ERASMUS UNIVERSITEIT ROTTERDAM
ERASMUS SCHOOL OF ECONOMICS

The effect of health provider density on health care utilization and inequity in Peru

MSc in Health Economics

Academic year 2015 – 2016

Name Student:	Erna Berends
Student number:	331416
Supervisor:	dr. S. Neelsen, Institute of Health Policy & Management
Co-reader:	Prof. dr. E. van Doorslaer, Department of Applied Economics

Abstract

Background: this study examines the effect of health provider density on the probability of health care utilization for the publicly- and uninsured population in Peru. Evidence on determinants of health care utilization in developing countries is low. The other objective is to examine to what extent health provider density creates inequality in health care utilization, and the underlying factors causing inequity for this part of the population in Peru. This thesis contributes to the knowledge about the effect of health provider density on health care utilization and inequity, in a developing country. It is specifically interesting for policy makers.

Methods: the study uses eight binary measures of health care utilization, which are utilization of GP consults, analysis, scans, other tests, ophthalmology, medicines and health care use in general. A logistic regression method is used to obtain the relations. The main explanatory variables are doctor density, nurse density and skilled health provider density (doctors + nurses). To measure inequality concentration curves are obtained showing the distribution of provider density against health care utilization. Health care utilization is standardized by age, gender and health status. Inequality is decomposed to obtain the factors causing inequity in utilization of GP visits and health care utilization in general.

Results: skilled health providers are significantly and positively related to utilization of GP consults and ophthalmology. The results are rather contrasting because doctors show a positive and nurses a negative relation with utilization of health care. Doctor density is positively and significantly associated with utilization of GP consults, analysis, other tests and health care overall. Nurse density is negatively and significantly related to utilization of other tests. The three groups of health provider density are not correlated with the other measures of health care utilization. Inequity exists favoring people living in high provider density regions for utilization of GP consults and health care overall. Decomposition shows that doctor density and wealth are the most important contributors of the inequity.

Conclusion: the results indicate that especially physician density has a positive relation with health care utilization, especially GP consults. The weakened results when controlling for confounding factors indicate that health provider density is not the only explanatory factor for health care utilization. It is plausible that the quality of provided services plays an important role. Furthermore, the inequity analysis shows that it would be effective to distribute health providers to areas with low provider densities.

Acknowledgement
I chose the MSc Health Economics because I am highly interested in the related topics. The master provided me a lot of new knowledge, tools and insights to understand and evaluate relevant matters. I enjoyed the program very much and this thesis is the final step in completing it.

I would like to thank my supervisor Sven Neelsen, who has supported me during the thinking and writing process. His help was very valuable and is much appreciated. Furthermore I would like to thank Eddy van Doorslaer for the opportunities to brainstorm about the thesis topic and his co-evaluation. When I started the master I had already decided to travel to Peru in the summer of 2015, therefore I enjoyed it very much that I could write my thesis about Peru.

Table of Contents

Abstract **2**
Acknowledgement **3**
List of Figures **5**
List of Tables **5**
List of Abbreviations and Acronyms **6**
1 Introduction **7**
2 Determinants of Health Care Utilization **9**
2.1 Theory 9
2.2 Empirical Evidence 11
3 Inequity in Health Care Utilization **13**
3.1 Theory 14
3.2 Empirical Evidence 14
4 Background on the Peruvian Health Care System **16**
5 Methodology **19**
5.1 Research Design 19
5.2 Data Collection and Model Specification 21
5.3 Measuring Inequality 28
5.4 Decomposing Inequity 28
6 Results **30**
6.1 Descriptive Statistics 30
6.2 Regression results 33
6.3 Results on Inequity 35
6.4 Additional Analysis 39
7 Conclusions and Discussion **41**
7.1 Summary of the Main Findings 41
7.2 Discussion of the Results 42
7.3 Limitations 45
7.4 Policy Implications 46
7.5 Suggestions for Future Research 47
References **48**
Appendix **52**

List of Figures

Figure 1. Revised health behavioral model of Andersen (1995) ... 10
Figure 2. Distribution of physicians by level of district poverty, rate of medical professionals / 10,000 inhabitants, retrieved from World Bank (2011) ... 18
Figure 3. Visualization of marginal effects in a nonlinear model by Long and Freese (2001) ... 20
Figure 4. Variations of skilled health workers per province in 2007. Source: own elaboration ... 32
Figure 7. Utilization of inpatient and outpatient health care per province in 2010. Source: own elaboration ... 32
Figure 5. Concentration Curves for the health care utilization measures USE01 and USEhc ... 37
Figure 6. Graphical decomposition of inequity for the health care utilization measures ... 38

List of Tables

Table 1. Description of the dependent variables ... 23
Table 2. Description of the aggregated measures USEophc and USEhc ... 23
Table 3. Descriptive statistics of the dependent variables ... 31
Table 4. Descriptive statistics of the explanatory variables ... 31
Table 5. Results of the univariate logistic regressions – average marginal effects ... 34
Table 6. Results of the multivariate logistic regressions – average marginal effects ... 34
Table 7. Inequality in health care utilization ... 37
Table 8. Decomposition of the CI for health care utilization ... 39
Table 9. Results for multivariate OLS regressions ... 40
Table 10. Pearson correlations ... 53
Table 11. Descriptive statistics of the control variables ... 54
Table 12. Results of the control variables for logistic regressions Model 1 with Skilled- average marginal effects ... 56
Tabel 13. Results of the control variables for logistic regressions Model 2 with Doctors & Nurses- average marginal effects ... 58
Table 14. Fitstat measures probit and logit for *USEhc* (left: models with skilled, right: models with doctors & nurses) ... 60

List of Abbreviations and Acronyms

AUS	Universal Health Insurance Law (*Aseguramiento Universal en Salud*)
BRM	Binary Regression Model
ENAHO	National Household Survey (*Encuesta Nacional de Hogares*)
GP	General Practitioners
HRO	Human Resource Observatory
INEI	National Institute for Statistics (*Instituto Nacional de Estadistica e Informatica*)
LAC	Latin American Countries
NGO	Non-Governmental Organization
NHS	National Health System
MDG	Millennium Development Goals
MINSA	Ministry of Health (*Ministerio de Salud)*
OECD	Organization for Economic Cooperation and Development
OLS	Ordinary Least Squares
PAHO	Pan American Health Organization
SE	Standard Error
SID	Supplier-Induced Demand
SIS	Integral Health Insurance (*Segura Integral de Salud*)
SP	Specialist Physicians
WHO	World Health Organization

1 Introduction

As stated in the Millennium Development Goals (MDG) of the United Nations, by 2015 child and maternal mortality should have been reduced sharply and infectious diseases should at least be halted (United Nations, 2005). Although the 20^{th} century knows many advanced developments in the field of health, poor countries are still confronted with health crises. Some researchers suggest that lack of human health resources are part of this problem. For instance Chen et al. (2004) suggest that Sub-Saharan countries should enlarge their health workforce three times in order to approximate accomplishing the MDG (Chen et al., 2004). Sundmacher and Busse (2011) state that a larger density of health providers leads to better health outcomes, like a decrease in cancer mortality (Sundmacher and Busse, 2011). Furthermore, evidence exists that health worker availability correlates with decreasing infant and maternal mortality in several countries (Anand and Bärnighausen, 2004).

However, many studies about health provider density concentrate on developed countries (e.g. Eibich and Ziebarth, 2013, Continelli et al., 2010, Busato and Künzi, 2008). Other studies concentrate on the effect on specific health outcomes (e.g. Sundmacher and Busse, 2011, Anand and Bärnighausen, 2004, Valdivia, 2004). A general outcome enhances the validity of a research for the health care system as a whole. Yet, only a small amount of quantitative research exists on the relation between health worker density and health care utilization (Kruk et al., 2009). Therefore, it is a relevant contribution to the existing literature to study the relationship between health provider density and health care utilization.

Since research on this topic is mostly performed in developed countries, I choose a developing country that invested in the public health workforce in recent years: Peru.

Peru belongs to the group of middle income countries, with a GDP of $4,552 per capita in 2010 (Atun et al., 2014). 28% of the population lived below the poverty of regionally differentiated poverty lines of $1 and $2 (Francke, 2013). The average life expectancy at birth for men was 71.4 and 76.6 for women in 2010 (Atun et al., 2014). In comparison to the South American region, Peru had a high rate of maternal mortality and under-5 mortality in 2010. Furthermore, the country had the lowest rate of births attended by professionals in comparison to other countries in Latin America (Carpio and Santiago Bench, 2015). These facts indicate there is a large opportunity to improve health care utilization in Peru.

During the last decade, the country made a considerable good effort to enlarge their health workforce. The total staff working for the NHS grew from 87,588 to 107,983 in 2007 to 2009 (MINSA, 2015) From a policy perspective, it is of great importance to know what the return on the investment in human resources will be. Baker and Liu (2006) argue that policy decisions about health planning are often made without valid evidence from scientific research and understanding of the characteristics determining health care utilization. This particularly holds for poorer and rural regions (Baker and Liu, 2006). Therefore the results are not only relevant from an academic perspective, but also for the purpose of policy decision making.

The MDG framework is lacking attention to equity, particularly on the goals related to health (Cometto and Witter, 2013). However, not only the total number of health providers counts. The distribution and access among socioeconomic groups is important as well (Gerdtham, 1997). Many countries show evidence of inequality in health care utilization (i.a. Gerdtham, 1997, Makinen et al., 2000, Ozegowski and Schumacher, 2013). In Peru, the distribution of health providers is not only distributed unequally across the different regions in Peru, but also across the different quintiles of wealth (World Bank, 2011). Valdivia (2004) found that for Peru the effect of investments in e.g. reducing waiting times and new facilities profited especially the poor. However in 2006, the distribution of health care workers was still distributed very unequal across the different income quintiles (World Bank, 2011). Inequity contains an ethical dimension, because differences in health delivery are considered as unfair (Whitehead, 1991). Considering the recent investments of the Peruvian government, it is interesting to research the effects on inequality in health care utilization as well.

The objective of this study is therefore to research the following two issues. First, what the impact of health provider density is on utilization of health care, for the publicly- and uninsured population in Peru. Second, examine to what extent health provider density creates inequality in health care utilization for this sample in Peru and to determine the factors causing inequity.

I measure the first topic using a logistic regression method measuring the dependent variable health care utilization by eight different measures. Health provider density is measured by a group of doctors, a group of nurses and an aggregate group of skilled health providers. I analyze the average marginal effects, as these reflect the probability with which an increase in provider density affects the health care utilization rate. For the second matter, I construct concentration curves to measure inequality. Subsequently, I perform a decomposition of inequity to derive whether and to what extent the inequality in health care utilization by provider density is driven by socio-economic factors.

I find a contrasting relationship between doctors (*positive*) and nurses (*negative*) with the probability of health care utilization. However the associations are not very significant. Doctor density significantly relates to the probability of using more GP consults, analysis, other tests and health care in general. Nurse density is only negatively and significantly related to utilization of other tests, like hemodialysis.

Inequity in health care utilization due to differences in provider density exists for utilization of GP consults and utilization of health care in general. The inequity is biased towards people living in areas with high physician density, but it only exists to a small extent. The most important contributors are doctor density and wealth.

The structure of the study is as follows. First, chapter 2 provides a theoretical framework and empirical evidence on the determinants influencing health care utilization. Chapter 3 describes the underlying theory and empirical evidence for equity in health care. The setting of the Peruvian health system is pictured in chapter 4. Chapter 5 explains the methodology of the research and chapter 6 analyzes the results. A discussion of the validity of the results and conclusions are provided in chapter 7.

