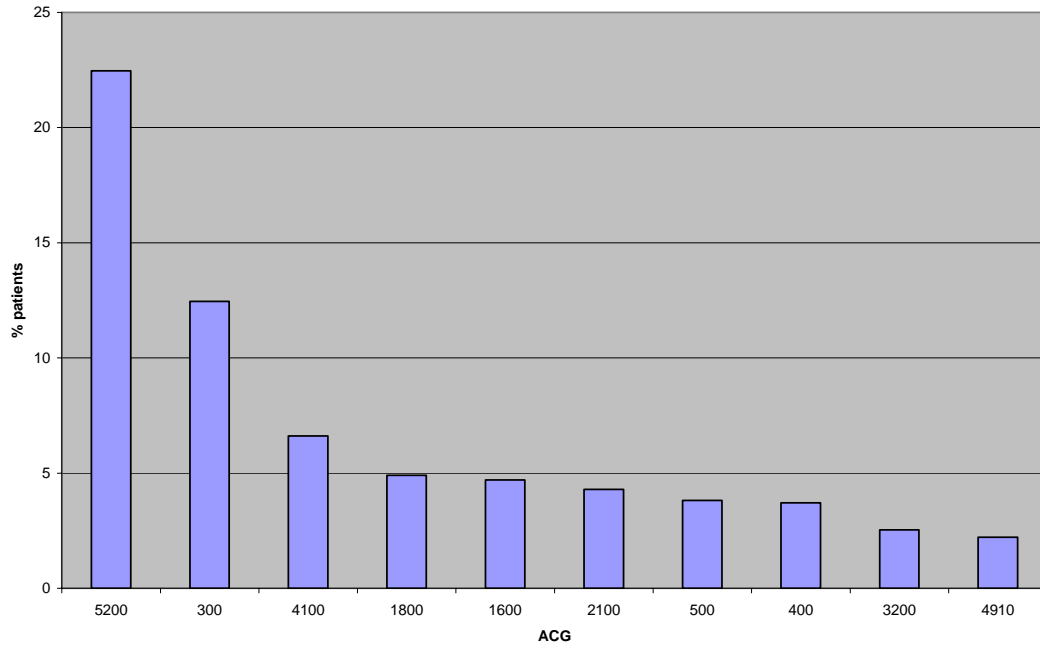**Figure 8  Percentage GPRD patients in ten most common ACGs**

**Figure 9  Observed vs predicted referrals by practice for model with age & sex as covariates**
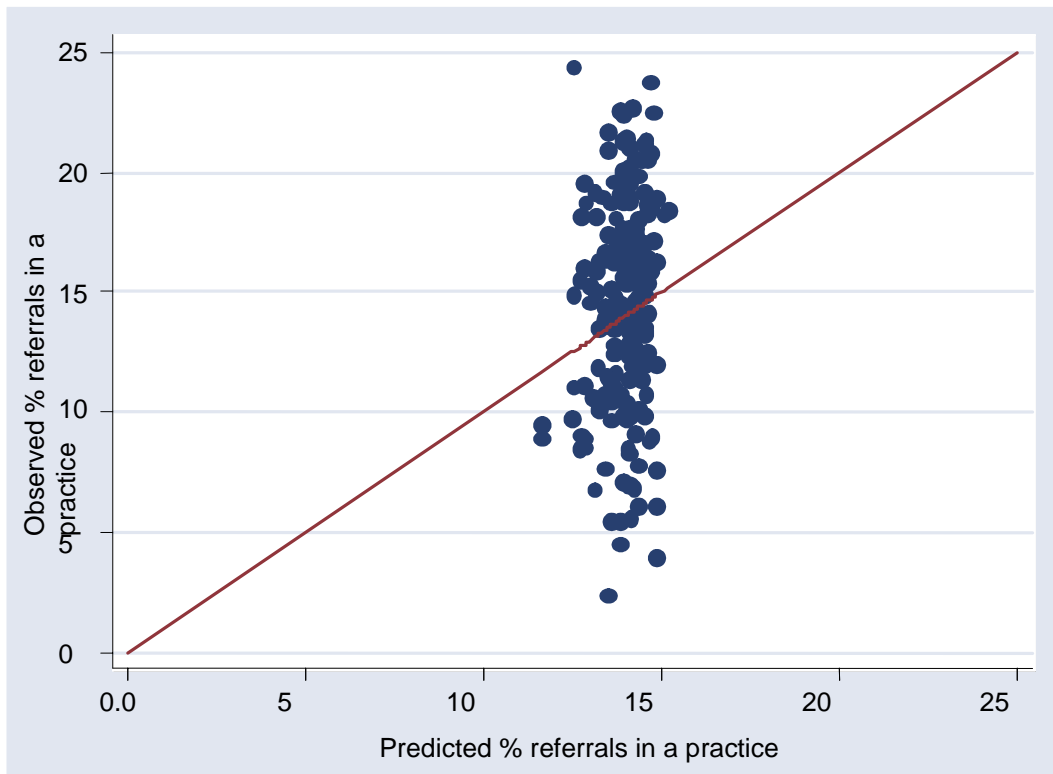
**Figure 10  Observed vs predicted referrals by practice for model with age, sex & morbidity as covariates**
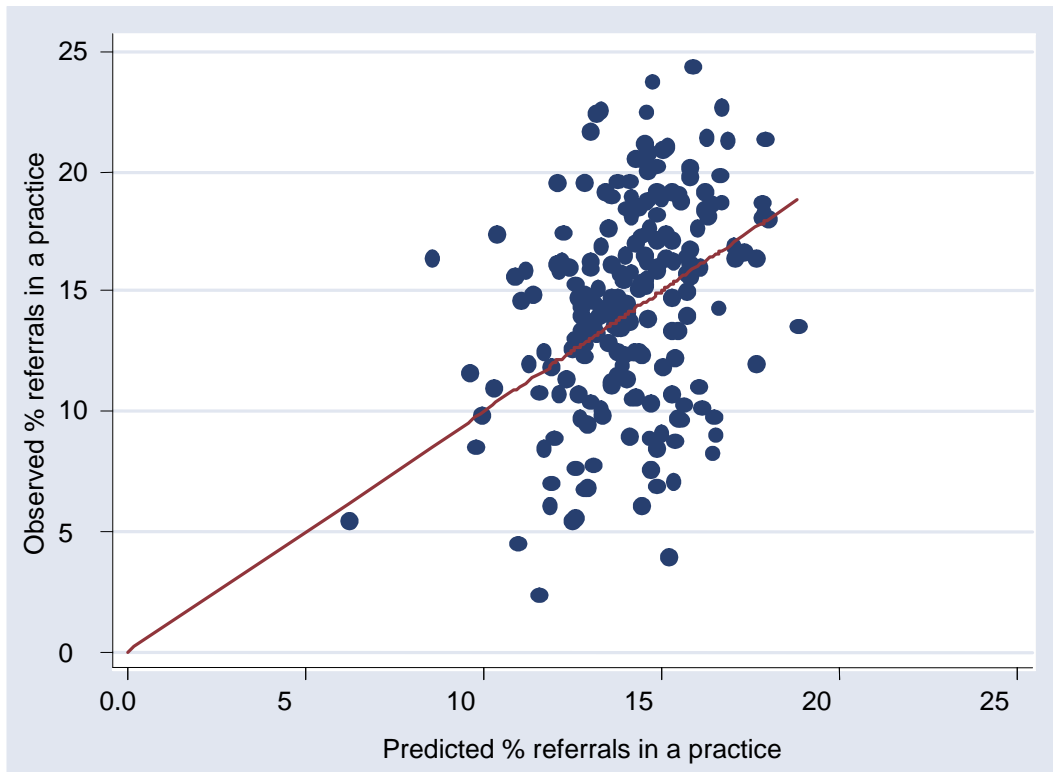
# Chapter 6    Prescribing

## 6.1 Introduction

The prescribing costs of general practitioners in the United Kingdom have increased rapidly in recent years, with a 60% real terms increase in spending and a 55% increase in the number of items dispensed between 1996 and 2006 (Information Centre for Health & Social Care: http://www.ic.nhs.uk).  Prescribing by general practitioners now costs around £7.8 b (9.9 euro; $15.3) a year, about 10% of the National Health Service's expenditure in England (Scoggins et al, 2006). General practitioners' prescribing decisions are coming under increasing scrutiny, with considerable pressure to prescribe cost effectively (Majeed A et al, 1999).  The development of new drugs, enhanced indications for existing drugs (such as statins), more rigorous management of chronic diseases, and the ageing of the population of England will all continue to increase the cost and volume of prescribing in primary care (NAO 2007).  Prescribing budgets for Primary Care Trusts are currently allocated using a formula that incorporates certain weightings for inequalities (based on Disability Free Life Expectancy; Limiting Long Term Illness) and needs (based on age, sex, Limiting Long Term Illness, Disability Living Allowance claimants, Low Income Scheme Index and Low birthweight births) (DH Exposition book 2009-10 and 2010-11).  Prescribing budgetary allocations from PCTs to general practices are, however, still largely based on historical prescribing patterns (Majeed A et al, 1996, (DH Exposition book 2009-10 and 2010-11).  When these patterns do not reflect clinical need, historical inequities in resource allocations are perpetuated.

To overcome these problems some Primary Care Trusts use needs based models to determine indicative prescribing budgets for general practices.  A limitation of these models is that they are largely based on the demographic profile of a practice population, sometimes with a weighting for local characteristics taken from the census.

The models do not generally contain any direct measure of morbidity within a practice. Previous research on such models has generally shown that they are poor predictors of prescribing costs in practices; and general practices with high prescribing costs often come under considerable pressure to reduce these costs (Favato G et al, PLoS ONE 2007). Consequently, general practices that look after populations with higher burdens of morbidity may be unfairly scrutinised or penalised for having high prescribing rates. Variation in prescribing could be due to differences in the case-mix of patients registered with the general practice, socioeconomic factors, or inefficient or inappropriate prescribing. More sophisticated models to better understand these variations are needed. Prescribing models that incorporate morbidity could be used to help predict expenditures for budgetary planning and to separate practices that have high prescribing costs because of a high burden of disease from those that have high costs because of inefficient prescribing. These models could also help identify practices that appear to have inappropriately low prescribing rates for their practice's morbidity burden to further investigate whether they may be under-treating patients.

In this chapter, the ACG system (Section 3.2) was used to investigate how well patient level morbidity based measures of case-mix explain the variability in prescribing among general practices in the UK. This is the only case-mix system specifically designed for use in primary care and it has been widely used in studies examining variations in primary care practice. (Starfield B et al 1991; Weiner JP et al, 1991; O'Sullivan C et al, 2005; O'Sullivan C et al, 2004).

## 6.2 Methods

Data was obtained from the UK General Practice Research Database (Lawson DH et al, 1998). General practices participating in the database follow set guidelines for the recording of clinical and prescribing data and submit anonymised patient based clinical records to the database at regular intervals. The accuracy and comprehensiveness of the data recorded in the database has been documented previously (Jick H et al 1991; Hollowell J et al 1997). The variables collected by the database include age; sex; registration details; medical diagnoses (Read and OXMIS codes) that are part of routine care or resulting from admissions to hospital, consultations or emergency care; referrals;

laboratory test; and prescriptions issued for each patient. Although the prescriptions issued by specialists are not picked up in the General Practice Research Database, most prescriptions for chronic disease in the UK are issued by general practitioners.  Data for the year 2001 were obtained only for practices that met the "up to standard criteria", a quality marker set on the basis of internal consistency of the practice, completeness of longitudinal recording, and compliance with the recording guidelines of the General Practice Research Database (www.gprd.com ).  All practices provided one full year's worth of data.  Patients were excluded if they were registered with a practice for less than 180 days.  Read and OXMIS codes were converted codes from the *International Classification of Diseases*, ninth revision (Bindman AB et al, 2007; Forrest CB et al, 2003) using a lookup table.