2 Determinants of Health Care Utilization

This chapter contains a discussion of the theory and important literature. It elaborates on the effect of supply side determinants on health care utilization. Paragraph 2.1 presents a behavioral model which can explain the determinants influencing health care utilization. The next section explicates the empirical evidence.

2.1 Theory

To explain what drives the utilization of health care of an individual, the revised health care behavioral model of Andersen (1995) can be used. Although the original model was already developed in 1968, the principles of this theoretical framework have been used in much of the health care utilization literature (Phillips et al., 1998). Figure 1 depicts the model which describes the societal and individual determinants that influence use of health services. The variable of interest, availability of health care workers, is embedded in the factor ‘health care system’. The health care system comprises the delivery of health services and contains the features ‘organization’ and ‘resources’. Resources comprise the human resources and available technology, and organization comprises the practical execution of the resources.

The underlying hypothesis of health providers influencing health care utilization is that health utilization increases as the ratio health workers/population increases, ceteris paribus (Andersen and Newman, 1973).

Factors directly influencing health care utilization are predisposing characteristics, enabling resources and need. The predisposing factors contain characteristics like race, age and gender, but also health beliefs, like knowledge and values towards health and the system. These aspects determine that some people have a greater tendency to use health care. Enabling resources comprise organizational factors varying per individual, e.g. access to health care and insurance, income and waiting times. Thus, these components are derived from different characteristics of the population, environment and health outcomes (Andersen, 1995). More health workers can for instance reduce waiting times, because more people can be treated at the same time. More supply leads to higher competition, which can reduce prices. Furthermore, increased provider density is probably correlated with an increase in facilities, which leads to a smaller travel distance. These are examples of enabling factors. Hence, in that way health providers influence enabling factors which influence the health behavior.

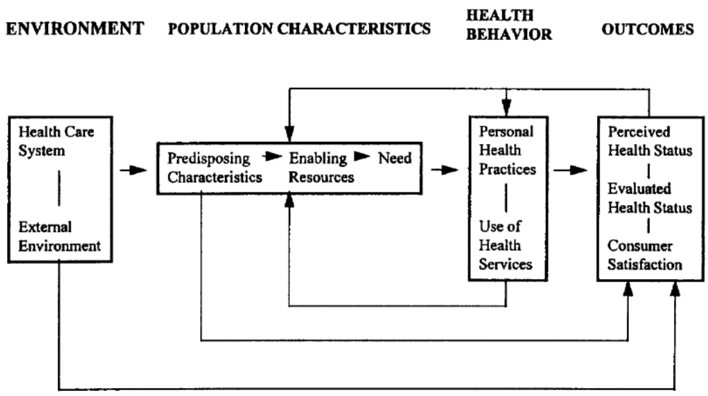

Figure 1. Revised health behavioral model of Andersen (1995)

Another theory that should be considered is the theory of supplier-induced demand (SID). Information asymmetry between patients and physicians causes health consumers to be unable to make rational decisions regarding health care utilization. (Wennberg, Barnes and Zubkoff, 1982). It means that physicians can induce patients to use health care services, instead of a patient deciding to use health care, determined by its own preferences. A normative and positive approach exist to explain this matter. The definition of Wennberg et al. (1982) is considered to be normative: the physician defines the need of a patient and evaluates prices.

The physician might therefore manipulate demand of the patient to his own advantage (Wennberg et al., 1982). On the other hand, 'the physicians ability to shift the patients' demand for medical care at a given price' as described by Hadley et al. (1989) is an example of a positive definition of SID (Labelle et al., 1993). However it is extremely challenging to ultimately prove the existence of supplier-induced demand empirically since, for example, one needs to disentangle patient preferences from typically unobserved physician inducement (Eibich & Ziebhart, 2013).

The hypothesized relationship between health worker availability and the degree of health care utilization is therefore rather complex. As is visible from the behavioral model, the explanatory variables of health care utilization are correlated. The components of the health care system not only influence use of health services, they also show a relationship with health outcomes, like consumer satisfaction. E.g. as the ratio health workers/population increases, patients receive more personal attention from physicians and consumer satisfaction increases, keeping other factors constant. This influences the population characteristics field, which is already influenced by the supply side of the health care system. Moreover, health care utilization may also influence the supply of health workers, as higher demand for health care attracts more health providers, perhaps resulting in SID.
Considering this, the relationship between health care utilization and health care workers and other explaining variables, presumably contains reverse causality. Furthermore, concerning the components of the behavioral model, omitted variable bias is likely to occur. The presence of unobservable factors which causally affect both health worker supply and health utilization cannot be precluded. Therefore, this study only provides correlations and no causal relationship.

2.2 Empirical Evidence

This section elaborates on evidence about the effect of health provider availability and other supply side determinants on health care utilization.
Research shows that for rural parts in Honduras, walking distance to a medical center has a strong and negative association with health care utilization (Baker and Liu, 2006). This suggests utilization of health care might improve with smaller walking distance and therefore by a higher health worker density. Physician supply is in fact found to be an important factor in determining health care utilization. Busato and Künzi (2008) conducted a large small-area variation research in utilization and provision of health care resources. It concentrated on the multicultural population in Switzerland. The authors made a distinction between primary care

physicians and specialist physicians as attributes for the health care system. The dependent variable they used was consultation rate, which is also one of the dependent variables this study uses. Most important findings are that the density of general practitioners (GP) was significantly and strongly positive associated with the utilization of GP consults. I.e., for every additional GP per 10.000 inhabitants, utilization of GP consults increased with 0.10. Specialist physicians (SP) had an opposing effect: every extra SP was associated with an average decrease of GP consultations of 0.01. The association persisted for every region, however the magnitude differed between the various language regions in Switzerland (Busato and Künzi, 2008).
Kruk et al. (2009) performed a large-scale research among 106 developing countries about the effect of health workers on six specific health outcomes. Health providers were divided into a group of doctors, nurses & midwives and an aggregated group of both; skilled health workers. The researchers found that skilled health provider density is positively correlated with use of skilled birth attendants and measles immunization. In the disaggregated analysis they found that doctor density is positively related with use of measles immunization and nurse density with the use of skilled birth attendants. The researchers did not find any significant correlations for the other four dependent variables. Multiple explanations for the lack of association are given. Besides methodological reasons, it could be that health services are performed by health providers not registered in the used dataset. It is common among developing countries to have health providers with low education levels who cannot be recognized as physicians. Hence, if these community workers are not included a relation cannot be established (Kruk et al., 2009).

Research from several regions in the US indicated that the GP supply has a positive correlation with having a primary care physician. Furthermore, this implied a larger use of preventive health services (Continelli et al., 2010). Other studies showed that having a primary care physician indeed implied more preventive care use, because utilization rates of specialists reduced simultaneously (Kravet et al., 2008, Eibich and Ziebarth, 2013). Having a primary care doctor eventually reduces utilization of health care in the US significantly according to Kravet (2008). The health care utilization indicators for which this effect occurred, were emergency department visits, hospitalization admissions and surgeries (both inpatient and outpatient). In general, these measures contain a large inpatient care element. These findings suggest primary physician can be useful in containing inpatient care utilization, which can be an incentive to contain health care expenditures (Kravet et al., 2008).

The analysis of Wright and Ricketts (2010) shows similar findings of a higher provision of primary care supply reducing inpatient admissions. On the other hand the researchers argue that one should be careful with drawing conclusions. Their results indicate that the finding does not hold on country level, because the association between primary care provision and outpatient consults and total number of surgeries was insignificant. This suggests that the method of analyzing geographic areas matters (Wright and Ricketts, 2010).

Eibich and Ziebarth (2013) used a comprehensive household panel dataset to explain regional differences in medical care utilization in Germany. The authors argue that the individual demand side determinants can help explain the regional variation in health care utilization. The density of hospital beds (to explain inpatient care) and physician density (to explain outpatient care) are used to represent the supply side. The model is supplemented with many individual characteristics. The coefficients for physician density show that there is a negative relation with inpatient care and an insignificant association with outpatient care. Furthermore, an inverse relation is found again between significantly high levels of hospitalization and low levels of doctor visits and vice versa per state. The results show no differences in utilization of inpatient and outpatient care between states when controlled for the individual characteristics (Eibich and Ziebarth, 2013). Higher levels of physician provision are correlated with better access to health care, but they are also correlated with specific characteristics like insurance status and income. When adjusted for the latter factors, a significant association between health worker availability and medical care access in urban California, US no longer exists (Grumbach et al., 1997). Therefore, it is important to add variables in the model to control for socioeconomic or other circumstances.

In general, based on the existing literature can be concluded that supply side determinants, and thus provider density, have a positive relationship with health care utilization. However, this mainly holds for general practitioners and the effects differ for inpatient and outpatient care. Controlling for confounding factors is highly important.

3 Inequity in Health Care Utilization

Chapter 3 discusses the concept of inequity. Similar to the previous chapter, paragraph 3.1 elaborates on a theory about the influence of health provider density and equity in health care utilization. Paragraph 3.2 provides empirical evidence related to this topic.

3.1 Theory

First I provide a brief explanation of inequity. Horizontal equity in health care delivery refers to a situation where people with equal needs are treated equally, regardless any other characteristics like race, gender or income. Vertical equity means an unequal treatment for people with unequal needs (Wagstaff, van Doorslaer and Paci, 1990). This study concentrates on horizontal equity. Inequity should not be interchanged with inequality in health. The latter indicates a variation in health in a certain population (Wagstaff and van Doorslaer, 1998). Equity on the other hand distinguishes a systematic inequality in health between different socioeconomic groups, which creates the general view of unfairness (Braveman and Gruskin, 2002).

The conceptual framework of Andersen (1995) has shown that socioeconomic characteristics of individuals influence the degree of health care utilization. Some people have a greater propensity of using health care than other based on predisposing characteristics, including demographic factors (Baker and Liu, 2006). The enabling components in the model allow an individual or family to satisfy their need for health, because of socioeconomic factors like health coverage, income and level of education (Andersen and Newman, 1973). In other words, people with certain attributes unrelated to need for health are more likely to use health care than others. Besides individual level characteristics, enabling characteristics on community level influence the use of health services as well, like health provider availability. A growing health worker population is expected to reduce prices and increase supply of services, hence enhancing access to health care (Andersen and Newman, 1973). Therefore, it can be hypothesized that horizontal inequity in health care use declines when provider density increases.

3.2 Empirical Evidence

The scientific research on the topic health equity experienced a large increase during the years 1980 until 2005, among others because of an increased demand from policy makers, NGO’s and others (O’Donell et al., 2008). Also in well developed countries, inequity in health care exists (i.a. Gerdtham, 1997, van Doorslaer, Masseria and Koolman, 2006, Ozegowski and Schumacher, 2013). This section provides an overview of published literature on the determinants of inequity in health care delivery, concentrating on health provider density.

Evidence from Sweden indicates horizontal inequity exists in the delivery of health care. Need, defined as morbidity, is associated with health care utilization. Socioeconomic factors like professional status, income and place of residence are significant as well (Gerdtham, 1997). Furthermore in 2000 reseachers found that in half of 21 OECD countries a need-standardized doctor use in favor of the rich exists. This horizontal inequity was measured by comparing the need-expected distribution of use among individuals, with the actual use of health care by income. Characteristics like age, gender and self-assessed health were included as predictors. Again, the probability of visiting a GP was more equitable than the probability of visiting a SP and the existence of private health insurance strengthens this inequity. (van Doorslaer et al., 2006). Research from eight developing countries shows that the underlying explanations for inequalities are different per country. Several inequality patterns in utilization of health care are visible. Rich people seeking health care were more likely to actually be treated by a doctor than ill poor people (except for Kyrgyzstan). There was no clear pattern for the countries across income quintiles for treatment in a hospital, which can be explained by differences between the countries. For instance Kyrgyzstan and Kazakhstan supplied more hospitals in rural and poor areas. Furthermore, in most countries rich persons used less hospital care, since most hospitals are facilitated publicly and contain less quality than the scarce private clinics (Makinen et al., 2000).