To construct the morbidity groups, the Johns Hopkins adjusted clinical group system software was used to initially assign the patients into one of the 81 mutually exclusive Johns Hopkins adjusted clinical groups, on the basis of age, sex, and a combination of recorded diagnoses over a one year period.  These groups were then assembled into six mutually exclusive "Resource Utilisation Bands (RUBs)" using the range of diagnoses pertaining to each patient.  These six categories are constructed by the software according to patients' expected resource use on the basis of a nationally representative database of 2 million patients aged less than 65 years in the United States.  For example, a patient with uncomplicated type 2 diabetes would be placed in group 2, whereas a patient with type 2 diabetes, heart failure, cellulites, and chest pain would be placed in group 5 (www.acg.jhsph.edu ).  In this paper these six groups represented morbidity groups of patients, with group 1 being the healthiest patients and group 6 the sickest. Age was grouped as children (0-15 years), young adults (16-34), older adults (35-64), and adults of pensionable age (>=65 years).

Using the rule of 10 events or observations required per coefficient estimated in a model and adjusting for the design factor (using intracluster correlation coefficient of 0.02 for prescribing and average cluster size of 8000), this study required a total of 14000 events or observations to estimate the models' coefficients with adequate precision (Machin D et al, 2005).  After exclusions, the dataset used from the General Practice Research Database had more than sufficient numbers of events or observations.

## 6.3 Statistical methods

A two level Poisson model with random intercepts was used to investigate the association between age, sex, morbidity, and the number of prescriptions issued (Goldstein H et al, 2003) (outcome and covariates were considered at the patient level), after accounting for clustering within the general practices.

A number of methods were used to estimate the extent of variation in prescribing. First variation at the practice level that is explained by the covariates was estimated using an adjusted R-squared measure based on a linear regression model (Weisberg S et al, 2005). This was a practice level analysis in which the outcome was the mean number of prescriptions issued by each practice. The practice mean age, and proportions for each sex and morbidity groups, were used as predictors. Then the variation in prescribing was partitioned into practice and patient levels using an R-squared measure derived from a two level logistic regression with random intercepts (Snijders TAB et al, 1999). For this purpose the number of prescriptions was converted to a dichotomous response according to whether or not a patient had received a prescription. As some information may be lost owing to collapsing number of prescriptions to categories or mean, sensitivity analyses were carried out to check the consistency of the results using another type of R-squared measure estimate from a two level linear regression model with random intercepts (Snijders TAB et al, 1999). A square root transformation of the number of prescriptions issued as the response was used to satisfy the assumptions of normality required by the linear regression model. The R-squared measures obtained from all three methods were compared across models fitted with no covariates, with age and sex, and with age, sex, and morbidity.

To assess how well the models discriminated between patients who had received a prescription and those who had not, the receiver operating characteristic areas from the logistic model were calculated (Hanley JA et al, 1982). The receiver operating characteristic area represents the proportion of patient pairs that is correctly ranked by a model according to the prescribing status of the patients.

Residual plots were used to investigate assumptions of normality of residuals required by the multilevel models. MLwiN v 2.02 software (Rasbash J et al, 2000) and Stata version 9.2 were used for the statistical analyses (StataCorp, 2005).

### 6.3.1   Results

Information on age, diagnosis, and prescribing was complete. Twelve patients with no recorded sex information were excluded from the analysis. Patients registered with a practice for less than 180 days were also excluded. After exclusions, 129 practices with 1,032,072 patients were eligible for inclusion. The median time that a patient had been registered with a general practitioner in 2001 was 11 years. Overall, 49.3% of the patients were male and 50.7% were female. Sixty four percent of patients were issued a prescription at least once during 2001. The median percentage of patients issued a prescription in the study year was 65% (90% range 11% to 75%). The median number (90% range) of prescriptions issued to a patient across the 129 practices was 2 (0 to 18). The median total number of prescriptions issued across the 129 practices was 9852 (3508 to 14 589).

The number of patients in the two sickest morbidity groups was relatively small and therefore these two groups were combined in all subsequent analyses. The results from Table 13 show that age, sex, and morbidity vary across practices along with the number of prescriptions issued across all practices for each of these groups. The sex distribution of the patients was similar across the practices. The age and morbidity distributions of patients varied, however, particularly for those in the oldest age group (>=65 years) and for morbidity groups 4-6. The median number of prescriptions issued increased with age and morbidity groups and was higher for females. The number of prescriptions issued by the practices varied considerably, with the highest variation occurring in patients aged 65 and over and in the sickest morbidity groups.

The number of prescriptions issued to a patient was strongly associated with the patient's age and morbidity (Table 14; P<0.001), increasing steeply with age and morbidity. Several scenarios below illustrate the relations observed in these models. The expected number of prescriptions for boys and girls aged 0 to 15 are estimated to be

1.6 and 2.2, respectively, whereas the expected numbers for men and women aged 65 or more are 9.2 and 12.7. For the healthiest boys and girls aged 0 to 15 the expected number of prescriptions is 0.05 (same for both). The corresponding values for the least healthy girls and boys are 6.2 and 6.8. The expected numbers of prescriptions for the least healthy men and women aged 65 and over are 21.1 and 23.3.

Table 15 presents the results on the extent of variation explained in prescribing from the practice level analysis. Adding morbidity explains considerably more of the variation in prescribing between practices than age and sex. This result is supported by the patient level analysis presented in Table 16 where variation is split into practice and patient levels. The inclusion of morbidity explained considerably more of the total variability than patients' age and sex alone (80% v 10%). Of the total variation, only 0.1% remained unexplained at the practice level and 19% remained unexplained at the patient level, after adjusting for age, sex and morbidity. When adjusting for age and sex the corresponding values are 4% and 86%. The results show that most (96%) of the total variation in the prescribing outcome occurs at the within practice level. The extent of variation explained in prescribing based on the sensitivity analyses was 60% at patient level and 74% at practice level when morbidity was included and 20% and 6% when only age and sex were included.

The receiver operating characteristic area for a model with age and sex was 0.648 (95% confidence interval 0.647 to 0.649), which increased to 0.972 (0.971 to 0.972) when morbidity was included. Thus morbidity significantly improved the ability of the model to discriminate between patients who had received prescriptions and those who had not.

# 6.4 Discussion

Patient morbidity explains considerably more of the variability in prescribing than patients' age and sex alone. About 4% of the total variation is due to differences between practices and most of the variation is due to differences between patients within practices.

## 6.4.1   Comparison with previous studies

Many studies in the UK and elsewhere have shown that prescribing in general practice varies considerably, with threefold to four fold variations commonly seen even after practices with outlying prescribing rates are excluded. Statistical models from those studies based in the UK have not included direct measures of morbidity and have generally explained only a small proportion of this variation.

Other than the morbidity burden of a practice, other factors that could influence prescribing rates include deprivation, doctors' knowledge, professional experience, role perception, and time pressures; the number of doctors in the general practice; and patients' expectations of receiving a prescription and their demands (Carthy P et al, 2000; Watkins C et al, 2003; Webb S et al, 1994; Britten N et al, 1997; Cockburn J et al, 1997; Ashworth M et al, 2007).

Examples of studies that do employ multilevel modelling techniques and ACGs in examining variation in prescribing practice patterns include work by Davis P and Gribben B (1995). Several of their papers are based on a survey representing a 1% sample of GP visits (around 10,000) in New Zealand at two time points. In a study investigating prescribing patterns, they control for patient, diagnostic and practitioner variables and conclude that these improve the predictive power of the model, but do not reduce the extent of variability between practitioners in prescribing. This work raised questions about the role of clinical uncertainty and professional autonomy in the doctor's role in relation to prescribing medication.

Further research by Davis P et al (2000) explores economic vs health services research theories on variation in medical practice where health economists stress the influence of income incentives while health services research emphasise clinical ambiguity in doctor's decisions. Income incentives, doctor agency and clinical ambiguity (measured as local doctor density, practitioner encounter initiation and diagnostic uncertainty respectively) are examined in relation to prescribing, test ordering and doctor request for follow-up. Davis P et al found no relationship between competition and decision making; that doctor initiated follow up consultations were associated with lower rates of intervention, and that diagnostic uncertainty is associated with higher investigations and follow-up. They concluded that, for the variables studied, a clinical, rather than economic, model of doctor decision-making provided a more plausible interpretation of variation in rates of clinical activity in general practice.

Davis further investigated the variability between doctors in their clinical activity, again measured as prescribing, ordering of investigations and doctor-initiated follow-up (Davis et al, 2002). They found large variation between doctors in each of these measures, even after adjusting for case-mix, patient and doctor variables. These variables explained between 15% and 29% of the total variance in the three outcomes. However, the variance components concluded that only between 4% and 11% of the remaining variability was at the doctor level. The work was extended by focussing on one diagnosis only, upper respiratory tract infection. Here, they found that the proportion of total variance explained by the model decreased, although the residual variance at the doctor level increased.

### 6.4.2   Strengths and limitations

In general, the amount and quality of diagnostic data collected in primary care is varies widely although prescribing data is of better quality (Thiru K. et al (2003)). This study used data from the General Practice Research Database, which has been extensively validated and shown to be of high quality. The practices submitting information to the database are reasonably representative of the age and sex profile of the UK population, with some under-representation of inner city practices. The average size of the practices is greater than the national average (Hollowell J et al, 1997; Jick H et al, 1991). In contrast with many previous studies of variation in prescribing, this study used data at

individual patient level rather than an ecological design.  The ecological design has the limitation of drawing inferences at the individual patient level solely on the basis of aggregate statistics.  This study also controlled for diagnosis based morbidity groupings specifically designed for use in primary care when examining variation in prescribing.

Among the limitations of the study is that the ACG system was developed for use in the United States and therefore might need some further adaptation to maximize its utility in the UK.  It is, however, now been used for an increasing number of UK based studies.  Finally, the ACG system depends on diagnostic codes recorded by the general practitioners during consultations.  Differences in the way that general practitioners record similar conditions on their practice computers could introduce bias into the estimates of their practices' morbidity scores.

### 6.4.3   Implications for practice

In this chapter a measure of patient morbidity was used to explain variation in general practice prescribing.  Including morbidity in the model considerably improves its explanatory power and therefore its potential utility for monitoring prescribing in general practice and the allocation of prescribing budgets.  With increasing availability of tools based on using general practice electronic medical records, computerised clinical data for activities such as assessment of morbidity is increasingly available.

This study shows how strong morbidity is in explaining variation in the number of prescriptions issued and in determining which group of patients is most likely to receive prescriptions. In practice, there are several ways that these findings could be implemented in a simple manner.  For example, patient age, sex and diagnostic information could be used to assign each patient to one of the morbidity groups and then one could allocate budgets according to the number of patients in each of the groups.  An example where a PCT might implement the findings might be to again assign each patient to one of the morbidity groups and to compare the practices in the PCT to see which practices are outliers i.e. have a higher burden of illness/elderly etc.  The findings might be used by the PCT, for example, to provide extra services or to target particular interventions.  Researchers at Keele University demonstrated the variation in volume of clopidogrel (defined daily doses per 1,000 age and sex weighted patients) prescribed by

PCTs in England as an average (range) of 149 (61 to 341) (August 2005 to July 2006) http://www.official-documents.gov.uk/document/hc0607/hc04/0454/0454.pdf. This study could be further weighted by morbidity to investigate whether the morbidity burden of PCTs further explains some of the variability. Another study from Keele University demonstrates the variation among English PCTs in diabetic prevalence versus diabetic test strips per thousand patients. One could similarly compare prescribing patterns across the various morbidity groups for practices or PCTs to identify areas that are significantly different from their peers for further investigation http://www.nao.org.uk/system_pages/search.aspx?&terms=technical+supplement.

The focus of this study was on prescriptions issued. Each prescription issued might, however contain several items and contain drugs for very different therapeutic areas. Hence, further work is required to investigate the association between morbidity and total prescribing volume (measured by number of items prescribed) and costs and how well morbidity explains variation in prescribing in specific therapeutic areas. The use of such patient based measures of case-mix could then be explored in setting budgets for health services, examining how efficiently health services are being used, and to produce measures of clinical performance and quality of care adjusted for case-mix.