The prior studies did not consider the additional effect of supply side determinants on inequity. Ozegowski and Schumacher (2013) did examine the impact of outpatient care supply on regional inequities in outpatient care provision in Germany. Total supply side determinants of health care accounted for half the explanatory power of the model. The density of SP's and GP's were both associated with a lower equity index. This implies that people living in a dense provider region receive more care than needed, hence explaining that people living in a region with lower provider density receive less health care due to the fact of lower provider density. Furthermore, the researchers found that in a situation with two persons with an equal need, the person with a higher frequency in GP contacts consumes less outpatient care in total. A high density of SP's is rather associated with more outpatient care supply than needed. (Ozegowski & Schumacher, 2013).

Specifically for Peru, Valdivia (2002) found large differences in utilization of outpatient care for different income groups for the year 1997. During the last decade of the 20th century, the Peruvian government expanded the public health infrastructure in the country and the author questioned whether this led to a more equitable situation. This was estimated by a random effects model including an interaction term between income and health infrastructure. Valdivia found that public health infrastructure expansion worked equity-enhancing, although the effect was rather small (Valdivia, 2002). Furthermore, the large investments in the public health infrastructure in Peru included reducing travel distances to health facilities, reducing waiting times and implementing health facilities on new locations. The effect of these interventions was associated with better chronic nutrition of children aged under five years and had a pro-poor bias (Valdivia, 2004).

Based on the empirical evidence it can be concluded that socioeconomic factors play a relevant role in the degree of equity in health care utilization. In general a pro-rich use of health care exists in most countries. Though, with a pro-poor enforcement of supply side elements, equity seems to improve.

4 Background on the Peruvian Health Care System

This chapter provides a background on the Peruvian health care system in terms of organization, problems and developments during recent years.

The Peruvian health care system is divided into several financing funds. The government serves the poorest and necessitous people through a National Health service System (NHS). The formal sector workers are secured through a social security system (*EsSalud*), while the military and police forces have their own health insurance systems. A private insurance scheme also exists, mainly accessible for the wealthy part of the population. The health system is fragmented and the subsystems work independently (World Bank, 2011). According to the Pan American Health Organization (PAHO), in 2010 62.6% of the Peruvian population was covered by health insurance: 37% was enrolled in the NHS, 21.1% in social security and 5.5% had a private insurance. (PAHO, 2012). This study concentrates on people entitled to the MINSA, i.e. NHS. This system targets the poor and has the most to gain in terms of health outcomes for the poor and mostly uninsured people. Moreover, the government invests valuable resources in order to enlarge this system. Hence it is useful to evaluate the effect of the additional health providers for political purposes.

The largest scheme of the NHS is the Integral Health Insurance (SIS) program. Originally SIS concentrated mainly on maternal and child health, but since 2006 it targets the total poor, uninsured population. It aims on practicing a fee waiver system for participants for all basic care in NHS facilities (World Bank, 2011). In 2009 the government passed a law, the *Aseguramiento Universal en Salud* (AUS). This law established a list with a basic package of care with guaranteed coverage, and was to be enforced after 2010 (World Bank, 2010).

Deficiencies in the supply of health care are a constraint in improving health care. Díaz et al. (2009) found that inadequate supply of health and nutrition services in Peru was a limiting factor in producing better outcomes for the SIS program (World Bank, 2011). Especially in the NHS system exists a gap in the amount of physicians and the desired quality of services. This is concerning considering that the NHS accounted for 58% of all nurses and 62% of the physicians in the country in 2007. Moreover, the NHS targets to provide coverage to approximately 73% of the Peruvian population, both SIS beneficiaries and the uninsured (World Bank, 2011). Peru was identified as one of the 57 countries worldwide where a serious deficiency of health workers exists, lacking medical specialists. According to the WHO it deteriorates quality and attractiveness of services which leads to a lower utilization rate (World Health Organization, 2010). A benchmark from the Health Human Resource Observatory (HRO) indicates there should be 25 professional health providers per 10,000 inhabitants in Peru. In 2006, there were only 9.4 professional health providers/10,000 inhabitants (World Bank, 2011).
A serious problem threatening provider density in Peru is outmigration of health professionals. In 2010, 78% of the medicine students indicated they wanted to migrate after their graduation (Jiménez, Mantilla, Huaynay et al. 2015). Compared to other LAC, the actual number of migration in Peru lies above the average (Carpio and Santiago Bench, 2015). Students indicated that the desire to migrate is strongly motivated by a better economic perspective outside of Peru (Jiménez et al. 2015). This is linked with the method of remuneration and contracting. In 2008 the Ministry of Economic and Financing began to introduce the Results Based Budget (PpR) as new remuneration system. It started with five pilot programs, increasing to ten in 2010. Salary and financing became linked to certain goals of performance. Examples are the improvement of children's growth and reducing maternal mortality (World Bank, 2011). In the NHS health workers are contracted in various ways, with large differences in remuneration as a consequence. This especially holds for physicians. Moreover, in 2013 only 47% of the health professionals had a permanent engagement, 33% had a temporary contract and the remaining part consisted of different kinds of engagements. (Jiménez et al., 2015). The data provided on the website of the NHS however contains all employees, disregarding type of contract. Hence in this study takes all health providers working for the NHS into account.

Besides a general shortage of health workers, supply is unevenly distributed across regions. Staff shortage is a binding constraint in especially rural areas. In 2006 the density ratio of

9.4/10,000 inhabitants diverged between provinces with an average of 16.8 in Lima to 3.3 in Cajamarca (World Bank, 2011). A goal of the HRO was to cut the difference in provider density between rural and urban regions by half in 2015 (MINSA, 2009).
Moreover, the distribution sorted by wealth is unequal. Figure 2 illustrates this unequal distribution for four different years, where quintiles are divided by the level of district poverty. So quintile I represents the poorest 20% of districts and quintile V the 20% richest districts. The x-axis indicates the different quintiles and the y-axis measures the medical professionals per 10,000 inhabitants. The number of physicians / 10,000 inhabitants living in the poorest quintile of districts, increased from 0.8 in 1992 to 4.2 in 2006 (World Bank, 2011). The years depicted in the figure are 1992 (blue), 1996 (red), 2004 (green) and 2006 (purple).

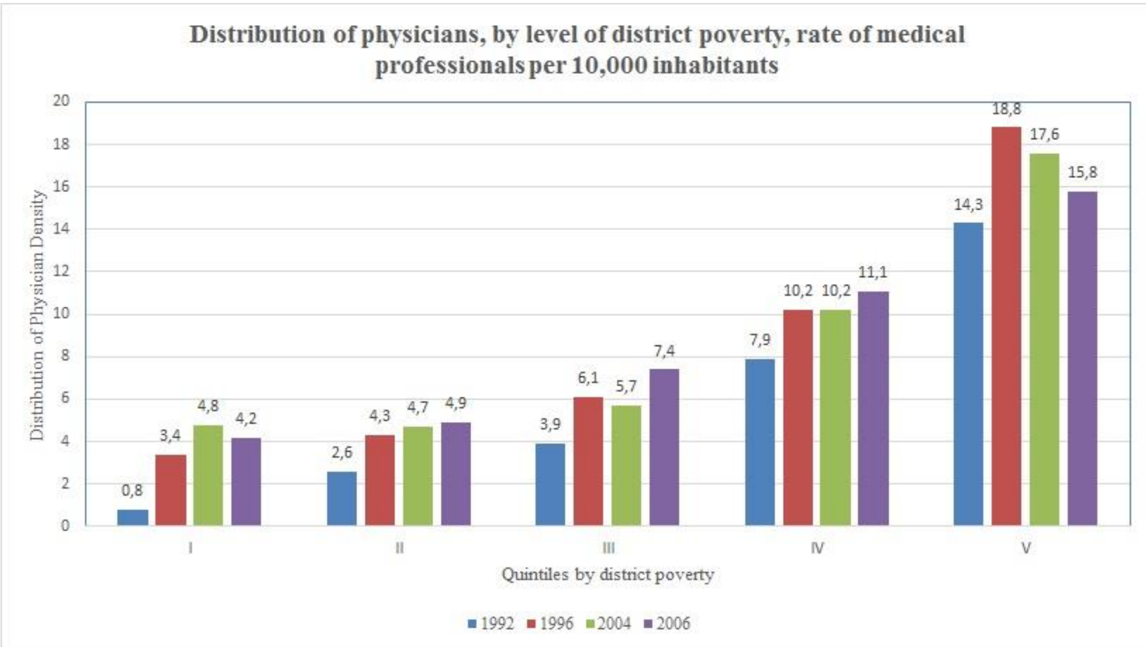

Figure 2. Distribution of physicians by level of district poverty, rate of medical professionals / 10,000 inhabitants, retrieved from World Bank (2011)

Besides the target of 25 health workers per 10,000 citizens, the HRO aims on achieving at least 40% of total medical workforce in physicians in primary care. As discussed, the density of health providers is associated with better health outcomes (Anand and Bärnighausen, 2004, Sundmacher and Busse, 2011). Peru attempts to reform the supply side, among others by enlarging the number of health workers. The number of physicians increased indeed between 2007 and 2009, but the amount of new doctors on permanent hire is rather low (World Bank, 2011).

5 Methodology

The following section explains the research design of the method that this study uses to identify what influences health care utilization. Paragraph 5.2 elaborates on the collection of the data for the research and defines the models. Paragraph 5.3 describes how to measure inequality in health care utilization and finally I explain about decomposing inequity.

5.1 Research Design

To measure the effect of health worker availability on health care utilization, I will use a binary regression model (BRM). This means I use a logistic regression method[1]. It provides a method to study how all explanatory variables affect the probability of an event occurring. The event in this case is whether a person used health care or not. The main reason to use a BRM is that the outcome variable is dichotomous. A person either used health care services or did not use this. Equation (1) below depicts a logistic regression.

$$y\,(x) = \frac{exp\,(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)}{1 + exp\,(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)} \qquad (1)$$

In this formula y(x) is bound to be between 0 and 1. An ordinary least squares (OLS) method is not restricted to estimations within this range, therefore predicting outcomes outside the area 0 to 1. This would give results that are actually not possible. In a logistic regression, y(x) is transformed to a logarithm, but it still contains the desirable properties of a linear regression model (Hosmer and Lemeshow, 2000).

A limitation of using a logistic regression model is that it requires a large sample size and a limited amount of missing data. For this research that will not be a problem as the sample contains 180,670 observations and the missing data is eliminated. However, another concern is that the magnitude of the change in the probability of the dependent variable depends on the levels of all the explanatory variables (Long and Freese, 2001). Hence to obtain the magnitude of the coefficient, the marginal effects need to be obtained. The estimation of a logistic regression provides only the sign and significance. A marginal effect provides the probability of an event occurring, although solely for specific values of explanatory variables. Figure 3, where 'd' represents a binary independent variable, visualizes this. For instance, Δ2

[1] Analysis of the log-likelihood of the logistic and probabilistic models will provide information for the decision of using either the logistic or probabilistic regression method.

is not equal to Δ3, while this would have been equal in a linear model. However as mentioned, in nonlinear models the magnitude of a change in a variable depends on the values of all variables in the model (Long and Freese, 2001). Therefore I need to obtain the *average* marginal effects, resulting in the average of the cumulated marginal effects. It provides a better insight in the general effect of provider density on the probability of health care utilization occurring. The interpretation on the other hand can be difficult to deduct (Williams, 2015).