## 6.5 Conclusions

Inclusion of a diagnosis based patient morbidity measure into prescribing models can explain a larger amount of the variability at both patient and practice levels. The use of patient based case-mix systems should be explored further when examining variation in prescribing patterns between practices in the UK. Some areas for future work include: investigating the association between morbidity and total prescribing volume and costs; examining how well morbidity explains variation in prescribing in specific therapeutic areas. Further modelling work includes examining the effect of including general practice characteristics in the model with patient level data. There is potential for development of a tool to help general practices and PCTs to predict their prescribing activity and cost. In the longer term, case-mix systems may prove useful in fairer allocation of budgets and in the production of case-mix adjusted measures of performance.

**Table 13  Number of patients and prescription issued by age, sex, and morbidity**

| Variable | Number of patients (median %*, (90% range) across practices) | Annual number of prescriptions (median (90% range) across practices) |
|---|---|---|
| | | |
| Age (yrs) | | |
| 0-15 | 202,303 (19.0, 15.3 to 25.6) | 392,437 ( 1, 0 to  8) |
| 16-34 | 257,806 (24.8, 18.4 to 35.0) | 624,181 ( 1, 0 to  10) |
| 35-64 | 407,051 (39.5, 32.7 to 43.7 | 1,768,563 ( 2, 1 to  17) |
| >=65 | 164,912 (15.9, 8.2 to 22.2) | 1,840,789 (10, 0 to 28) |
| | | |
| Sex | | |
| Male | 508,545 (49.3, 47.4 to 52.3) | 1,831,839 (1, 0 to 17) |
| Female | 523,527 (50.7, 47.7 to 52.6) | 2,794,131 (3, 0 to 19) |
| | | |
| Morbidity | | |
| 1 (healthiest) | 338,890 (31.1, 23.9 to 46.0) | 24,648 (0) |
| 2 | 140,972 (13.7, 8.7 to 20.5) | 483,762 (2, 0 to 13) |
| 3 | 251,278 (25.0, 20.2 to 28.1) | 1,177,099 (3, 0 to 15) |
| 4 | 274,814 (27.1, 13.6 to 35.0) | 2,602,883 (7, 1 to 25) |
| 5 and 6(sickest) | 26,118 (2.5, 1.1 to 4.5) | 337,578   (9, 1 to 36) |
| | | |
| Overall | 1,032,072 | 4,625,970 (2, 0 to 18)* |

*Percentage of patients in each age, sex, and morbidity groups were calculated for each practice

**Table 14  Association between age, sex and morbidity and number of prescriptions issued (results from two level Poisson regression models using patient level data)**

| Rate Ratios (95% CI) | Model 2 (age and sex) | Model 3 (age, sex and morbidity |
|---|---|---|
|  |  |  |
| Age (yrs) |  |  |
| 0-15 | 1 | 1 |
| 16-34 | 1.26 (1.25 to 1.26) | 1.13 (1.12 to 1.13) |
| 35-64 | 2.26 (2.25 to 2.27) | 1.85 (1.84 to 1.86) |
| 65+ | 5.65 (5.63 to 5.67) | 3.38 (3.37 to 3.39) |
|  |  |  |
|  |  |  |
| Sex |  |  |
| Male | 1 | 1 |
| Female | 1.38 (1.37 to 1.38) | 1.10 (1.10 to 1.11) |
|  |  |  |
|  |  |  |
| Morbidity |  |  |
| 1 (Healthiest) |  | 1 |
| 2 |  | 43.42 (42.83 to 44.02) |
| 3 |  | 58.21 (57.53 to 58.89) |
| 4 |  | 97.03 (95.89 to 98.18) |
| 5 and 6 |  | 134.56 (132.73 to 136.42) |

**Table 15  Percentage of variation in prescribing explained using data summarised at practice level**

| Regression models* | Variation (%) explained at practice level |
|---|---|
| Model 1: no predictors | 0 |
| Model 2: age and sex | 4 |
| Model 3: age, sex and morbidity | 57 |

*Mean number of prescriptions issued by each practice was used as response.  Predictors were summarized to express mean (for age) and percentage (for sex and morbidity) for each practice

**Table 16  Percentage of variation in prescribing explained using logistic regression model based on patient level data**

| | Model 1* | Model 2† | Model 3†† |
|---|---|---|---|
| Variation | % | % | % |
| Percentage of total variance explained | 0 | 9.7 | 80.1 |
| Level at which % of total variance was unexplained: | | | |
|    Practice level | 3.9 | 4.1 | 0.1 |
|    Practice level | 96.1 | 86.2 | 19.0 |

Prescribing was dichotomised as prescription issued or not issued for each patient
*   No predictors
†  Age and sex
†† Age, sex and morbidity group

# Chapter 7    Discussion

## 7.1 Introduction

This thesis describes the first large-scale studies in the United Kingdom to control for diagnostic-based morbidity when examining variation in home visits, specialist referrals and prescribing patterns in general practice.  Chapters 2 and 3 set out the literature review and methodology and Chapters 4, 5 and 6 describe examples of its application in examining variation in home visit, specialist referral and prescribing rates in the UK using large general practice datasets from the Morbidity Statistics in General Practice Survey and the General Practice Research Database.  Section 7.2 below summarises each of the chapters separately.  The scope and limitations of the thesis are reviewed in 7.3, and finally, section 7.4 sets out recommendations for health services and future research.

## 7.2 Summary of thesis

In Chapter 2 the literature on variations in general practice resource use in the UK was reviewed and revealed a body of evidence documenting large variations in measures such as consultations, referrals and prescribing practice patterns in general practice both in the UK and beyond.  A brief overview of the history of case-mix and the reasons why case-mix systems based on patient diagnoses in particular have been developed were presented.  The main systems for measuring case-mix were highlighted and key differences between the systems for measuring case-mix were outlined.  The motivation for using the Johns Hopkins Case-Mix System was, briefly, that its components provide measures for each patient's overall morbidity as morbidity has been shown to be a better predictor of health services resource use than other measures, for example, measures based on specific diseases.  Another important feature of the ACG system is that it was originally developed specifically for use in primary care settings.  Aims and objectives for this work were set out, namely, to apply the Johns Hopkins ACG Case-Mix system to general practice populations in the UK and investigate whether these diagnostic-

based measures of morbidity can explain variation between GP outcomes and, in doing so, to explore methods for appropriately dealing with the challenging methodological issues and produce results which can be communicated easily to clinicians and policy makers.

Chapter 3 described the methodological aspects of this thesis, beginning with a description of the ACG grouping mechanism and the development and validation of the morbidity groups. The techniques that others have used for examining variation in general practice outcomes and for identifying factors that might explain the variation were reviewed. Given the clustered nature of general practice data, the particular advantages of multilevel models over other methods were highlighted. The various methods used after fitting multilevel models to enable us to quantify the effect of the model covariates on variability between general practices were presented. Such methods allowed the variation to be partitioned into that at the practice level and that at the patient level.

In Chapter 4, the first study is described where the effect of age, sex, morbidity and social class on variation in general practice home visits was examined, using data from general practices included in the Morbidity Statistics for General Practice (MSGP4) survey. This work produced several interesting findings and demonstrated that the Johns Hopkins ACG System can be applied to UK general practice. Morbidity and social class were both found to be strong determinants of home visits. The odds of a home visit increased with higher morbidity and also, although with a less stark difference, for the most disadvantaged social classes. Age and sex explained slightly more of the total variability in home visits than morbidity (14% vs 10%), and social class only explained only a small proportion of total variability in home visits (1%). For all models fitted, most of the total variability in home visits remained unexplained. This unexplained variation in home visits was mainly attributed to differences *within* practices rather than between practices. These results support other work suggesting that the amount of variability between practices may be less than implied by previous studies based on aggregate information once morbidity is taken into account (Davis P et al, 2002).

When both morbidity and social class were included in a model simultaneously, the effect of social class changed markedly, with the middle classes (those of intermediate and skilled occupations (manual and non-manual) having lower odds of home visits after adjustment for morbidity than the highest (professional) and lowest social classes (partly skilled and unskilled). This contrasts with findings of Julian le Grande that state that the middle classes receive the best level of care in the NHS (Le Grande J (2006)).

Chapters 5 and 6 describe the investigation of the effect of morbidity on variation in specialist referral and prescribing outcomes respectively, using data extracted from general practices participating in the General Practice Research Database. The social class measure previously used in the study of home visits was not recorded in the GPRD dataset. Findings from Chapter 5 show that morbidity explains almost a third (30%) of the total variation in the referral outcome compared to only 5% explained by age and sex. Chapter 6 shows that including morbidity with age and sex explains substantially more (80%) of the total variation in the prescribing outcome compared to 10% for age and sex. In terms of the main focus of interest, which was variation *between* practices, morbidity explains a similar proportion for the referral and home visit outcomes (3.6% and 3.5% respectively remains unexplained compared to 4.5% and 4.7% by age and sex respectively) but morbidity explains substantially more for the prescribing outcome (0.1% remained unexplained compared to 4.1% for age and sex). Morbidity significantly improved the ability of the model to discriminate between patients who had received a referral and those who had not (ROC 0.77) and even more for discriminatory ability of the prescribing model between patients who had received a prescription and those who had not (ROC 0.97).

The success of the morbidity measure in explaining variability in the prescribing outcome compared, say, home visits, may be due to the fact that prescriptions tend to be given for particular conditions and are directly related to patients' diagnoses, the basis for the ACG system, whereas historically home visits tend to be made to the young and elderly.

Although this research has shown that morbidity explains more of the variation *between* general practices then age and sex for all three outcomes investigated, Chapters 4 and 5

illustrate that most of the total variation in two of the outcomes, home visits and referrals, remains unexplained. Prior to fitting models with covariates to the general practice outcomes, the variation was split into practice and patient level variation and most of it was found to be between patients within practices rather than between practices. Even after fitting the various models, the unexplained variation occurs largely within practices so that, for example, the GP decision of whether to visit or refer a patient varies, even for patients with similar age, sex and morbidity. Most of the variation is from those patients with higher morbidity/comorbidity, and this may reflect uncertainty in dealing with more complex cases.

Exploring use of the various statistical methodologies described in Chapter 3 for the applications to general practice home visits, referrals and prescribing patterns described in Chapters 4 to 6, it was found that a combination of methods to quantify the effect of model covariates on variability between practices provided more insight than any one single measure (O Sullivan C et al, (2005)). This research has contributed to the application of statistical methods to explain variation in general practice outcomes using large and complex primary care data sets. It should also contribute to raising awareness among primary care researchers of the value of techniques such as multilevel modelling, and the need for further development of derived measures of variation based on multilevel models (e.g. confidence intervals for R-squared measures) in order to answer questions raised through this work.

For both the referral and prescribing outcomes, morbidity significantly improved the ability of the models to discriminate between patients with and without the outcome compared to age and sex. Including morbidity in the models considerably improves explanatory power of variation in the outcomes explained. The thesis gives examples of how it could be used in practice, for example, for fairer comparisons of prescribing and referrals in general practice; in determining which group is more likely to receive prescriptions; in assigning each patient to a morbidity group and comparing to see which practice has a higher burden of illness etc.

In recent years there have been many new tools developed to aid comparison of outcomes between general practices and Primary Care Trusts. The Association of

Public Health Observatories and London Health Observatory practice profile tools www.lho.nhs.uk, the NHS Comparators website www.nhscomparators.nhs.uk and the NHS Atlas of Variation www.rightcare.nhs.uk are among a suite of tools that have been developed with the aim of assisting GPs and primary care commissioners in providing and commissioning healthcare services for their local population, enabling them to compare and investigate aspects of local activity and outcomes.  The findings of this thesis demonstrate the important role of patient morbidity in explaining referral and prescribing patterns between general practices and in determining which group of patients is most likely to be referred or to receive a prescription.  A serious problem associated with lack of appropriate adjustment for clinical case-mix is misidentification of outliers.  Many practices identified as outliers when adjusted for age and sex, are no longer outliers when case-mix adjustment is applied. Attentions may be misdirected to problems that are less serious than perceived, while ignoring the real problem areas. This leads to a waste of time, money and resources.  Practices with a relatively higher or lower burden of morbidity may be wrongly perceived as over or under using services, or not achieving quality targets, and, as a result, may be inappropriately scrutinised or penalised.  Thus, for any comparison of general practice outcomes, careful consideration should be made of the case-mix measures that may affect the outcome.

## 7.3 Scope and limitations

Previous studies have shown that UK general practice home visit, referral and prescribing rates vary considerably. Statistical models have not included direct measures of the morbidity of patients and have generally explained only a small proportion of this variation.  Chapters 4 to 6 describe the first large-scale studies in the UK to use patients' morbidity to explain variation in general practice home visits, referrals and prescribing patterns.

The application of the ACG System in the UK is of particular interest because the vast majority of the population is registered with a practice, while studies elsewhere have tended to focus on, say, members of a single health plan, or the elderly (DH Departmental Report 2008).  These were primary care based studies that used detailed age, sex and diagnostic data over a one year period for all patients from a large number

of general practices contributing to either the MSGP4 or the GPRD, both which have been well validated and shown to be of high quality. The studies have the advantage of using detailed demographic and diagnostic data for a large number of general practices and take into account the inherent clustered nature (Section 3.6) of general practice data. Hence they are not ecological in design, in contrast to many previous studies of variation in general practice patterns.

In this thesis, most of the general practice outcomes investigated are treated as binary (yes/no) outcomes. Interpretation of measures of variation in such outcomes is not straightforward when applying multilevel modelling techniques. This research showed that a single measure was not sufficient to explain such variation and that a combination of these methods provided better insight.

In applying the ACG system, one must assume that the data recorded is complete and of high quality. In reality, while it is likely that practices record age and sex to a fairly high standard, it is possible that there is variability in quality and completeness of recording of diagnoses. Certainly this was the case for some GPRD practices, which is why only data was used from practices that met the "up to standard criteria", a quality marker set on the basis of internal consistency of the practice, completeness of longitudinal recording, and compliance with the recording guidelines of the General Practice Research Database (www.gprd.com).

Only data from consultations in general practice was used to generate the measures of case-mix. Using data from secondary care in addition to primary care data may have resulted in more accurate measures of case-mix.

The system uses diagnostic codes recorded by general practitioners over an extended period of time (usually one year) and is therefore a proxy measure of patient morbidity. There may be an underlying tendency for certain practices to see their patients as more sick and therefore code them that way. Hence, differences in the way that general practitioners record or code similar conditions on their practice computers could introduce bias into the estimates of the practices' morbidity scores. This is a potential source of confounding that cannot be evaluated with this data.

Another limitation is that the method of assigning case-mix was developed in the USA and may need some modification to maximise its utility in the United Kingdom.

For the home visits and specialist referrals from general practice, even after adjustment for age, sex and morbidity (and social class for the home visits outcome), much of the variability remains to be explained. Other than morbidity and social class within a practice, other factors that could influence such outcomes include doctors' knowledge, professional experience, role perception and time pressures; the number of doctors in the general practice; deprivation; local provision of specialist services and patients' expectations of receiving a prescription or referral and their demands. This information is not available from routine general practice databases. In Section 7.4 some possible further work in this context is suggested.

The focus of Chapter 6 was on prescriptions issued. Each prescription issued might, however contain several items and contain drugs for very different therapeutic areas. In Section 7.4 a recommendation for further work is made and suggests using an alternative outcome measure such as number of items prescribed or, better still, number of items dispensed.

The limitations described in this section are addressed in the form of recommendations in Section 7.4.

Since the papers based on chapters 4 to 6 of this thesis have been published there has been an explosion of interest in the UK for measuring diagnostic-based morbidity in general practice and for having patient focussed analyses and interventions. Valderas's communication to the BMJ in 2009 spoke about 'multimorbidity' as a research priority for the UK. One of the five core research programmes National Health Research's School for Primary Care Research focuses specifically on comorbidity research (Valderas et al (2009)). The work of this thesis provides a valuable contribution to the evidence base of understanding how morbidity measures that take into account multiple conditions of patients over time can contribute to explaining variation in general practice outcomes.

A recent King's Fund study by Appleby J et al (2010) highlighted the great 'potential to improve performance by focusing on decision making and reducing variations in clinical practice across the NHS'. … With the 'scale of the quality and productivity challenge facing the NHS' they identify that 'tackling variations in clinical practice is one of the most important areas to focus on'. The work of this thesis has been at the forefront in developing an understanding of the important role of patient morbidity in variation in general practice and the findings will have important implications in helping to address current challenges for the NHS.

# 7.4 Recommendations

Several recommendations are highlighted following on from this thesis. These include both recommendations immediately relevant to health services (e.g. GPs, PCTs etc.) and also suggestions for future research and are grouped as such below.

### 7.4.1   Recommendations for Health Services

The findings of this thesis demonstrate the importance of morbidity in explaining referral and prescribing patterns between general practices and in determining which group of patients is most likely to be referred or to receive a prescription. A serious problem associated with lack of appropriate adjustment for morbidity is misidentification of outliers. Many practices identified as outliers when adjusted for age and sex may no longer be outliers when case-mix adjustment is applied. Attentions may be misdirected to problems that are less serious than perceived, while ignoring the real problem areas. This leads to a waste of time, money and resources. Practices with a relatively higher or lower burden of morbidity may be wrongly perceived as over or under using services, or not achieving quality targets, and, as a result, may be inappropriately scrutinised or penalised. This work has shown that, for comparisons of general practice outcomes such as referrals to secondary care and prescribing, morbidity should be taken into account.

**Recommendation 1:  Morbidity of patients should be taken into account in comparisons of referral and prescribing outcomes between general practices,**

The general practice outcome measures examined in this thesis were based on activity and it is possible that outcomes in the form of cost data could give further insight into the variation in these outcomes across practices.  There may be potential for using such a system to support distribution of resources to general practices, for example, in exploring potentially more equitable allocation of prescribing budgets to general practices using prescribing cost data.  Other uses (that could be based on either activity and/or cost outcomes) could assist in better understanding of the health status of the patient population; identification of those who are likely to use services in the future; examining how efficiently health services are being used; and producing measures of clinical performance and quality of care adjusted for morbidity.

**Recommendation 2 : Explore potential to use ACGs in the UK for various applications, e.g. to gain a better understanding of the health status of the patient population; support distribution of resources to general practices; case finding those who are likely to use services in the future; examining how efficiently health services are being used; and producing morbidity adjusted measures of clinical performance and quality of care**

While morbidity has been shown to be an important factor in comparisons of general practice outcomes, case-mix adjustment systems can be time-consuming and costly to implement and administer (Dunn et al (1996)).  In comparisons of age-sex, ACG and DCG methods, Dunn et al found that, although the diagnostic based methods were a big improvement on age-sex for predictive accuracy, the age-sex adjusted method fared best for criteria such as ease of administration, cost, resistance to gaming and ease of audit.  In adjusting doctor's payments for case-mix, Kuttner (1998) concludes that it seems a complex way of achieving some straightforward policy goals.  Before widespread adoption in the UK, a pilot study should be set up to evaluate the cost-effectiveness of implementing and administrating such a system.

**Recommendation 3: Evaluate cost effectiveness of implementing and administering in the UK with a pilot study**

A limitation of most of the current studies in the UK is that they depend on the availability of good quality data.  In general, the amount and quality of diagnostic data collected in general practice varies widely although prescribing data is of better quality

(Thiru K et al (2003)).  However, the need for improvement of general practice data is clear and real problems in data quality are documented in clinical records in primary care (House of Lords (2001)).  Data quality is of utmost importance for the application of diagnostic based case-mix systems.  Poorly recorded, entered or coded data and missing data may lead to patients being wrongly classified and conclusions from studies relying on such poor quality data may be erroneous. It is therefore important to ensure that the data have been validated to an acceptable degree of accuracy.  It is also necessary to ensure careful monitoring of the implementation of case-mix systems.

**Recommendation 4: Before applying case-mix measurement systems, ensure data is validated to acceptable degree of accuracy.**

**Recommendation 5: Ensure careful monitoring of implementation of case-mix systems.**

## 7.4.2   Recommendations for Research

The ACG system was derived for use in the USA and is likely to need some further adaptation to maximise its utility in the United Kingdom.  This would require input from clinicians in order to thoroughly examine the classification of diagnoses into morbidity groups as described in Chapter 3 and to adapt the system accordingly where appropriate.

**Recommendation 6: Examine the classification of UK general practice diagnoses into morbidity groups with input from clinicians and adapt ACG system where appropriate.**

In this thesis, only general practice data was used to generate the morbidity measures for each patient. While hospital patient data is available to commissioners for applying risk adjustment models such as the PARR++ (Patients At Risk of Readmission) model, this is not (at time of publishing) generally the case for general practice data.  With some commissioners now being given access to general practice electronic medical records (for example, Croydon Primary Care Trust's Virtual Ward Project) computerised clinical general practice data for measurement of morbidity will be increasingly available to enable increased understanding of the morbidity of general practice populations for doctors and other healthcare professionals.  Linking patient data

from general practice to corresponding secondary care data (e.g. hospitals) may enable calculation of more comprehensive measures of morbidity for each patient.

**Recommendation 7: Pursue possibility of using data from secondary care (e.g. hospitals) in addition to general practice data to obtain more comprehensive measures of morbidity in the UK.**


For databases such as the General Practice Research Database (Section 5.2.1) where relatively high quality general practice data is available, a limited number of variables are collected.  In this thesis, the choice of possible adjustment factors when investigating variation between practices has been limited by using the GPRD.  Sometimes it is possible to combine various datasets, e.g. census data and general practice data at the enumeration district level, to obtain more information (Scrivener G and Lloyd DCEF (1995)).  However, linking of such data in the UK is not straightforward, mainly for reasons of anonymity.  It is often not possible to relate data from different sources for reasons of confidentiality.  It would be worthwhile to explore further the possibility of linking other relevant data, for example, patient/general practice/secondary care variables, with the GPRD data or extending the number of variables collected by the GPRD in order to allow more in-depth investigations.   For example data on doctors' professional experience, role perception and time pressures; the number of doctors in the general practice; deprivation; local provision of specialist services and patients' expectations or receiving a prescription or referral and their demands.  Once the data is linked, the multilevel models considered could be extended to explore the inclusion of measures at the general practice and other levels that may influence the general practice outcomes.  These could be explored, together with patient age, sex and morbidity to see how these contribute relative to one another in explaining general practice variation. Clinical judgement and an awareness of the important influences on resource use in general practice must be exercised when selecting factors for inclusion in the models.  Linkage of cost data to referrals and prescribing data (E.g. linking cost of referrals together with pharmacy cost) would allow a more comprehensive understanding of use of resources.

**Recommendation 8: Explore possibility of linking other relevant data (e.g. patient/general practice/secondary care data) with GPRD data or extending number of variables collected**

**in order to allow more in-depth investigation. Linkage of cost data would allow a more comprehensive understanding of variation in use of resources.**

The prescribing outcome used in this thesis was based on the total number of prescriptions for a particular patient. Each prescription issued might, however contain several items and contain drugs for very different therapeutic areas. Hence, further work is required to investigate the association between morbidity and total prescribing volume (measured by number of items prescribed) and costs and how well morbidity explains prescribing variation for specific therapeutic areas. An improvement on the prescribing outcome used in this thesis would be number of items prescribed or, even better, number of items dispensed.

**Recommendation 9: Further work is required to investigate the association between morbidity and total prescribing volume (measured by number of items prescribed) and costs and how well morbidity explains variation in prescribing in specific therapeutic areas.**

**Recommendation 10: Further work is also required to investigate the impact of morbidity on variation in referrals by specialty.**

It is important for commissioners and policy makers to understand the association between morbidity and the cost implications of variation in general practice. Adjustment for morbidity may be useful, for example, in setting of general practice prescribing budgets.