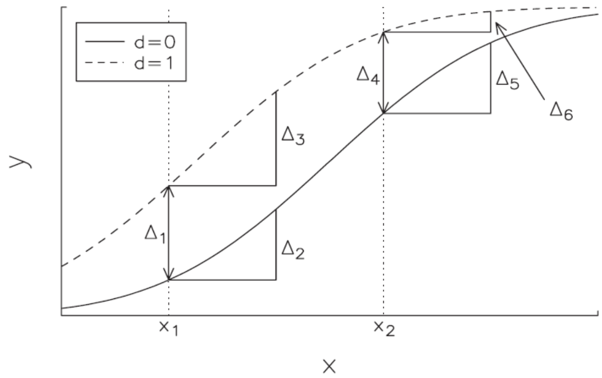

Figure 3. Visualization of marginal effects in a nonlinear model by Long and Freese (2001)

Causal relationships are hard to obtain due to several reasons of endogeneity. First, the opportunity of reverse causality exists, as already mentioned in the section about the health behavioral model. It is possible that for instance wealth has an effect on health care utilization, but health care utilization affects wealth as well through an improvement in health status and therefore productivity. Second, the potential presence of correlated omitted variables forms a threat. This means that a variable, possibly unobservable, is excluded which actually has a causal relationship with both provider density and health care utilization. It therefore confounds the results I obtain from the analysis in this study. An example can be a reform of health care coverage or a health shock, but also personal preferences. Health care coverage may separately cause an increase in both health provider supply and health care utilization. Although I will try to address these issues, the possibility of their presence cannot be eliminated entirely. Therefore, this study will not claim to establish a causal relationship, the results presented are merely associations.

5.2 Data Collection and Model Specification

I derive the data for this study from both the *Instituto Nacional de Estadistica e Informatica* (INEI), the Peruvian institute for statistics and from the website of the Ministry of Health in Peru (MINSA). The latter contains data on the health worker availability per administrative level. Peru has four administrative levels: national, departmental, provincial and district level. The most detailed level is per district. Two departments are eliminated because the administrative organization changed during the observed time period. The effect of these changes is that the number of districts per province cannot be matched in these departments, leading to unreliability of the data. Therefore this dataset contains 803 districts, compared to 158 provinces and 23 departments.

The health provider data is available on every administrative level. It is a panel data set, available for the years 2004, 2006, 2007, 2009, 2010 and 2012. In this research, only 2007, 2009 and 2010 are taken into account. Considering the implementation of the SIS in 2006 the demand for health care became more stable from 2007. The stable demand allows to conduct a better impact evaluation of the effect of a growing provider density. I apply the same reasoning in excluding 2012 data, since the implementation of the law on a guaranteed basic coverage package (AUS) is likely to increase demand for health care. The health workers are divided into several groups. Physicians and dentists cumulated represent the group 'doctors' and nurses and midwives form together the group 'nurses'. The reasoning behind this is that physicians and dentists perform more specialized care than nurses and midwives. The effect of a growth in density on the degree of health care utilization might therefore be different. Accumulated these two groups represent the 'skilled health providers'. For reasons of comparability it is better to present the density of health workers in a certain area, as opposed to the total number of health workers. I choose to measure the number health providers on province level. The departments are too large, with only 25 in the whole country it is plausible that many geographical and cultural difference exist within the departments. The districts are rather too disaggregated to measure, resulting in too few observations per district to create a reliable result. Therefore I measure health providers as the amount per 1000 persons available on the province level. This administrative level is detailed, but contains enough observations to derive conclusions from.

This study uses the data on micro household level from the dataset *Encuesta Nacional de Hogares*; National Household Survey (ENAHO) for the outcome and control variables. The ENAHO dataset exists of surveys conducted on households by the INEI. The total ENAHO sample contains 22,640 households, of which 13,824 situated in urban areas (61%). The population is randomly selected. The survey provides detailed information on the individual level. The survey population is distributed proportionally in twelve subsamples throughout the year, in order to guarantee a homogenous representation in each interference level (INEI, 2009).

The sample I use in this study consists of 180,670 observations in total for the targeted three years. Besides health related questions, the surveys contains information about predisposing and enabling factors that Andersen (1995) describes. A respondent is able to answer the questions in the survey through choosing between several closed options. This results in merely categorical variables.

An important limitation of these data is that the health surveys are self-reported. People might forget their health conditions or visits during the past time. Furthermore, perceptions of health can differ which may lead to biased results (O'Donnell et al., 1998).

5.2.1 Dependent Variables

The ENAHO dataset provides 15 different measures of health care utilization. All measures are derived from the multiple option question *'which health services did you receive?'*. I choose to use seven of the 15 binary variables to measure the outcome variable *health care utilization*. These variables are *USE01-USE05, USE07* and *USE13*. The variables excluded from the analysis are not relevant enough to observe, because these variables require a specific set of the population. For instance, labor and delivery services are only used by pregnant women. The remaining measures are health care services that theoretically can be used by the entire population in the sample. provides the description of the dependent variables. The variables are all unconditional on being ill, meaning they do not make a distinction between ill and healthy people using health care, as opposed to other measures the ENAHO survey contains. It is preferred to use the unconditional measures, because provider density appears to be correlated with the number of people reporting any illness. I choose to include as many dependent variables as feasible to ensure the generalizability of the results.

Table 1. Description of the dependent variables

Variable	Label	Coding
USE01	Utilization of GP consults	1 = if had a consult in the past 4 weeks 0 = if not had a consult in the past 4 weeks
USE03	Utilization of analysis	1 = if had an analysis in the past 4 weeks 0 = if not had an analysis in the past 4 weeks
USE04	Utilization of X-rays or tomography	1 = if had a scan in the past 4 weeks 0 = if not had a scan in the past 4 weeks
USE05	Utilization of tests, like hemodialysis	1 = if had a test in the past 4 weeks 0 = if not had a test in the past 4 weeks
USE07	Utilization of ophthalmology	1 = if used ophthalmology care in the past 3 months 0 = if not used ophthalmology care in the past 3 months
USE13	Utilization of inpatient care	1 = if hospitalized or had surgery in the past 12 months 0 = if not hospitalized or had surgery in the past 12 months
USE02	Utilization of medicines	1 = if used medicines in the past 4 weeks 0 = if not used medicines in the past 4 weeks

The first five measures, *USE01*, *USE03 – USE05* and *USE07* are variables containing outpatient health care services that require visiting a health professional. I observe the variables separately so I can analyze possible differences in effect of provider density for the different health services. *USE13* contains inpatient health care services and should be examined separately, because prior literature found a different relationship between provider density and utilization of inpatient and outpatient care respectively. For instance, Kravet et al. (2008) found that provider density is negatively correlated with inpatient health care and positively with GP consults (Kravet et al., 2008).

Furthermore I analyze utilization of medicines (*USE02*). The utilization is relatively large with 38.10% of the sample indicating to have used medicines. Medicine use can be related to provider density because receiving a prescription is conditional on seeing a health worker. Finally, I create two aggregated variables *USEophc* and *USEhc*. Table 2 visualizes the aggregated variables descriptions.

Table 2. Description of the aggregated measures USEophc and USEhc

Aggregated variable	Label	Coding
USEophc	Utilization of outpatient care services Includes USE01, USE03-05 & USE07	1 = if used outpatient care in the past 4 weeks or 3 months 0 = if not used outpatient care in the past 4 weeks or 3 months
USEhc	General measure of health care utilization Includes USE01, USE03-05, USE07 & USE13	1 = if used health care in the past 4 weeks or 3/12 months 0 = if not used health care in the past 4 weeks or 3/12 months

The study uses *USEhc* as an overall measure of health care to derive the effects of an increasing provider density on the probability of health care utilization in general. It is interesting to observe if any changes in the relationship occur for purposes of policy decision making. Furthermore, the utilization numbers for some measures are very low. From a policy perspective it can be more relevant to observe the results of health care utilization in general. I exclude *USE02* from this measure because medicine use may also be the result of buying over the counter medicines, or prescribed by a non-professional health worker not included in the dataset.

I generate *USEophc* to allow for an analysis of the differences between inpatient and outpatient health care utilization per region. I do not include this in the regression analysis because it does not contain additional value along the side of the other dependent variables.

It is possible for people to use more than one service, for instance having an X-ray and going to the doctor in the same period. However, I treat *USEhc* and *USEophc* merely as measures *if* health care is used, so no extra weights are put on people who use health care more than once. If otherwise, a multinomial logistic regression should be used, which is not feasible considering the difficult interpretation.

5.2.2 Explanatory Variables

Health provider density is the variable of interest in this research. The available data on health providers contains valuable information, since the numbers are available per province and occupation per health worker. Provider density is divided into a high skilled group *doctors*, consisting of physicians and dentists and a lower skilled group *nurses*. The latter contains nurses and midwives. The group *skilled* is the sum of doctors and nurses. The numbers are presented as health workers per 1000 province population, which improves the comparability. The method of analyzing health providers both aggregated (*skilled)* and disaggregated (*doctors* and *nurses)* is also performed by Anand & Bärnighausen (2004) and Kruk et al. (2009). The objective is to investigate whether there are differences in effects between the groups. E.g. I expect doctors to have a different, stronger, association with utilization of GP consults than nurses, as GP's are doctors. I construct two different models. The first contains skilled health workers as explanatory variable, while the second includes both doctors and nurses. If I would create separate models for doctors and nurses as well it is likely that the results will be biased. Doctors and nurses are highly correlated (0.6810 based on a Pearson correlation). Furthermore they both affect health care utilization and it is plausible that an increase in doctors is related to a change in the amount of nurses and vice versa.

5.2.3 Control Variables

Based on the theoretical framework of Andersen (1995) and other literature, I include a set of control variables. Excluding these variables will cause a bias in the results, as it is obvious that health provider density is not the only factor influencing health care utilization. In a model solely incorporating the explanatory variable one omits the context of the reality, hence deriving a conclusion based on insufficient information. The purpose of the these variables is therefore to control for factors that can confound the results.

In terms of demographic variables *Age-Sex* groups, *marital status* and an *urban* dummy are included to the model. The categorical variable 'AgeSex' contains male and female groups of different age categories. The variable improves the measures of fit, which is quite logical as people from different age groups and different genders have another demand for health. This variable is statistically preferred in comparison to the continuous variable of age or a gender dummy. Marital status is a factor which might influence utilization of health care through demand for health care. These are predisposing factors that influence demand for health (Andersen, 1995). A dummy for urbanization is included inter alia because differences in the use of maternal health care services were found between rural and urban areas (Elo, 1992).

Socio-economic variables I include are *educational level, a log of monthly non-health consumption per capita* and *sector of occupation.* Previous research indicates that maternal schooling was significantly and positively associated with a higher use of maternal health care services for women in Peru (Elo, 1992), which supports inclusion of education. Wealth has an influence on utilization of health care as well, as a larger budget allows for more health expenditures (Kruk et al., 2009). In the model monthly non-health consumption per capita is added instead of monthly income per capita as it appears that this improves the predictive power of the model significantly. I take the logarithm of monthly non-health consumption to correct for skewedness as large differences exist in consumption per capita. The average monthly consumption is 233.41 Peruvian Sol (PEN) with a standard deviation of 179.56 PEN supports this statement. Sector of occupation is included in the model because I expect that need for health care changes per sector that a person works in. For instance, agricultural work is heavier than working in the service industry.

To address the effect of the health system as Andersen (1995) describes, other than provider density, I control for *health insurance, CRECER, health status* and *insurance*illness*, which is an interaction between health insurance and health status. Health insurance is obtained from the ENAHO dataset as whether someone is enrolled in SIS or not. People with any insurance type other than SIS are excluded from the dataset. CRECER is a nutrition program implemented by the government (World Bank, 2011). 56,536 of the 180,670 observations live in a district where CRECER is implemented. Such a program influences the need for health through health status and therefore utilization of health care. Hence, not including this can bias the results. *Health status* is based on the question whether a person was confronted with any illness during the past 4 weeks. Health insurance and health status are important contributors of health care utilization (Elbich and Ziebarth, 2013, Grumbach et al., 1997). The interaction variable *insurance*illness* observes the extra effect of having a health insurance and being ill on health utilization. All these factors are likely to influence both demand for health care and health worker availability.

In the second model examining the effect of *doctors* and *nurses* instead of *skilled* provider density, I include several additional variables. The reason is that doctors and nurses provide different services, therefore other factors can have an effect on the increase in their density. Doctors tend to perform more specialized care which is more expensive. Therefore I include extra variables that control for development and poverty. Specifically I include a categorical variable for *job position* and a dummy that measures whether the non-health consumption is *below the poverty line*. Job position provides a reflection of intellectual development. Someone working as a professional (category two) performs more difficult work than someone working as a clerk (category three), which will also result in a higher salary. The dummy controlling for consumption below the poverty line is a measure of extreme poverty. Both variables improve the model statistically.

Finally, *year and regional dummies* are included to control for fixed effects. The provinces are taken into account, inter alia because health status, degree of utilization and provider density differ between regions. Furthermore the landscape in Peru has very different structures with a part of coast, the Andes mountain range and an Amazon area. The model contains a set of dummy variables of provinces to control for region invariant effects. In this way the structural differences between regions, e.g. geographical structure, cannot confound the effects of provider density on health care utilization. Excluding them may lead to biased results. For instance, a rural region surrounded by mountains as geographical structure probably needs more health providers in order to change the utilization rate than a rural region with a flat landscape and many cities. Time is included since over the years the error term of the model should be uncorrelated with the explanatory variables. The year dummies can control for unobserved characteristics that affect all individuals in a given year. Finally, I use robust standard errors to prevent any possible heteroscedasticity.

5.2.4 Model Specification

The logistic regression model that I use to identify the effect of skilled provider density on several outcomes of health care utilization is depicted in equation two:

$$\Pr(y=1|x) = \Phi\,(\beta_0 + \beta_1{*}\textit{skilled} + \beta_2{*}\ \textit{i.age*sex}_i + \beta_3{*}\textit{i.education}_i + \beta_4{*}\ln(\textit{monthly non-health consumption}_i) + \beta_5{*}\textit{i.marital status}_i + \beta_6{*}\textit{health insurance}_i + \beta_7{*}\textit{i.occupational sector}_i + \beta_8{*}\textit{i.urban}_i + \beta_9{*}\textit{i.any illness}_i + \beta_{10}{*}\textit{i.insurance*illness}_i + \beta_{11}{*}\textit{CRECER}_i + \beta_{12}{*}\textit{i.year} + \beta_{13}{*}\textit{i.province}_i + \varepsilon,\ \textit{robust}) \qquad (2)$$

Model three controls for possible separate effects between doctors and nurses. It contains more control variables and is shown below:

$$\Pr(y=1|x) = \Phi\,(\beta_0 + \beta_1{*}\textit{doctors}_\mathrm{i} + \beta_1{*}\textit{nurses}_\mathrm{i} + \beta_3{*}\ \textit{i.age*sex}_i + \beta_4{*}\textit{i.education}_i + \beta_5{*}\ln(\textit{monthly non-health consumption}_i) + \beta_6{*}\textit{i.poverty line}_i + \beta_7{*}\textit{i.marital status}_i + \beta_8{*}\textit{health insurance}_i + \beta_9{*}\textit{i.occupational sector}_i + \beta_{10}{*}\textit{i.job position}_i + \beta_{11}{*}\textit{i.urban}_i + \beta_{12}{*}\textit{i.any illness}_i + \beta_{13}{*}\textit{i.insurance*illness}_i + \beta_{14}{*}\textit{CRECER}_i + \beta_{15}{*}\textit{i.year} + \beta_{16}{*}\textit{i.province}_i + \varepsilon,\ \textit{robust}) \qquad (3)$$

These models have the highest likelihood ratios and measures of fit[2]. These are shown in table 13 in the Appendix.

I examine eight outcome measures of health care utilization: utilization of GP consults (*USE01*), utilization of analysis (*USE03*), utilization of scans (*USE04*), utilization of other tests (*USE05*), utilization of ophthalmology (*USE07*), utilization of inpatient care services (*USE13)*, utilization of health care in general (*USEhc*) and utilization of medicines (*USE02*). For each of these outcomes I estimate the model separately. However, I also use two different groups of explanatory variables: skilled health workers and doctors & nurses. This means that I estimate 16 different models.

---

[2] Fitstat measures indicated that a logistic regression is more suitable than a probabilistic regression for all measures of utilization of health care.

5.3 Measuring Inequality

A concentration curve provides an overview of the inequality in the distribution of provider density against health care utilization. As the concentration curve is a rank-based measure, it can also be used to rank provider density for a population (Van der Poel, O'Donnell and van Doorslaer, 2008). The x-axis is broken into deciles, ranked from the persons exposed to the lowest provider density in their region up to the highest provider density per decile. The y-axis shows the cumulative health care utilization measured as percentage of the total. In other words, the concentration curve measures the amount of cumulative health care utilization by the poorest *'x'* percent in terms of provider density. If this distribution would be completely equal, the concentration curve would lie on the 45-degree line: every person uses the same amount of health care, regardless the amount of health providers available. If however, the concentration curve lies below (above) the 45-degree curve, people with low access to health providers use less (more) health care than people living in a region with high provider density (O'Donnell et al., 2008). Hence, the curve designates value to the people with poor access when it lies above the line of equality.

From the concentration curve one can derive the concentration index. It is calculated as twice the area between the concentration curve and the 45-degree line. If the concentration curve lies below (above) the line of equality, the value of the concentration index is positive (negative) (O'Donnell, 2008). I explain the exact mathematical form in the next section. Although utilization of health care is a binary variable, it can be used to draw a concentration curve since it satisfies the principle of transfers, even though that is bound to 1 (Wagstaff, 2011). The principle of transfers implies for this study that if health providers are transferred from a high density region to a low density region, but keeping the order of the rank, inequality decreases.

5.4 Decomposing Inequity

As dependent variable in the decomposition I will use the measures of health care utilization for which provider density is significant. The aim is to analyze the inequity caused by provider density. If provider density appears to not have an influence on health care utilization, the inequity analysis will not provide useful information.

To measure inequity in need-standardized health care utilization, it is preferred to use a generic outcome measure of health care utilization. This gives a broader overview of the inequality in a country, contrary to a specific measure like use of institutional births (O'Donnell et al., 2008).

The decomposition of equity is performed by an Ordinary Least Squares (OLS) regression method. Since the outcome variable is binary, it might give problems considering the bounds of dichotomous variables. Binary outcomes are best estimated by a nonlinear method. However, according to van der Poel et al. (2008), usage of OLS gives a more intuitive interpretation and decomposition methods for binary outcomes all have their own limitations. Therefore OLS is preferred (Van der Poel et al., 2008).

Decomposition of equity requires standardizing health care utilization by need. The aim is to find out what the inequality would be if utilization caused by 'justifiable' need factors is eliminated. Subsequently I measure whether and to what extent the inequality in health care utilization by provider density is driven by socio-economic factors. There exist two ways of standardizing the need by use: the indirect and direct method. The indirect method is more feasible on the individual level and is therefore chosen (O'Donnell et al., 2008). Equation four visualizes how to calculate the concentration index.

$$y_i = \alpha + \sum \beta_j x_{ji} + \sum \gamma_k z_{ki} + \varepsilon_i \qquad (4)$$

In this equation, y_i is the health care utilization, and i denotes the individual; α, β and γ are vector parameters, each indicating different characteristics of the individual. The x_j variables are the variables that indicate need by age, gender and having an illness, which needs to be standardized for. The z_k variables are the confounders which are of particular interest in the research for inequity. To standardize utilization by need, $\hat{y}_i^{IS}$ in equation five should be obtained:

$$\hat{y}_i^{IS} = y_i - \hat{y}_i^X + \bar{y} \qquad (5)$$

$\hat{y}_i^{IS}$ denotes the need-standardized health care utilization, $\hat{y}_i^X$ is the value estimated by OLS and $\bar{y}$ is the overall sample mean of health care utilization. The outcome $\hat{y}_i^{IS}$ can be interpreted as the distribution of health care utilization that would be expected to be observed, regardless of differences in the distribution of the x's across provider densities (O'Donnell et al., 2008).

Subsequently, an OLS regression is performed using ADePT software on the need-standardized health care utilization. I examine to what extent the inequality in health care utilization driven by provider densities, is caused by socioeconomic factors. The same variables are used as in the logistic regression, which leads to equation six.

$$y_i = \alpha + \Sigma\beta_j[i.AgeSex_{ij} + i.any\ illness_{ij}] + \Sigma\gamma_k\,[i.provider\ density_{ik} + i.education_{ik} + \ln(monthly\ non\text{-}health\ consumption_{ik}) + i.poverty\ line_{ik} + i.marital\ status_{ik} + health\ insurance_{ik} + i.occupational\ sector_{ik} + i.job\ position_{ik} + i.urban_{ik} + \beta_9{*}i.any\ illness_{ik} + CRECER_{ik} + i.year_k + i.province_{ik}] + \varepsilon_i \quad (6)$$

6 Results

The first section provides the descriptive statistics of the data sample, followed by the results of the linear probability model estimations. Paragraph 6.3 presents the decomposition of inequity. Furthermore, an additional analysis is performed on the results of the logistic regressions in the last paragraph.

6.1 Descriptive Statistics

The total sample consists of 180,670 observations measured during the years 2007, 2009 and 2010. The number of observations per year shows a slight decrease, but is in general very similar. Table 3 provides the descriptive statistics of the eight types of dependent variables. During the years the percentage of people using health care increased for all categories. Utilization of tests has the largest increase with 55.5% from 2007 to 2010, followed by utilization of GP consults with 31.8%. The amount of observations for utilization of tests is however very small with only 98 observations in 2010.
Utilization of GP consults, health care overall and medicines are the variables with the most observations.

Table 4 shows the descriptive statistics of the health providers. The number of health workers increased each year. The minimum amount of health workers is 0 in 2007 for every group, but grows with the years, with an increase of 28.1% from 2007 to 2010 in the mean on skilled health workers per 1000 province population. Skilled health worker availability is positively correlated with the variables utilization of a GP (*USE01*), use of scans (*USE04*), ophthalmology (*USE07*), health care overall (*USEhc*) and inpatient health care (*USE13*) On the other hand, skilled provider density is negatively correlated with utilization of analysis (*USE03*), use of tests (*USE05*) and medicines (*USE02*). The density of health providers varies heavily per province. Figure 4 captures this variation.