**Recommendation 11: Further work is required to understand the association between morbidity and the cost implications of variation in general practice outcomes such as prescribing. Adjusting for morbidity may prove useful, for example, in setting of general practice prescribing budgets.**

**Recommendation 12: An improvement on the prescribing outcome used in this thesis would be number of items prescribed or, even better, number of items dispensed.**

**Recommendation 13: Extend multilevel models to include measures at practice and other levels that may influence the general practice outcomes**

More research is needed to ensure that the findings remain consistent for the outcomes examined across different general practice populations in the UK.

**Recommendation 14: Further research is necessary to ensure findings consistent across different general practice populations.**


Chapter 4 examined variation in home visits and compared morbidity with social class using the MSGP4 data. Such a social class measure is not available in the GPRD. Further work envisaged includes comparing the morbidity measure used with another simpler morbidity measure that may be easier and/or less costly to implement in practice.

**Recommendation 15: Further research is necessary to compare the explanatory power of ACG morbidity measures with other morbidity measures, for example the Charlson Comorbidity Index, in UK populations.**


Measuring variation for discrete outcomes is not straightforward and it was found that a combination of methods to quantify the effect of model covariates on variability between practices provided more insight than any one single measure. The Snijder's and Boskers R-squared measures for multilevel discrete outcomes were particularly useful for the purpose of this research, as they allow the unexplained variation to be quantified separately into that due to practices and that due to differences between patients within practices. Further work is needed to investigate the assumptions required for this measure, and how it can be extended to cover count outcomes and to obtain confidence intervals.

**Recommendation 16: This work suggests that a combination of methods to quantify the effect of multilevel model covariates on variability between practices (for discrete outcomes) provided more insight than any one single measure. The Snijder's and Boskers R-squared measures for multilevel discrete outcomes were found to be particularly useful, as they allow the unexplained variation to be quantified separately into that due to practices and that due to differences between patients within practices. Further work is needed to investigate the assumptions required for this measure, and how it can be extended to cover count outcomes and to obtain confidence intervals.**

# 7.5 Relevant publications and oral presentations

## Awards

This thesis was funded by a Department of Health UK National Primary Care Researcher Development Award.

## Publications from this thesis

**O'Sullivan C**, Omar RZ, Ambler G, Majeed A. Case mix and variation in specialist referrals in general practice. BJGP 2005;55(516):529-533.

**O Sullivan C**, Omar RZ, Forrest CB, Majeed A. Adjusting for case mix and social class in examining variation in home visits between practices. Fam Prac 2004,21(4): 355-363.

Omar, RZ & **O'Sullivan C** (joint first authors), Petersen I, Islam A, Majeed A.  A model based on age, sex, and morbidity to explain variation in UK general practice prescribing: a cohort study. *BMJ* 2008;337:a238

*Unpublished paper*

**O Sullivan C**, Omar RZ, Majeed A.  Performance monitoring in primary care: a practical comparison of methods of estimating variation in performance for binary outcomes

## Publications related to variation in general practice

Majeed A, Bardsley M, Morgan D, **O'Sullivan C**, Bindman AB. Cross sectional study of primary care groups in London: association of measures of socioeconomic and health status with hospital admission rates. BMJ 2000;321:1057-1060.

Gray J, Millett C, **O'Sullivan C**, Omar R Z and Majeed A (2006) Association of age, sex and deprivation with quality indicators for diabetes: population‐based cross sectional survey in primary care, *Journal of the Royal Society of Medicine* 99 (11) : 576‐581

## Relevant oral presentations

Royal Statistical Society (RSS) and Statisticians in the Pharmaceutical Industry (PSI) joint conference, Cardiff, July 2005: 'Estimating variability in general practice outcomes and identifying factors that explain variability'

Invited speaker at the RSS half day primary care meeting: 'Methods of estimating variation in performance between general practices', London, 2003

The Society for Academic Primary Care Conference Annual Science Meeting, Manchester, 2003: 'Effect of adjustment for clinical case mix on variation in referrals between practices'

Berlin ACG Users Group: 'Use of ACGs in the UK', 2003

Johns Hopkins ACG Case-Mix users conference: 'Adjusting for Case Mix and Social Class in Comparisons of Home Visiting Rates across Practices'. Baltimore, USA 2002

## Relevant poster presentations

Society for Academic Primary Care conference: 'Psychosocial consultations & psychiatric referrals: the role of case mix', Glasgow, Scotland, 2004

Royal Statistical Society seminar: 'Performance monitoring in primary care: comparison of three methods of estimating variation in performance for binary outcomes', London, 2003

# Appendices

Number of patients who had at least one consultation in one year by ADG and number (%) that are home visits

**Table 17  Home visits by ADG**

| ADG | | No home visit (%) | Home visit (%) | Total |
|---|---|---|---|---|
| 1. | Time Limited: Minor | 57 453 (80.00) | 14 359 (20.00) | 71 812 |
| 2. | Time Limited: Minor—Primary Infections | 110 125 (75.69) | 35 376 (24.31) | 145 501 |
| 3. | Time Limited: Major | 2272 (48.48) | 2414 (51.52) | 46 86 |
| 4. | Time Limited: Major—Primary Infections | 8121 (65.87) | 4207 (34.13) | 12 328 |
| 5. | Allergies | 13 855 (85.50) | 2349 (14.50) | 16 204 |
| 6. | Asthma | 14 025 (73.97) | 4936 (26.03) | 18 961 |
| 7. | Likely to Recur: Discrete | 39 536 (74.49) | 13 539 (25.51) | 53 075 |
| 8. | Likely to Recur: Discrete—Infections | 44 412 (76.90) | 13 344 (23.10) | 57 756 |
| 9. | Likely to Recur: Progressive | 1279 (30.95) | 2854 (69.05) | 4133 |
| 10. | Chronic Medical: Stable | 50 011 (76.22) | 15 606 (23.78) | 65 617 |
| 11. | Chronic Medical: Unstable | 13 494 (57.29) | 10 060 (42.71) | 23 554 |
| 12. | Chronic Specialty: Stable—Orthopaedic | 8726 (79.97) | 2186 (20.03) | 10 912 |
| 13. | Chronic Specialty: Stable—Ear, Nose, Throat | 1774 (80.49) | 430 (19.51) | 2204 |

| 14. | Chronic Specialty: Stable—Eye | 3059 (72.49) | 1161 (27.51) | 4220 |
|---|---|---|---|---|
| 16. | Chronic Specialty: Unstable—Orthopaedic | 2169 (78.90) | 580 (21.10) | 2749 |
| 17. | Chronic Specialty: Unstable—Ear, Nose, Throat | 340 (70.39) | 143 (29.61) | 483 |
| 18. | Chronic Specialty: Unstable—Eye | 2839 (69.70) | 1234 (30.30) | 4073 |
| 20. | Dermatological | 27 540 (87.46) | 3950 (12.54) | 31 490 |
| 21. | Injuries/Adverse Effects: Minor | 30 821 (80.46) | 7483 (19.54) | 38 304 |
| 22. | Injuries/Adverse Effects: Major | 42 676 (78.49) | 11 694 (21.51) | 54 370 |
| 23. | Psychosocial: Time limited, Minor | 6239 (70.31) | 2634 (29.69) | 8873 |
| 24. | Psychosocial: Recurrent or Persistent, Stable | 17 004 (73.89) | 6009 (26.11) | 23 013 |
| 25. | Psychosocial: Recurrent or Persistent, Unstable | 3535 (62.14) | 2154 (37.86) | 5689 |
| 26. | Signs/Symptoms: Minor | 31 102 (71.99) | 12 100 (28.01) | 43 202 |
| 27. | Signs/Symptoms: Uncertain | 34 812 (75.06) | 11 567 (24.94) | 46 379 |
| 28. | Signs/Symptoms: Major | 16 536 (69.38) | 7297 (30.62) | 23 833 |
| 29. | Discretionary | 18 822 (80.23) | 4637 (19.77) | 23 459 |
| 30. | See and Reassure | 3715 (72.81) | 1387 (27.19) | 5102 |
| 31. | Prevention/Administrative | 115 246 (81.88) | 24 912 (18.12) | 137 458 |

| 32. | Malignancy | 1734 (50.04) | 1731 (49.96) | 3465 |
| 33. | Pregnancy | 64 (70.33) | 27 (29.67) | 91 |
| 34. | Dental | 1848 (75.86) | 588 (24.14) | 2436 |

## 2 level logistic regression model

Two level logistic regression models were used to adjust for various sets of predictors of our home visit, referral and prescribing outcomes. Writing $y_{ij} = 0/1$ as the binary outcome representing whether the $i^{th}$ patient in the $j^{th}$ practice experienced the outcome or not. Assume $y_{ij} \sim$ Binomial $(1, \pi_{ij})$ where $\pi_{ij}$ is the probability of the $i^{th}$ patient in the $j^{th}$ practice experiencing the outcome. The model is expressed as:

$\alpha_j$ - log odds of outcome for patient with covariate values zero

$\beta_{1k}$ - log(odds ratio) for the $k^{th}$ covariate (of n covariates)

$x_{ijk}$ - $k^{th}$ covariate for the $i^{th}$ patient in the $j^{th}$ practice

# Examples of multilevel logistic models applied in this thesis

**Model 1  A model with no covariates**

$$\text{logit}(\pi_{ij}) = \alpha_j$$

**Model 2  A model including age and sex only**

$$\text{logit}(\pi_{ij}) = \alpha_j + \beta_1 male_{ij} + \beta_2 age(2)_{ij} + ... + \beta_8 age(8)_{ij}$$

**Model 3  A model including morbidity**

$$\text{logit}(\pi_{ij}) = \alpha_j + \beta_1 morb(2)_{ij} + ... + \beta_7 morb(8)_{ij}$$

$male_{ij}$ - sex indicator of the $i^{th}$ patient in the $j^{th}$ practice

$\beta$  - log odds ratio for the corresponding covariate

$age(2)_{ij}$ - $2^{nd}$ age group indicator for the $i^{th}$ patient in the $j^{th}$ practice

$u_j$ - the random effect corresponding to the $j^{th}$ practice

This model allows the log odds of the outcome to vary between practices. The $u_j$ are assumed to be distributed normally with mean 0 and variance $\sigma_u^2$.

## Confidence intervals for ICCs

It is important to take into account the uncertainty of model derived estimates when making comparisons (Goldstein and Spiegelhalter (1996)). Predictive models can only approximate real situations using data with which there will almost always be noise random fluctuation. Confidence intervals can be used to assess the precision of predicted outcomes and are dependent on the size of the dataset and the predictive ability of the covariates. Confidence intervals for the ICC are not straightforward to obtain for discrete response models. It is, however, possible to use bootstrapping or MCMC methods for obtaining 95% confidence intervals estimates for the ICCs. The bootstrap method is described below.

## Bootstrapping to obtain ICC confidence intervals

A single sample of data gives one estimate for each parameter in a model and allows us to calculate the sample estimate of the ICC (Rasbash et al, 2000). Repeated sampling with replacement of the dataset is called bootstrapping, and allows multiple estimates of the outcome to be calculated (Meijer et al (1995); Goldstein (1996)). These estimates can then be used to assess bias and to produce estimates of the model uncertainty.

There are both parametric and non-parametric methods for generating bootstrapped data (Carpenter et al 1999, 2000). This research uses parametric methods.

## Parametric bootstrapping

Parametric bootstrapping uses assumptions about the distribution of the data to construct bootstrap datasets for models. Estimates of model parameters and practice level variance are obtained on fitting a multilevel logistic regression model to general practice data. A series of practice level residuals can then be sampled from a Normal distribution with zero mean, and variance equal to the between practice variance estimate. The parameter of interest, in our case the ICC, can be calculated for each of the sample bootstrapped datasets as it was calculated for the original dataset. This provides a large number of estimated values for the parameter of interest which can be used with, for example, the percentile confidence interval, to estimate a corresponding confidence interval.