Table 3. Descriptive statistics of the dependent variables

	Utilization of GP consults			Utilization of analysis			Utilization of scans, like X-rays			Utilization of tests, like hemodialysis		
Year	*2007*	*2009*	*2010*	*2007*	*2009*	*2010*	*2007*	*2009*	*2010*	*2007*	*2009*	*2010*
Observations	62,786	59,772	58,162	62,786	59,772	58,162	62,786	59,772	58,162	62,786	59,772	58,162
Used health Care (%)	20.53	25.06	27.06	2.28	2.86	2.79	0.84	1.10	1.11	0.10	0.13	0.17
Used health Care	12,892	14,964	15,737	1,429	1,710	1,622	529	569	647	63	76	98
Mean	0.21	0.25	0.27	0.02	0.03	0.03	0.01	0.01	0.01	0.001	0.001	0.001
Standard Deviation	0.40	0.43	0.44	0.15	0.17	0.16	0.09	0.10	0.10	0.03	0.04	0.04
	Utilization of ophthalmology			**Utilization of inpatient care**			**Utilization of health care overall**			**Utilization of medicines**		
Year	*2007*	*2009*	*2010*	*2007*	*2009*	*2010*	*2007*	*2009*	*2010*	*2007*	*2009*	*2010*
Observations	62,786	59,772	58,162	62,786	59,772	58,162	62,786	59,772	58,162	62,786	59,772	58,162
Used health Care (%)	1.57	1.82	1.90	3.97	4.43	4.68	24.37	29.08	31.08	35.98	37.28	41.03
Used health Care	984	1,089	1,104	2,495	2,645	2,720	15,304	17,367	18,077	22,590	22,265	23,866
Mean	0.02	0.02	0.02	0.04	0.04	0.05	0.24	0.29	0.31	0.36	0.37	0.41
Standard Deviation	0.12	0.13	0.14	0.20	0.21	0.21	0.43	0.45	0.46	0.48	0.48	0.49

Table 4. Descriptive statistics of the explanatory variables

	Doctors & Dentists/per 1000 province population			Nurses & Midwives/per 1000 province population			Skilled health workers/per 1000 province population		
Year	*2007*	*2009*	*2010*	*2007*	*2009*	*2010*	*2007*	*2009*	*2010*
Observations	62,786	59,772	58,162	62,786	59,772	58,162	62,786	59,772	58,162
Mean	0.498	0.564	0.565	0.734	0.888	1.021	1.237	1.452	1.585
Standard Deviation	0.27	0.25	0.25	0.41	0.44	0.49	0.63	0.64	0.67
Minimum	0	0.127	0.098	0	0.195	0.270	0	0.406	0.426
Maximum	1.856	2.095	1.792	3.009	3.652	3.312	4.640	5.143	4.781

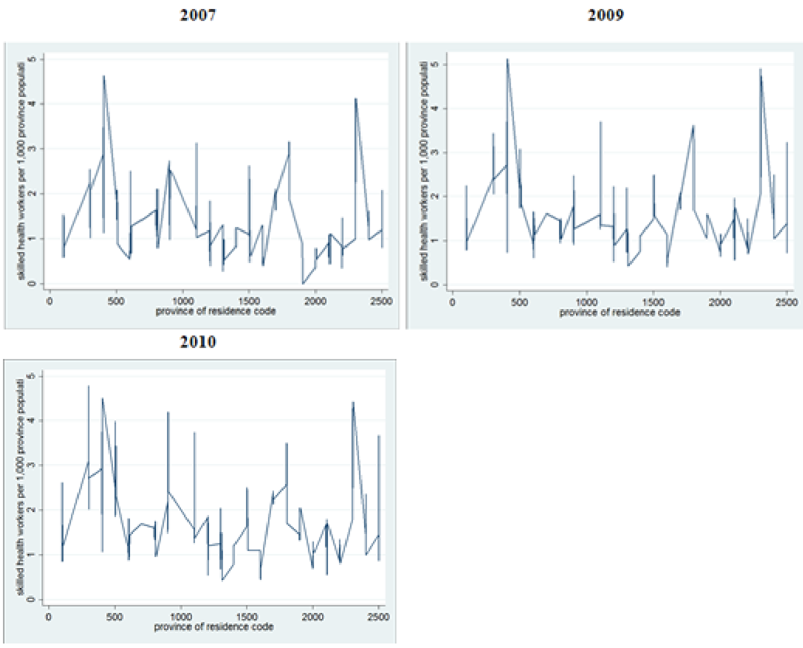

Figure 4. Variations of skilled health workers per province in 2007. Source: own elaboration

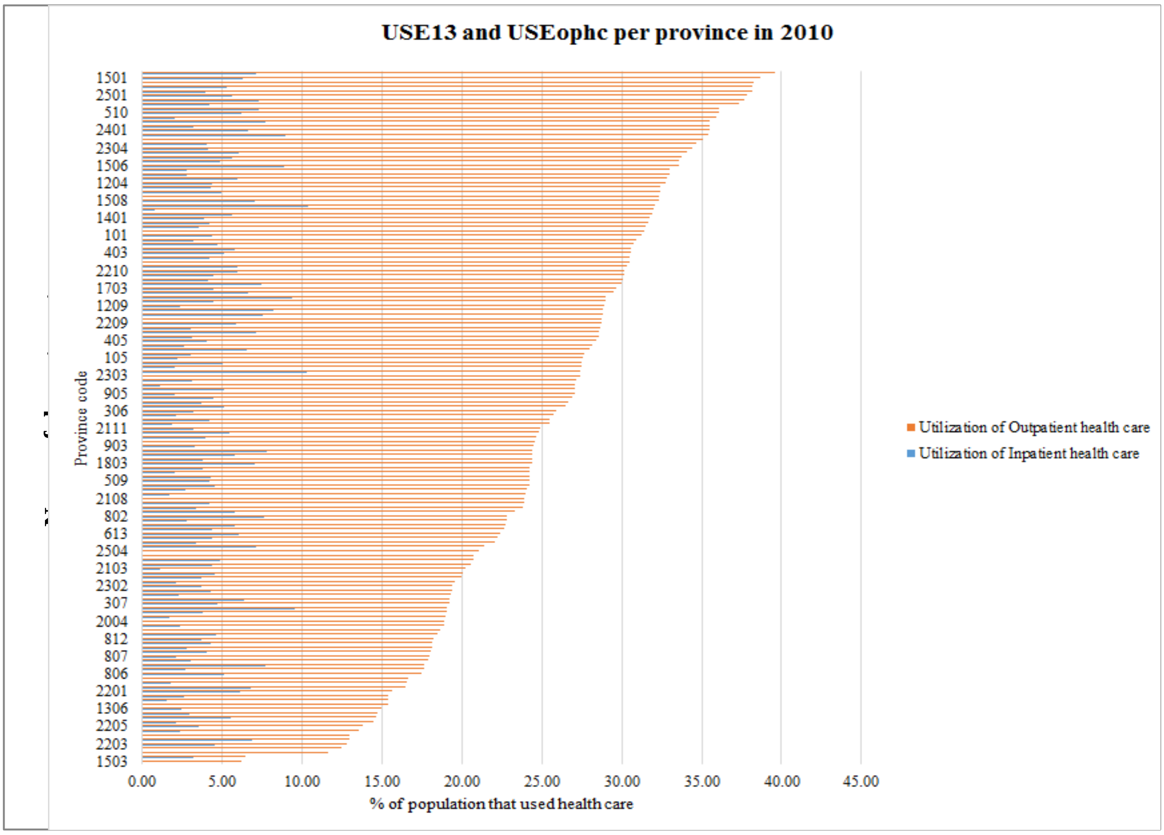

Figure 5. Utilization of inpatient and outpatient health care per province in 2010. Source: own elaboration

Eibich and Ziebarth (2013) found a substitute relationship between outpatient and inpatient care in regions. This implies that outpatient care consumption was relatively high in regions with low levels of inpatient care and vice versa. As indicated I create the variable *USEophc* to analyze the relationship between inpatient and general health care per region in Peru. However I cannot conclude a clear opposite trend between the two types of care per department. Figure 5 in visualizes the amount of inpatient and outpatient care per province in 2010. In provinces where utilization of outpatient care is relatively high, inpatient care is not lower than in provinces where outpatient care is relatively low.

Table 11 in the appendix provides the descriptive statistics of the control variables. Almost 67% of the sample is below an age of 35 years, this is in line with demographic numbers of the entire population in 2007 (World Bank, 2011). Males and females are almost equally proportioned and the majority of the sample is married. Only 16% of the respondents indicates not to have any legal partnership. It appears that the Peruvians in the dataset are quite educated, as almost 60% has finished at least secondary education, although approximately 10% of the adults has not finished any education at all. Concerning the occupations of the sample, more than 45% does not work. This might be explained by the fact that 50% of the sample is younger than 23 years old. Therefore probably many of the observations still participate in educational activities instead of being employed. Agriculture (25%) and services (22%) are the dominant occupational sectors for the sample. One third of the sample is living in a region where the CRECER program is implemented and almost 46% of the population in the sample is enrolled in the SIS program.

6.2 Regression results

I first perform a univariate analysis. This is primarily to obtain whether the relationship between provider density and health care utilization is existent at all and to detect a possible change when including control variables. For instance, Grumbach et al. (1997) found an effect of provider density, only when *not* controlling for other factors. The socioeconomic and individual characteristics like health insurance and wealth correlate with provider density, explaining that outcome. I controlled for multi-collinearity in the sample, but there are no critical correlations between the measures of provider density and the other variables. Table 10 containing Pearson correlation in the appendix shows this. Table 5 provides the results of the logistic regressions without the control variables. It shows the estimated coefficients, the robust standard error (SE) and the significance level for each variable. Note that the variables on the left are the different dependent variables and that I estimated a model containing skilled and a model containing doctors and nurses.

Table 5. Results of the univariate logistic regressions – average marginal effects

Univariate models	Skilled		Doctors		Nurses	
	Coefficient	*SE*	*Coefficient*	*SE*	*Coefficient*	*SE*
Utilization of GP consults	0.016***	0.001	0.089***	0.005	-0.021***	0.003
Utilization of analysis	-0.001***	0.000	0.032***	0.002	-0.020***	0.001
Utilization of scans	0.001***	0.000	0.019***	0.001	-0.008***	0.001
Utilization of other tests	-0.000	0.000	0.004***	0.000	-0.003***	0.000
Utilization of ophthalmology	0.003***	0.000	0.034***	0.002	-0.014***	0.001
Utilization of inpatient care	0.001***	0.001	0.034***	0.002	-0.016***	0.002
Utilization of overall health care	0.017***	0.002	0.133***	0.005	-0.043***	0.003
Utilization of medicines	-0.029***	0.002	0.010***	0.006	-0.096***	0.004

N=180,670 for all models. *** coefficient is significant at α<=0.001 ** coefficient is significant at α<=0.005 * coefficient is significant at α<=0.10. Standard errors are robust standard errors.

The coefficients appear to be highly significant, although the signs differ. The coefficients for doctors and nurses show a strong contrasting result, the coefficients for doctors are all positive and all negative for nurses. The coefficients for skilled seem to reflect the average effect of doctors and nurses. Is implies that doctors have a positive relation with the probability of using all types of health care, while nurses have a negative association with the likelihood of health care utilization. The negative coefficients for nurses are surprising, but the multivariate analysis should be analyzed first before deriving conclusions.

Table 6. Results of the multivariate logistic regressions – average marginal effects

Multivariate models	Skilled		Doctors		Nurses	
	Coefficient	*SE*	*Coefficient*	*SE*	*Coefficient*	*SE*
Utilization of GP consults	0.007*	0.004	0.029**	0.013	-0.004	0.007
Utilization of analysis	0.002	0.002	0.015**	0.006	-0.004	0.004
Utilization of scans	0.001	0.001	-0.001	0.004	0.002	0.002
Utilization of other tests	-0.001	0.001	0.005*	0.003	-0.004**	0.002
Utilization of ophthalmology	0.003*	0.002	0.003	0.005	0.003	0.003
Utilization of inpatient care	-0.002	0.002	0.001	0.007	-0.003	0.004
Utilization of overall health care	0.005	0.004	0.030**	0.013	-0.008	0.008
Utilization of medicines	-0.004	0.003	-0.003	0.011	-0.005	0.006

*** coefficient is significant at α<=0.001 ** coefficient is significant at α<=0.005 * coefficient is significant at α<=0.10. Standard errors are robust standard errors.