## Confidence interval for bootstrapped data

The percentile confidence interval is a simple method that uses bootstrapped datasets to obtain a confidence interval for the parameter of interest (Carpenter et al (2000); DiCiccio et al (1996(a)(b))).  The estimates are ordered from the smallest value to the largest.  Values at the 2.5[th] percentile and the 97.5[th] percentile can be quoted respectively as the lower and upper range of the 95th percentile interval.

The confidence intervals for the ICC estimated with the Turner (2001) method were calculated in two steps. Firstly the estimated between practice variances are ordered from smallest to largest and then the ICC is estimated for both the 2.5[th] and 97.5[th] percentiles of this as the ICC always increases with increasing between practice variance for the Turner method.  These two ICC values represent the lower and upper limits respectively for an approximate 95% confidence interval for the ICC.

## ICC – Goldstein's methods

**Normal response model**

With this method, the binary variable is considered as a continuous response and a 2-level normal response model with a random intercept is fitted. An estimate of ICC can be obtained provided that the probability of observing the outcome is not extreme (Goldstein; Collet, D. (1991)). The ICC is estimated based on the definition given in Equation 2, using the between and within practice variance estimated from the 2-level random intercept model. The residuals at the practice and patient levels are assumed to be distributed as $N(0, \sigma_u^2)$ and $N(0, \sigma_e^2)$ respectively.

**Method of model linearisation**

This method linearises the logistic regression model with the first order Taylor Series expansion (Goldstein, H. et al (2002)). Given a set of predictor values, the between and within practice variance can be expressed as a linear function of the practice-specific response probabilities. Sample estimates are substituted for the parameters, and the ICC is given by:

$$\rho_{ij} = \sigma_u^2 \pi_{ij}^2 ([1 + \exp(\beta_0 + \sum_{k=1}^{n} \beta_{1k} x_{ijk})]^{-2}) \Big/ \{\sigma_u^2 \pi_{ij}^2 ([1 + \exp(\beta_0 + \sum_{k=1}^{n} \beta_{1k} x_{ijk})]^{-2}) + \pi_{ij}(1 - \pi_{ij})\}$$

(4)

with notation defined as in above.

This method does not produce a single value for the ICC. It estimates an ICC for each of the predictors included in the model. For example, it will produce an ICC for each of the 8 morbidity groups, one for each group. To provide a single measure for each model, the mean of all the ICCs estimated for all sets of combinations of possible predictor values was calculated.

**Method of model simulation**

With the simulation method, a 2-level logistic regression model is fitted to the data. A large number of values for the between practice residuals are simulated from a Normal distribution with mean zero and variance equal to the estimated $\sigma_u^2$ from the model (Goldstein, H. et al (2002)). The corresponding response probabilities $\pi_{ij}$ are calculated. The variance of the response probabilities is equivalent to the between practice variance. The level 1 variance is computed for each of these probabilities. The expected value of these within practice variances is an estimate of the within practice variance for the simulated data. The ICC for each set of predictor values is calculated as in definition (1). Similarly to the method of linearisation, the mean ICC, calculated from all sets of combinations of possible predictor values, is presented as the overall estimate of ICC.

**Comparison of methods for estimating ICC**

For the home visits study, Goldstein's model linearisation and simulation methods resulted in similar, although slightly higher estimates of ICC of 2.7 to 2.8% for age and sex and 2% for morbidity. Treating the binary 0/1 as a normally distributed variable produced ICCs very similar to that obtained from the Turner method for model 2 but lower than that that from model 1. However, the overall conclusion regarding the variability in performance remained the same.

## Calculation of Coefficient of Variation for home visits study (using Woolf adjustment)

Initially adjustment is made for age and sex only using Woolfs method (Breslow et al (1980) and a weighted average is calculated in the following way:

$$\hat{\alpha} = \sum w_i \hat{\alpha}_i / w_i \text{ with variance } 1/\sum w_i$$

where $\hat{\alpha}_i$ is the estimated log odds for the ith practice. The weights $w_I$ are the inverse of the variance $v_i$ of $\hat{\alpha}_i$ as $v_i = 1/a_i + 1/b_i$ where $a_i$ and $b_i$ are the number of patients with and without home visits in the ith age-sex group. There are 8 age groups thus providing a total of 16 age-sex grouping for each practice. This method is a fixed effect method widely used in meta-analysis.

The log odds adjusted for age and sex is then exponentiated and converted to the probability of a making a home visit for each practice. One can then calculate the average probability of home visit and the corresponding standard deviation for 60 practices. This allows calculation of coefficient of variation after taking account of differences in age and sex between the practices. Similarly, one can use adjust for the eight morbidity classes and calculate a CV. If the CV adjusted for age and sex is lower than that obtained by adjustment via morbidity class, it implies that adjusting for clinical case-mix explains more of the variability.

# References

Aitchison J, Shen SM. Logistic-normal distributions: some properties and uses. *Biometrika* 1980;67(2):261-272

Armstrong D, Fry J, Armstrong P. Doctors' perceptions of pressure from patients for referral. *British Medical Journal* 1991;302:1186-8

Ash AS, Ellis RP, Yu W, et al.: *Risk Adjustment for the Non-Elderly.* Boston University, Department of Medicine. Report to the Health Care Financing Administration. Boston. June 1998.

Ashworth M, Lloyd D, Smith RS, Wagner A, Rowlands G. Social deprivation and statin prescribing: a cross-sectional analysis using data from the new UK general practitioner 'Quality and Outcomes Framework.' *J Public Health* 2007;29:40-7

Averill RF, Goldfield NI, Eisenhandler J, Muldoon JH, Hughes JS, Neff JM, Gay JC, Gregg LW, Gannon DE, Shafir BV, Bagadia FA, and Steinbeck BA et al. Development and Evaluation of Clinical Risk Groups (CRGs). 1999. 3M Health Information Systems. Ref Type: Report

Aylin P, Majeed FA, Cook DG. Home visiting by general practitioners in England and Wales. *British Medical Journal* 1996;313:207-210

Barber J, Thompson S. Multiple regression of cost data: use of generalised linear models. *J Health Serv Res Policy*. 2004 Oct;9(4):197-204.

Bindman AB, Forrest CB, Britt H, Crampton P, Majeed A. Diagnostic scope of and exposure to primary care physicians in Australia, New Zealand, and the United States: cross sectional analysis of results from three national surveys. *BMJ* 2007;334:1261-4

Bland, M. *An introduction to medical statistics*. 2nd edn. Oxford Medical Publications 1995. pp267-268

Blendon RJ, Schoen C, Donelan K, Osborn R, DesRoches CM, Scoles K et al. Physicians' views on quality of care: a five-country comparison. *Health Affairs* 2001;20 (3):233-43

Bolanos-Carmona V, Ocana-Riola R, Prados-Torres A, Gutierrez-Cuadra P. Variations in health services utilization by primary care patients. *Health Serv Manage Res* 2002 May;15(2):116-25

Breslow NE, Xihong L. Bias correction in generalised linear mixed models with a single component of dispersion. *Biometrika* 1995;82(1):81-91

Breslow NE, Clayton DG. Approximate inference in generalized linear mixed models. *Journal of the American Statistical Association* 1993;88(421):9-25

Breslow, N.E., Day, N.E. Statistical Methods in Cancer Research. International Agency for Research on Cancer: Lyon 1980

Breslow, N.E., Day, N.E. (1980) *Statistical Methods in Cancer Research*. International Agency for Research on Cancer: Lyon.

Britten N, Ukoumunne O. The influence of patients' hopes of receiving a prescription on doctors' perceptions and the decision to prescribe: a questionnaire survey. *BMJ* 1997;315:1506-10

Carlisle R, Groom LM, Avery AJ, Boot D, Earwicker S. Relation of out of hours activity by general practice and accident and emergency services with deprivation in Nottingham: longitudinal survey. *BMJ* 1998;316(7130):520-3

Carlsson L: Clinical and economic features of categories of patients in defined populations. PhD thesis. Karolinska institutet, Stockholm; 2005.

Carlsson L, Börjesson U, Edgren L: Patient based 'burden-of-illness' in Swedish PHC. Applying the Johns Hopkins ACG case-mix system in a retrospective study of electronic patient records. *Int J Health Plann Mgmt* 2002 , 17:269-82

Carlsson L, Strender L-E, Fridh G, Nilsson GH: Clinical categories of patients and encounter rates in primary health care – a three year study in defined populations. *BMC Public Health* 2006, 6:35

Carlsson, L. (2004). "Burden of Illness in Defined Populations." Department of Clinical Services, Center of Family Medicine, Karolinska Institute

Carlsson L, Strender L-E, Fridh G, Nilsson G: Types of morbidity and categories of patients in a Swedish county. Applying the Johns Hopkins Adjusted Clinical Groups System to encounter data in Primary Health Care. *Scand J Prim Health Care* 2004 , 22:174-79.

http://taylorandfrancis.metapress.com/app/home/contribution.asp?wasp=42f3515c6f774b308efb5dccf3a34cfe&referrer=parent&backto=issue,11,14;journal,4,31;linkingpublicationresults,1:102104,1

Carlsson L, Boriesson U, Edgren L. Patient based 'burden of illness' in Swedish Primary Health Care. Applying the Johns Hopkins ACG Case-Mix System in a retrospective study of electronic patient records. *International Journal of Health Planning and Management* 2002;17(3):269-82

Carpenter J, Bithell J. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine* 2000;19:1141-1164

Carpenter J, Goldstein H, Rasbash J. A non-parametric bootstrap for multilevel models. *Multilevel Modelling Newsletter* 1999;11(1):2-5

Carr-Hill RA, Jamison JQ, O'Reilly D, Stevenson MR, Reid J, Merriman B. Risk adjustment for hospital use using social security data: cross sectional small area analysis. *British Medical Journal* 2002;324:1-4

Carr-Hill RA, Rice N, Roland M. Socioeconomic determinants of rates of consultation in general practice based on fourth national morbidity survey of general practices. *British Medical Journal* 1996;312:1008-1012

Carr-Hill R., Sheldon T.A., Smith P., Martin S., Peacock S. and Hardman G., Allocating resources to health authorities: development of method for small area analysis of use of inpatient services, *British Medical Journal* 1994;309: 1046-9

Carthy P, Harvey I, Brawn R, Watkins C. A study of factors associated with cost and variation in prescribing among GPs. *Fam Pract* 2000;17:36-41

Cockburn J, Pit S. Prescribing behaviour in clinical practice: expectations and doctors' perceptions of patients' expectations—a questionnaire study. *BMJ* 1997;315:520-3

Cohen J, Cohen P, West SG, Aiken LS. Applied Multiple Regression/Correlation Analysis for the Behavioural Sciences (3rd ed) 2003

Collet, D. Modelling Binary Data. Chapman and Hall: London 1991

Commenges D, Jacqmin H. The intraclass correlation coefficient: distribution-free definition and test. *Biometrics* 1994;50(2):517-526

Crampton, C., Davis, P., Lay-Yee, R., Raymont, A., Forrest, C., Starfield, B. (2005). "Does Community-Governed Nonprofit Primary Care Improve Access to Services? Cross-Sectional Survey of Practice Characteristics." *International Journal of Health Services*, 35(3), 465-478

Crampton, C., Davis, P., Lay-Yee, R., Raymont, A., Forrest, C., Starfield, B. (2004). "Comparison of Private For-Profit with Private Community-Governed Not-For-Profit Primary Care Services in New Zealand." *Journal of Health Services Research and Policy*, 9(2), 17-22.

Cumming RB, Knutson D, Cameron BA, and Derrick B. A comparative analysis of claims based methods of health risk assessment for commercial populations. 24-5-2002. Ref Type: Report

Davis P, Gribben B, Lay-Yee R, Scott A. How much variation in clinical activity is there between general practitioners? A multi-level analysis of decision-making in primary care. *Journal of Health Services Research and Policy* 2002;7(4):202-208

Davis P, Gribben B, Scott A., Lay-Yee R. The "supply hypothesis" and medical practice variation in primary care: testing economic and clinical models of inter-practitioner variation. *Social Science & Medicine.* 2000;50(3):407-18

Davis P, Gribben B. Rational prescribing and interpractitioner variation. A multilevel approach. *Int J Technol Assess Health Care* 1995;11(3):428-42

Department of Health. *Delivering IT in the NHS*. 2002
www.doh.gov.uk/ipu/whatnew/deliveringit/index.htm

Department of Health. Saving lives: our healthier nation. London: Stationery Office; 1999 (Cm 4386)

Department of Health. *Departmental Report 2008.* The health and personal social services programme. London: Stationery Office; 2008 (Cm 7393)

DiCiccio TJ, Efron B. Bootstrap confidence intervals. *Statistical Science* 1996;11(3):189-212

DiCiccio TJ, Efron B. Bootstrap confidence intervals: Rejoinder. *Statistical Science* 1996;11(3):223-228

Diez Roux A V. A glossary for multilevel analysis. *Journal of Epidemiology and Community Health*. 2002; 56:588-594

Donald, A., Donner, A. Adjustments to the Mantel-Haenszel chi-square statistic and odds ratio variance estimator when the data are clustered. *Statistics in Medicine* 1987; 6:491-499

Donner A, Koval JJ. The estimation of intraclass correlation in the analysis of family data. *Biometrics* 1980;36(1):19-25

Donner, A., Klar, N. Design and analysis of cluster randomization trials in health research. Arnold 2000

Duckett, S. J., and Agius, P. A. (2002). "Performance of Diagnosis-based Risk Adjustment Measures In a Population of Sick Australians." *Australian and New Zealand Journal of Public Health,* 26(6), 500-507
http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_ids=12530791&dopt=Abstract

Duncan C, Jones K, Moon G. Context, composition and heterogeneity: using multilevel models in health research. *Soc Sci Med* 1998,46(1):97-117

Dunn L, Rosenblatt A, Taira DA Latimer E, Bertko J, Stoiber T, Braun P, and Busch S. A comparative analysis of methods of health risk assessment. 18-6-1996.

Ellis R, Pope G, Iezzoni L. Diagnosis-based risk adjustment for medical capitation payments. *Health Care Financing Review* 1996;17:101-128. (www.dxcg.com)

EPIC General Practice Research Database 'A guide for researchers' August 2003

Fakhraei, SH, Kaelin JJ, Conviser R. Comorbidity-based payment methodology for Medicaid enrollees with HIV/AIDS. *Health Care Financing Review* 2001;23(2):53-68.

Favato G, Mariani P, Mills RW, Capone A, Pelagatti M, Pieri V, et al. ASSET (age/sex standardized estimates of treatment): a research model to improve the governance of prescribing funds in Italy. PLoS ONE 2007;2(7):e592

Fleming D. The measurement of morbidity in general practice. *J Epi & Comm Health* 1991;45(3):180-83

Forrest C, Majeed A, Weiner JP, Carroll K, Bindman AB. Primary care physicians' specialty referral rates in the United Kingdom versus the United States. *British Medical Journal* 2002; 325:370-371

Forrest CB, Majeed A, Weiner JP, Carroll K, and Bindman AB. Referral of children to specialists in the United States and the United Kingdom. *Arch Pediatr Adolesc Med* 2003:157(3):279-285

Forrest CB, Majeed A, Weiner JP, Carroll K, Bindman AB. Comparison of specialty referral rates in the United Kingdom and the United States: retrospective cohort analysis. *British Medical Journal* 2002;325:370-371

Fowles JB, Weiner JP, Knutson D, Fowler E, Tucker AM, Ireland M. Taking health status into account when setting capitation rates. A comparison of risk-adjustment methods. *JAMA* 1996;276:1316-21

Franks P, Nutting PA, Clancy CM. Health care reform, primary care, and the need for research. *JAMA* 1993;270:1449-1453

Franks P, Williams CG, Zwanziger J, Mooney C, Sorbero M. Why do physicians vary so widely in their referral rates? *Journal of General Internal Medicine* 2000; 15:163-8

Franks P, Zwanziger J, Mooney C, Sorbero M. Variations in primary care physician referral rates. *Health Services Research* 1999;34:323-9

General Practice Research Database. 2007. www.gprd.com

Goldstein H, Browne W, Rasbash J. Multilevel modelling of medical data. *Statistics in Medicine* 2002;1521(21):3291-315

Goldstein H, Spiegelhalter DJ. League tables and their limitations: statistical issues in comparisons of institutional performance. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 1996;159(3):385-443

Goldstein H,.Rasbash J. Improved approximations for multilevel models with binary responses. *Journal of the Royal Statistical Society, Series A (Statistics in Society)* 1996;159:505-13.

Goldstein H. Consistent estimators for multilevel generalised linear models using an iterated bootstrap. *Multilevel Modelling Newsletter* 1996;8(1):3-6

Goldstein H. Multilevel mixed linear model analysis using interative generalized least squares. *Biometrika* 1986;73(1):43-56

Goldstein H. Restricted unbiased iterative generalized least-squares estimation. *Biometrika* 1989;76(3):622-623

Goldstein H. Multilevel statistical models. 3rd ed. London: Edward Arnold, 2003

Goldstein H. Nonlinear multilevel models, with an application to discrete response data. *Biometrika* 1991;78(1):45-51.

Goldstein, H., Browne, W., Rasbash, J. Partitioning variation in multilevel models. *Understanding Statistics* 2002;1(4):223-232

Greenhalgh T, Controversies in primary care. Setting prescribing budgets in general practice. Effective prescribing at practice level should be identified and rewarded, *BMJ*, 316(7133), 750-3, 1998

Grumbach K,.Fry J. Managing primary care in the United States and in the United Kingdom. *New England Journal of Medicine* 1993;328(13):940-5

Halling, A., Fridh, G., Ovhed, I. (2006). Validating the Johns Hopkins ACG Case-Mix System of the Elderly in Swedish Primary Health Care. *BMC Public Health,6*(171)

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

Herrett E, Thomas SL, Marieke Schoonen W, Smeeth L, Hall AJ. 2009  Validation and validity of diagnoses in the General Practice Research Database: a systematic review. *British Journal of Clinical Pharmacology* 2010;69(1):4-14

Hippisley-Cox J, Hardy C, Pringle M, Fielding K, Carlisle R, Chilvers C.  The effect of deprivation on variations in general practitioners' referral rates: a cross sectional study of computerised data on new medical and surgical outpatient referrals in Nottinghamshire. *BMJ* 1997;314:1458-1461

Hollowell J. The General Practice Research Database: quality of morbidity data. *Population Trends* 1997;87:36-40

Hornbrook MC, Goodman MJ, Fishman PA, Bachman DJ, Rosetti MO, and Nelson AF. Global risk assessment models: two-part, three segment approach; risk adjusters: where are we?  117-140. 1996. Schaumberg IL. Society of Actuaries

House of Lords.  *Select committee on science and technology: fourth report.  2001.* http://www.parliament.the-stationery-office.co.uk/pa/ld200001/ldselect/ldsctech/57/5708.htm#a50 http://www3.interscience.wiley.com/cgi-bin/fulltext/97518189/PDFSTART

Hox J, Multilevel Modelling: When and Why? (1998) Pp 147-154 in: I. Balderjahn, R, Mather & M. Schader (Eds.). Classification, data analysis, and data highways.  New York: Springer Verlag.

Hull S, Rees Jones I, Moser K. Relation of rates of self referral to A&E departments to deprivation. *BMJ* 1998;317:538

Iezzoni LI, Ayanian J: Paying more fairly for Medicare capitated care.  *New England Journal of Medicine* 339(26):1933-1938, 1998

The Information Centre for Health and Social Care. Prescriptions Dispensed in the Community Statistics for 1996 to 2006: England, Published in 2007 http://www.ic.nhs.uk

Jick H, Jick SS, Derby LE. Validation of information recorded on general practitioner based computerised data resource in the United Kingdom. *British Medical Journal* 1991;302:766-8

Johns Hopkins University ACG Case Mix Adjustment System. Baltimore, Md: Johns Hopkins University School of Hygiene and Public Health; 2000. http://acg.jhsph.edu.

Journal of Health Planning and Management 2002,17(3),269-282.

Juncosa S, Bolibar B. A Patient Classification System for Our Primary Care: The Ambulatory Care Groups (ACGs). *Gac Sanit* 1997,11(2),83-94

Juncosa S, Bonaventura B, Roset M, Tomas R. Performance of an Ambulatory Case-Mix Measurement in Primary Care in Spain: Ambulatory Care Groups (ACGs). *European Journal of Public Health* 1999,9,27-35

Juncosa S, Carrillo E, Bolibar B, Prados A, Grevas J. Classification Systems in Case-Mix Groups in Ambulatory Care. Perspectives for Our Primary Health Care. *Aten Primaria* 1996:*17*(1);76-82

Kahn K, Rubenstein L, Draper D. The effects of the DRG based prospective payment system on quality of care for hospitalized Medicare patients. *Journal of the American Medical Association* 1990;264(15):1953-1997

Knutson D (1998). "Case-Study: The Minneapolis Buyers Health Care Action Group" Inquiry, 35, 171-17

Krieger N, Fee E. Social class: the missing link in US health data. *Int J Health* Serv 1994;24:25-44

Kronic R, Gilmer TP, Dreyfus T, Ganiats TG. CDPS-Medicare: The Chronic Illness & Disability Payment System modified to predict expenditures for Medicare benificiaries. *Final report to CMS* June 24, 2002 http://cdps.ucsd.edu/CDPS_Medicare.pdf.

Kronick R, Gilmer T, Dreyfus T, Lee L. Improving health-based payment for Medicaid beneficiaries: CDPS. *Health Care Financing Review* 2000; 21(3):29-64

Kronick R, Dreyfus T, Lee L, Zhou Z. Diagnostic risk adjustment for Medicaid: the disability payment system. *Health Care Financing Review* 1996;17(3):7-33

Kuk AYC. Asymptotically unbiased estimation in generalized linear models with random effects. *Journal of the Royal Statistical Society. Series B (Methodological)* 1995;57(2):395-407

Kuttner R. The risk-adjustment debate. *The New England Journal of Medicine* 1998;1952-1956

Kuttner R. Must good HMOs go bad? First of two parts: the commercialization of prepaid group health care. *The New England Journal of Medicine* 1998;338:1558-63

Larsen K, Merlo J. Appropriate assessment of neighbourhood effects on individual health: integrating random and fixed effects in multilevel logistic regression. *Am J of Epi* 2005;161(1):81-8)

Lawrence L, McLemore T. 1981 summary: National Ambulatory Medical Care Survey. *Adv Data* 1983;16(88):1-9

Lawson DH, Sherman V, Hollowell J. The General Practice Research Database. *QJM* 1998;91:445-52

Le Grand, J. The Blair Legacy? Choice and Competition in Public Services. Transcript of Public Lecture. London School of Economics 2006

Leyland AH, Goldstein H. (Eds.). (2001) *Multilevel modelling of health statistics.* Wiley Series in Probability and Statistics. Chichester: John Wiley & Sons Ltd.

Liang K, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika* 1986;73(1):13-22

Longford N. Standard errors in multilevel analysis. *Multilevel Modelling Newsletter* 1999;11(1):10-13

Machin D, Campbell MJ. Design of studies for medical research. New York: Wiley, 2005.