Table 6 delivers the results of the multivariate models, which are more interesting to observe as the likelihood of omitted variable bias is largely reduced. In this section I only present the results of the variable of interest; provider density. The results for the control variables can be found in the appendix in tables 12 and 13.

These average marginal effects are not as significant results as the univariate models. The coefficient of the skilled provider density reflects an average effect for both doctors and nurses. Only two coefficient are significant, for the dependent variables utilization of GP consults and utilization of ophthalmology. It can be interpreted that when the number of skilled health provider per 1000 persons increases with one, on average the probability of consulting a GP increases with 0.7 percentage points (p.p.), ceteris paribus. To put this in perspective, the share of people that indicated to having consulted a GP in 2010 was 27.06%. Doctor density shows on average more significant results than nurse density. As expected the relation between doctor density and the probability of health care utilization is positive for all measures of health care utilization. The association is significant for utilization of GP consults, analysis, other tests and health care in general. For health care in general the coefficient implies that when the number doctors per 1000 persons increases with one, on average the probability of using health care in general increases with 3 p.p., ceteris paribus. However, it should be noted that the effect for utilization of GP consults mainly determines this magnitude, as the coefficient of GP consults forms the largest part in the health care overall services and has a value of 0.029. In the relation between nurse density and health care utilization only one association is significant, which is with utilization of other tests, like hemodialysis. It implies that when the number of nurses per 1000 persons increases with one, the utilization of other tests decreases with 0.4 p.p., ceteris paribus. All other results are insignificant, hence including control variables is substantive in obtaining more valid results. The results of the control variables are provided in tables 12 and 13 in the appendix.

6.3 Results on Inequity

Considering the results of the regression analysis I use doctor density as indicator of living standards measure. I do include nurse density in the model as well to prevent omitted variable bias. Furthermore I use utilization of GP consults and utilization (*USE01*) of health care in general (*USEhc*) as dependent variables. The reason is that the results are particularly more significant and stronger for doctor density. The magnitude of the coefficients in the regressions on *USE01* and *USEhc* is larger, probably because the number of people that used these services is relatively large. For the inequality and decomposition analysis the relations between these variables are therefore the most interesting to observe. Based on the regression results in the previous section, policy makers should enlarge supply of doctors if their aim is to improve health care utilization. The inequity analysis provides information on how the additional doctor workforce should be distributed to enhance or maintain equality in health care utilization.

The Y-axis of the concentration curves depicted in Figure 6 measures the cumulative percentage of the utilization variables. The X-axis measures the cumulative percentage of population, ranked from poorest to richest. In this case, poor means the lowest amount of skilled health providers per 1000 people. Table 7 shows the correlation of the different quintiles of provider density distribution with the different categories of health care utilization. The number of 0.2088 in the lowest quintile for utilization of GP consults care indicates the following: in the quintile with the lowest provider density, 20.88% of the sample population consulted a GP. As explained, the concentration index (CI) shows the inequality in the distribution of utilization. For *USE01* the CI is 0.059, therefore the distribution is biased towards the rich. This means that on average people living in a province with low provider density, use less outpatient health care than people living in a province with high provider density. The results in the table show that the difference is increasing in each quintile. The CI for utilization of health care in general is more pro-rich with a value of 0.062. In the quintile with the highest provider density on average 31.76% used health care, compared to 23.96 in the quintile with the lowest provider density.

On the one hand, the values of the CI's are very close to zero, which means total equality. However, the numbers per quintile show an increasing trend indicating that low doctor density is truly related to less utilization. The decomposition of equity will reveal the underlying factors causing the inequality.

Figure 6. Concentration Curves for the health care utilization measures USE01 and USEhc

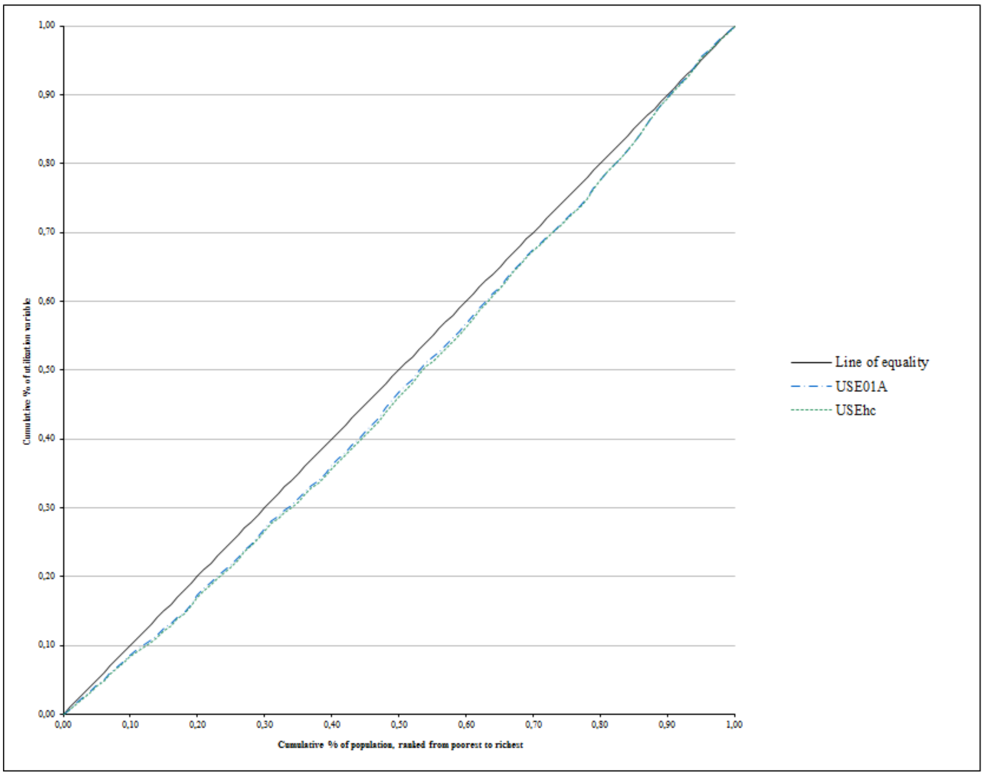

Table 7. Inequality in health care utilization

Inequality	Utilization of GP consults (USE01)	Utilization of care overall (USEhc)
Lowest quintile	0.2088	0.2396
2	0.2207	0.2567
3	0.2485	0.2877
4	0.2547	0.3028
Highest quintile	0.2737	0.3176
Total	0.2413	0.2809
Standard CI	0.0590	0.0619

Figure 7 and table 8Table 7 demonstrate the decomposition of inequity for the four measures of health care utilization. It captures the different parts that lead to the values of the CI's that are shown above. Furthermore, it makes a distinction between the justifiable and unjustifiable contribution to the inequality in health care utilization. The most important contributor to the unjustifiable inequality is doctor density, which has a positive value for both measures. It means that the density of doctors in a region is a factor contributing to an unequal utilization of both GP consults and health care overall. For utilization of *USEhc* the CI is 0.0619, and doctor density contributes to this CI with 0.038 which is more than 50%. The positive sign implies a pro-rich inequality impact, therefore it can be concluded that high doctor density favors people living in regions with a high provider density to utilize more health care services. Furthermore, wealth is a large contributor to the pro-rich inequality. Nurse density on the other hand seems to be biasing towards the poor. This means that for people living in regions with low doctor provider density, the number of nurses per person is related to more equal utilization of health care. Although the absolute numbers are not very substantial, the unjustified inequality represents on average a large share in the CI's. Doctor density and wealth are the most important contributors to this inequity. However, in general, the inequality is quite low and although differences per quintile of provider density exists, there are no extreme values of inequality.

Figure 7. Graphical decomposition of inequity for the health care utilization measures

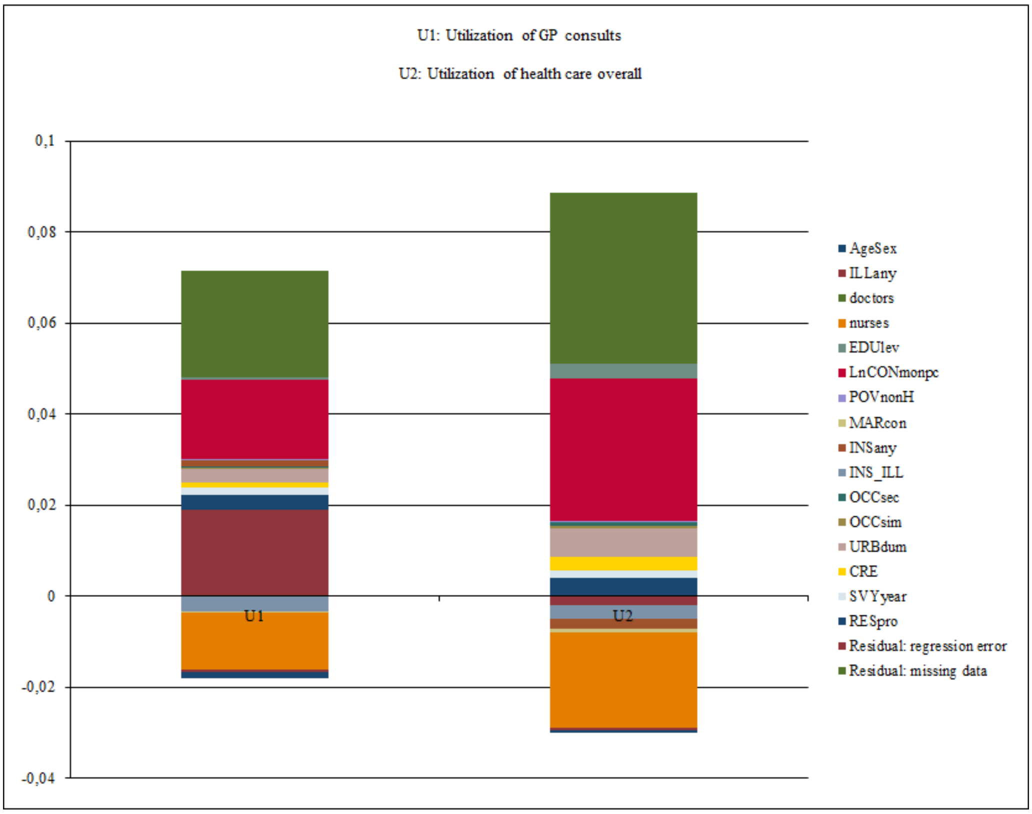

Table 8. Decomposition of the CI for health care utilization

	Utilization of GP consults (USE01)	Utilization of health care overall (USEhc)
Standardizing (need) variables		
Age-gender group	-0.001	-0.001
Any illness	-0.001	0.000
Subtotal	**-0.002**	**-0.001**
Control variables		
Doctor density	0.023	0.038
Nurse density	-0.013	-0.021
Education level	0.001	0.003
Monthly non-health cons.	0.017	0.031
Poverty line	0.000	0.000
Marital status	0.000	-0.001
Health insurance	0.001	-0.002
Insurance and ill	-0.003	-0.003
Occupational sector	0.000	0.001
Job position	0.000	0.001
Urban area	0.003	0.006
CRECER	0.001	0.003
Province	0.003	0.004
Year	0.002	0.004
Subtotal	**0.036**	**0.062**
Residual: regression error	0.019	-0.002
Residual: missing data	0.000	0.000
Inequality (total)	0.053	0.058
Inequity/Unjustified inequality	0.055	0.059

6.4 Additional Analysis

The average marginal effects in a BRM are not always reliable when using a continuous explanatory variable. Because the relationship between health provider density and health care utilization is nonlinear, it is not guaranteed that the coefficients are providing the exact correlation. This largely depends on how the explanatory variable is scaled. If the variable is scaled in small units, the chance is more likely that the average marginal effect is correct (Williams, 2015). Although the scale of skilled health workers consists rather of small units, I still control the robustness of the results in the logistic regressions. I do this using an Ordinary Least Squares (OLS) regression. The interpretation is straightforward, unlike the interpretation of the average marginal effects of the logistic regressions. A change of one unit in health provider density will reflect the direct change in health care utilization.