Majeed A, Bindman AB, Weiner JP. Use of risk adjustment in setting budgets and measuring performance in primary care I: how it works. *British Medical Journal* 2001;323:604-607

Majeed A, Bindman AB, Weiner JP. Use of risk adjustment in setting budgets and measuring performance in primary care II: advantages, disadvantages, and practicalities. *British Medical Journal* 2001;323:607

Majeed A, Malcolm L. Unified budgets for primary care groups. *BMJ* 1999;318:772-6

Majeed A. Allocating budgets for fundholding and prescribing. *BMJ* 1996;13:1274-572

Majeed FA, Cook DG, Evans N.  Variations in general practice prescribing costs: implications for setting and monitoring prescribing budgets. *Health Trends* 1996;28: 52-55

Majeed FA, Voss S. (1995)  Performance indicators for general practice. . *British Medical Journal* 1995;311:209-210

Majeed A, Bindman AB, Weiner JP. Use of risk adjustment in setting budgets and measuring performance in primary care I: how it works. *BMJ* 2001; 323: 604-607

Mann RD, Hall G, Chukwujindu J.  Research implication of computerised primary care. *Post-Marketing Surveillance* 1992;5:259-268

Mant, D.  National Working Group on R&D in Primary Care: final report. London NHS Executive 1997

McCormick, A., Fleming, D., Charlton, J. Royal College of General Practitioners, the Office of Population Censuses and Surveys, and the Department of Health. *Morbidity Statistics from General Practice. Fourth national study 1991-1992*, HMSO: London 1995

Meijer E, van erl Leeden R, Busing FMTA.  Implementing the bootstrap for multilevel models. *Multilevel Modelling Newsletter* 1995;7(2):7-11

Merlo J Multilevel analytical approaches in social epidemiology: measures of health variation compared with traditional measures of association. *Journal of Epidemiology & Community Health* 2003:550-552

Merlo J, Chaix B, Ohlsson H, Beckman A, Johnell K, Hjerpe P, Rastam L, Larsen K.  A brief conceptual tutorial of multilevel analysis in social epidemiology: using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology & Community Health*. 2006;60(4):290-7

Mills R, Fetter RB, Riedel DC, Averill R.  AUTOGRP: An interactive computer system for the analysis of health care data. *Medical* Care 1976;14(7):603-615

Moerbeek M.  The Consequence of Ignoring a Level of Nesting in Multilevel Analysis.  Multivariate Behavioural Research 39(1) 2004, p129-149

Morrell D, Gage H, Robinson N. Referral to hospital by general practitioners. *J R Coll Gen Pract* 1971;21:77-85

Muldoon JH, Neff JM, Gay JC. Profiling the health service needs of populations using diagnosis based classification systems. *The Journal of Ambulatory Care Management* 1997;20:1-18

Nazareth I, King M, Haines A, Rangel L, Myers S.  Accuracy of diagnosis of psychosis on a general practice computer system. *BMJ 1993*;307:32-34

Navein JF.  Guidelines partly explain differences in referral rates. *BMJ* 2002.  Internet communication NHS Executive  Quality and performance in the NHS: high level performance indicators. Leeds: NHS Executive 1999

O'Donnell CA.  Variation in GP referral rates: what can we learn from the literature? *Family Practice* 2000;17(6):462-471

O'Sullivan C, Omar RZ, Ambler G, Majeed A. Case-mix and variation in specialist referrals in general practice. *Br J Gen Pract* 2005;55:529-33

O'Sullivan C, Omar RZ, Forrest CB, Majeed A. Adjusting for case mix and social class in examining variation in home visits between practices. *Fam Pract* 2004;21:355-63

O'Donnell CA. Variation in GP referral rates: what can we learn from the literature? *FamPract* 2000;17:462-71

Office for National Statistics. *Key health statistics from general practice.* 1998. The Stationery Office, London. Series MB6 No. 2

Omar, RZ & O'Sullivan C (joint authors), Petersen I, Islam A, Majeed A. A model based on age, sex, and morbidity to explain variation in UK general practice prescribing: a cohort study. *BMJ* 2008;337:a238

Omar, R.Z., Thompson S.G. (2000) Analysis of a cluster randomized trial with binary outcome data using a multi-level model. *Statistics in Medicine*;19:2675-2688

Orueta J F, Lopez-De-Munain J, Baez K, Aiarzaguena JM, Aranguren JI, Pedrero E. Application of the Ambulatory Care Groups in the Primary Care of a European National Health Care System: Does It Work? *Med Care* 1999:37(3),238-248. http://www.medicalcare.com/pt/re/medcare/fulltext.00005650-199903000-00004.htm

O Sullivan C, Omar RZ, Ambler G, Majeed A, Case-mix and variation in specialist referrals in general practice *Br J Gen Pract.* 2005 July 1; 55(516): 529–533

Park Nicollet Medical Foundation, The Johns Hopkins University. *Final Report: A comparison of alternative approaches to risk measurement.* Washington, DC: The Physician Payment Review Commission. 1994

Peckham, S., Exworthy, M. Primary care in the UK: Policy, organisation and management. Palgrave Macmillan 2003

Pope GC, Randall MS, Ellis P, Ash AS, Ayanian JZ, Bates DW, Burstin H, Iezzoni LI, Marcantonio E, Wu SMB. Diagnostic Cost Group Hierarchical Condition Category Models for Medicare Risk Adjustment. Final Report for Health Care Financing Administration 2000

Public Health Service and Health Care Financing Administration: *International Classification of Diseases, 9th Revision, Clinical Modification.* U.S. Department of Health and Human Services. Washington, DC. U.S. Government Printing Office, September 1980.

Prescribing costs in primary care. National Audit Office: London, 2007. www.nao.org.uk/publications/nao_reports/06-07/0607454.pdf

Rasbash J, Browne W, Goldstein H, et al. A user's guide to MlwiN. Multilevel Models Project, Institute of Education, University of London 2000.

Rasbash J, Browne W, Goldstein H, Yang M, Plewis I, Healy M, et al. A users guide to MLwiN, version 2.1. London: Institute of Education, 2000

Reid R, MacWilliam L, Roos NP, Bogdanovic B, Black C. Measuring morbidity in populations: performance of the Johns Hopkins Adjusted Clinical Group (ACG) Case-Mix Adjustment System in Manitoba. 1999.

Reid RJ, MacWilliam L et al. Performance of the ACG Case-Mix System in two Canadian provinces. *Med Care* 2001;39(1): 86-99

Reid RJ, Roos NP, MacWilliam L, Frohlich N, Black C. Assessing population health care need using a claims-based ACG morbidity measure: a validation analysis in the province of Manitoba. *Health Serv Res* 2002;37(5): 1345-64

Reid, F., Cook, D.G., Majeed, F.A. Explaining the variation in general practice admission rates: cross-sectional study. *BMJ* 1999,319:98-103

Ridout, M.S., Demetrio, C.G.B., Firth, D. (1999) Estimating intraclass correlation for binary data. *Biometrics*;55:137-148

Robinson J.W., Zeger S.L., Forrest C.B. (2006). "A Hierarchical Multivariate Two-Part Model for Profiling Providers' Effects on Health Care Changes." *Journal of the American Statistical Association*, 101(475), 911-923

Rodriguez G, Goldman N. An assessment of estimation procedures for multilevel models with binary responses. *Journal of the Royal Statistical Society. Series A (Statistics in Society)* 1995;158(1):73-89

Russel, L. Medicare's new hospital payment system. The Brookings Institute, Washington D.C., 1989

Saint-Yves I F. The Read clinical classification. *Health Bulletin* 1992;50(6):422-427

Salem-Schatz S, Moore G, Rucker M, and Pearson SD. The Case for Case-Mix Adjustment in Practice Profiling. When Good Apples Look Bad. *JAMA* 1994;272(11): 871-874

Scoggins A, Tiessen J, Ling T, Rabinovich L. Prescribing in primary care, Understanding what shapes GPs' prescribing choices and how might these be changed. Technical report prepared for the National Audit Office. Cambridge, UK: RAND Corporation, 2006. www.ppa.org.uk/pdfs/publications/SMT_V&C_report_200606.pdf.

Scott A, Shiell A. Analysing the effect of competition on general practitioners' behaviour using a multilevel modelling framework. *Health Economics* 1997(a);6:577-588

Scott A, Shiell A. Do fee descriptors influence treatment choices in general practice? A multilevel discrete choice model. *Journal of Health Economics.* 1997(b);16:323-342

Scrivener C, Lloyd DCEF. Allocating census data to general practice populations: implications for study of prescribing variation at practice level. *BMJ* 1995; 311:163-165

Secretary of State for Health. The NHS plan: a plan for investment, a plan for reform. London: Stationery Office 2000. (Cm 4818-I.)

Sevcik AE, Abu-Jaber T, Marek L. Understanding approaches to case-mix assessment and case-mix adjustment. *JHQ Online*, Sept/Oct, pp. W5-24–W5-29. National Association for Healthcare 2004

Shenkman EA and Breiner JD. Characteristics of risk adjustment systems. 2001. Ref Type: Report

Signorini DF, Weir NU. Any variability in outcome comparisons adjusted for case mix must be accounted for. *British Medical Journal* 1999;318:128.

Snijders TAB, Bosker RJ. Multilevel analysis: an introduction to basic and advanced multilevel modelling. London: *Sage Publishers*, 1999

Starfield B, Hankin J, Steinwachs D, Horn S, Benson P, Katz H, Gabriel A. Utilization and morbidity: Random or tandem? *Pediatrics* 1985;75(2):241-247

Starfield BH, Weiner JP, Mumford L, Steinwachs D. Ambulatory Care Groups: A categorization of diagnoses for research and management. *Health Services* 1991;26(1):53-74

StataCorp. Stata statistical software: Release 9. College Station, TX: StataCorp, 2005

Sutch S, Lemke K, Weiner J, Kinder K. The application of ACG predictive models to the English population for the purposes of funding allocation. *BMC Health Services research* 2008, 8(Suppl 1):A15 (Meeting abstract)

The Johns Hopkins University ACG Case-Mix System. 2008. www.acg.jhsph.edu/

Thiru K, Hassey A, Sullivan F. Systematic review of scope and quality of electronic patient record data in primary care. *British Medical Journal* 2003;326;1070-0

Turner RM, Omar RZ and Thompson SG. Bayesian methods of analysis for cluster randomised trials with binary outcome data. *Statistics in Medicine* 2001;20:453-472

Young WW, Swinkola RB, Hutton MA. Assessment of the AUTOGRP Patient Classification System. *Medical Care* 1980;18(2):228-244

Valderas JM, Starfield B, Roland M. Multimorbidity's many challenges. A research priority in the UK. *BMJ* 2007;334:1128

Watkins C, Harvey I, Carthy P, Moore L, Robinson E, Brawn R. Attitudes and behaviour of general practitioners and their prescribing costs: a national cross sectional survey. *Qual Saf Health Care* 2003;12:29-34

Webb S, Lloyd M. Prescribing and referral in general practice: a study of patients' expectations and doctors' actions. Br J Gen Pract 1994;44:165-9

Weiner J P, Starfield B H, Steinwachs D M, Mumford L M. Development and application of a population-oriented measure of ambulatory care case mix. *Medical Care* 1991;29:453-472

Weiner JP, Abrams C. ACG Virtual Library: Version 4.5 (February 2000) ACG Software Documentation/Users Manual. *Health Services Research & Development Centre, The Johns Hopkins University.* 1991-2001

Weiner JP, Starfield BH, Steinwachs DM, Mumford LM. Development and application of a population-oriented measure of ambulatory care case-mix. *Medical Care* 1991;29:452-72

Weiner JP, Tucker AM, Collina AM, et al.: The development of a risk-adjusted capitated payment system: The Maryland Medicaid Model. *Journal of Ambulatory Care Management* 21(4):29-52, 1998.

Weisberg S. Applied linear regression. 3rd ed. New York: Wiley, 2005

Woodhouse G, Yang M, Goldstein H, Rasbash J. Adjusting for measurement error in multilevel analysis. *Journal of the Royal Statistical Society. Series A(Statistics in Society)* 1996;159(2):201-202

Wrightson, CW (2002). Financial Strategy for Managed Care Organisations. Rate Setting, Risk Adjustment, and Competitive Advantage. Health Administration Press

Zielinski A et al 2009. Validation of ACG Case-mix for equitable resource allocation in Swedish primary health care. *BMC Public Health*, 9:347