The variables included in the models are identical to the ones used in the logistic regressions. First, this allows for a better comparison and second, it proved to have the best possible explanatory power. The OLS results are provided in table 9.

Table 9. Results for multivariate OLS regressions

OLS regressions	Skilled		Doctors		Nurses	
	Coefficient	*SE*	*Coefficient*	*SE*	*Coefficient*	*SE*
Utilization of GP consults	0.008**	0.003	0.050***	0.012	-0.014**	0.007
Utilization of analysis	0.001	0.001	0.018***	0.003	-0.008***	0.003
Utilization of scans	0.001	0.001	0.004	0.003	-0.001	0.002
Utilization of other tests	-0.000	0.000	0.003***	0.001	-0.002***	0.001
Utilization of ophthalmology	0.002**	0.001	0.008**	0.004	-0.002	0.002
Utilization of inpatient care	-0.002	0.002	0.003	0.006	-0.005	0.004
Utilization of overall health care	0.006*	0.004	0.050***	0.012	-0.017**	0.007
Utilization of medicines	-0.005	0.004	0.012	0.012	-0.015**	0.007

N=180,670 for all models . *** coefficient is significant at α<=0.001 ** coefficient is significant at α<=0.005 * coefficient is significant at α<=0.10. Standard errors are robust standard errors.

The difference between the results obtained in the logistics and OLS results mainly relate to the significance of the coefficients. The sign and magnitude of the OLS results are quite similar to the ones of the logistic regressions, although OLS shows somewhat stronger magnitudes than the BRM. Furthermore, the relation between all three provider groups and utilization of health care in general are highly significant using OLS, while in the logistic regressions this was only the case for doctor density. The results for utilization of GP consults, analysis, others tests, ophthalmology and medicines are also more significant for at least one group of providers. The interpretation differs as well. For example, if skilled provider density increases with one unit, utilization of GP consults would increase with 0.8%, ceteris paribus. I.e., the interpretation is stronger for OLS than for the logistic regressions.

However, the outcomes of OLS for a binary outcome variable should be treated carefully as well. Health care utilization in this research is binary, either someone used health care or not. As a consequence, the minimum value of health care utilization is 0 and the maximum value is 1. OLS does not take this limitation into account and predicts values outside this range, which is actually impossible and hence undermines the reliability. This is the main reason for choosing the BRM as principle method for this research.

7 Conclusions and Discussion

This study had as objectives to analyze the relationship between health provider density and health care utilization and the extent of the attributing inequity, for the uninsured and publicly insured population in Peru. The next sections summarize the main findings and discuss how the results fit into the existing literature. Subsequently this final chapter provides a discussion on the validity of the results, the policy implications and suggestions for future research.

7.1 Summary of the Main Findings

This research examined the association between health provider density and several measures of health care utilization, and to what extent this relationship incorporates unjustifiable inequality. Specifically, the effect of skilled health workers per 1000 province population on utilization of GP consults, analysis, scans, other tests, ophthalmology, inpatient health care, health care in general and medicines is investigated. Also, the group of skilled health workers is divided into doctors and nurses to examine whether specific differences exist between categories of health providers. Concerning the inequity decomposition, I analyzed what the inequality in overall health care utilization and utilization of GP consults is due to a different distribution of physicians. Subsequently I investigated the underlying factors explaining this inequality.

The results of the univariate analysis of provider density on health care utilization show highly significant coefficients for all measures. I expected the correlation of all health provider groups to be positive for all measures of health care utilization, but the correlations for nurse densities are all negative. Therefore the coefficients of skilled health providers have ambiguous signs. The logistic regression results of the multivariate analysis indicated that skilled health provider density has a positive and significant relationship with the probability of using GP consults and health care overall. In observing doctors and nurses in the second model I find that doctor density is positively and significantly correlated with the probability of using GP consults, analysis, other tests and health care in general. Nurse density is only negatively and significantly related to utilization of other tests, like hemodialysis. The results of the multivariate analysis indicate that control variables change the outcomes and are therefore of utmost importance. The additional analysis using OLS confirmed the results by showing similar outcomes.

The inequality analysis was only conducted using doctor density as explanatory variable and utilization of GP consults and health care overall as dependent variables. The strong and significant results in the logistic regressions are the reason. The concentration curves show an inequality biased towards people living in regions with high provider density, and therefore has a pro-rich bias. The decomposition analysis learns that doctor density and wealth are important contributors to this unjustifiable inequality for both measures of health care utilization. Nurse density seems to have a pro-poor contribution to inequity.

7.2 Discussion of the Results

In the existing literature some studies make a distinction between general practitioners and specialist physicians by including GP density and SP density separately in one model (e.g. Busato and Künzi, 2008). Other studies use total provider density and density of GP's as proxies (e.g. Kravet et al. 2008, Continelli et al., 2010). And again others make no distinction between health workers in using provider density (Eibich and Ziebhart, 2013). Hence, it is rather common to use different definitions of provider density in the literature, probably depending on the available data and purpose of the research. This study makes a distinction between nurses and midwives on the one hand and doctors and dentists on the other hand. Together they form skilled health providers. This is comparable to the researches of Anand and Bärnighausen (2004) and Kruk et al. (2009).

Continelli et al. (2010) indicated that an increased supply of GP's is correlated with more people indicating to have a GP. Busato and Künzi (2008) found that GP density had a strong positive association with utilization of GP consults. I also find that an increase in skilled health providers and doctors is related to an increase in GP consults, thus confirming findings in prior literature. Furthermore, the results indicate that health provider density has an insignificant influence on utilization of inpatient health care. These results are comparable to this study. It can be explained that this provides empirical evidence that the utilization of GP consults results in preventive care, as at least no additional inpatient care utilization is used, like Continelli et al. (2010) argue.

Moreover, many studies find weaker effects when including control variables in their model, just like this study does (e.g. Grumbach et al. (1997), Eibich and Ziebhart (2013). This endorses the statement that confounding variables are present and that provider density is definitely not the only explanatory factor of health care utilization.

On the other hand some results I find do not match with results from other studies. For instance Kravet et al. (2008), found for the US that an increase in GP density decreased utilization of primarily inpatient health care services. Wright and Ricketts (2010) found similar results and Eibich and Ziebhart (2013) also indicated that when inpatient care use was significantly high in a certain state, the level of doctor visits was low and vice versa. As depicted in Figure 5 there was no inverse relationship between utilization of inpatient and outpatient health care per province in 2010. In this study, doctor density showed a positive and significant relationship with inpatient health care use for the univariate regressions and a lack of effect for the multivariate regressions. Therefore it is not certain that utilization of GP consults truly results in less inpatient care use.

A remarkable finding of this study is that the density of nurses shows a negative relation with the probability of using health care. Although in the multivariate analysis the significance is largely reduced through the effect of the control variables, the coefficients in the univariate analysis are all negative and significant. Kruk et al. (2009) studied the effect of provider density on several outpatient health care services in 106 developing countries. The researchers only found one significant association for nurse concentration. It was a positive relation with the probability of using skilled birth attendants, the other coefficient for nurse density were insignificant but positive (Kruk et al., 2009). Anand and Bärnighausen (2004) studied the relation between nurse density and specific health outcomes, they only found a negative significant relation between concentration of nurses and maternal mortality. Because this is a health outcome, the negative association implies a positive effect of nurse density. However I find that when the number of nurses increases, the probability of health care utilization decreases (*USE05*) or does not change. The lack of association might be explained by the fact that the health care services observed in this study rather require a physician than a nurse. Kruk et al. (2009) state that in middle-income countries it is more common for physicians to provide health care to infants than in low-income countries. Peru is a middle-income country so this might apply to Peru as well. Another explanation can be that the quality of services provided by nurses and midwives is insufficient, resulting in utilization of health care in private facilities.

Several explanations exist for the insignificant observations. It is possible that other factors concerning the supply side of health care have a (combined) influence on health care utilization. For instance, the availability of technologies, drugs, facilities or quality of care are also important factors in influencing utilization of health care (Kruk et al., 2009). E.g. walking distance to a medical facility is found to be an important factor in using public health facilities, in this case for Honduras (Baker and Liu, 2006). The quality of care can be subverted by absenteeism. This study provides no insight in health providers actually being present in their facilities during working shifts. Even if enough health workers would be hired, absenteeism can still undermine health utilization rates. On average 25% of the health suppliers in primary health centers were absent in Peru in 2006 (Chaudhury et al., 2006). Furthermore, Banerjee and Duflo (2006) found that absenteeism is more common among rural and poor areas in India, which can also explain regional differences in health care utilization. Moreover, the way the supply side is governed might also matter. Stange (2013) found for the US that in states where nurses have more authority for prescription, utilization of health care was slightly higher. Gosden et al. (2001) found in their review of the literature that primary care physicians paid by a fee-for-system had more incentive to provide a higher quantity of health care. This was in contrast to physicians paid by salary or a per capitation system. The mentioned reasons can be magnified by the indication of Jiménez et al. (2015) that only half of the contracted physicians has a permanent engagement.

Another explanation can be that in developing countries it is rather common to have health providers with low education levels who cannot be recognized as physicians. Hence, if these community workers are not included a relation cannot be established (Kruk et al. 2009). This study neither includes every health worker, but solely the providers that are confirmed to be a professional physician or nurse.

Considering utilization of analysis, scans, ophthalmology and inpatient care, it might be that the relationship is insignificant due to the very low utilization rates of these services.

Finally the results could be subject to measurement error. As explained, omitted variable bias remains to exist and survey data are not completely reliable.

The results on inequity of this study contribute to the literature by measuring inequality by quintiles in terms on provider density. In general studies analyze the distribution of health care utilization across quintiles of income (e.g. van Doorslaer et al., 2006). By examining the distribution using provider density, it becomes clearer what the (unjustified) inequality is for people facing different levels of health worker density.

Van Doorslaer (2006) found that doctor visits was distributed more among the rich in general, especially for SP visits, in 21 OECD countries. I discovered a similar pattern for the distribution of inequality. However, the inequalities are very close to zero. This might be explained by the fact that the analyzed group did not capture the whole Peruvian population, but only the people insured by the public system or people without health insurance. The part of the population insured by employers or having a private health insurance, can in general be considered as a group that is wealthier. It can be that contradictions in my sample are less evident than when one analyzes a group representing a whole population.

Multiple studies used ‘visits to a GP or SP’ as outcome variable (van Doorslaer et al., 2006, Ozegowski and Sundmacher, 2013), while I use visit to a GP and a general measure of health care utilization. However, this does not seem to make a large difference for the comparability of the results in terms of decomposing inequity. Van Doorslaer et al. (2006) found that wealth is a driver of the in general a pro-rich utilization of doctors. Private health insurance had a reinforcing effect on the inequity as well. Gerdtham et al. (1997) reported horizontal inequity for doctor visits in Sweden, due to among others income and place of residence. For eight developing countries researchers found that richer quintiles would both receive more medicines and visit a physician more frequently (Makinen et al. 2000). In this study the evidence that wealth is a very important factor is confirmed. Place of residence and having a health insurance actually do have very low values and do not contribute much to inequity. I do find a negative contribution to inequity of nurses. It means that in areas with a low doctor density, nurses decrease inequality in utilization of health care. This implies that nurses can be a substitute for doctors when the number of physicians is lacking.
The results are furthermore contradicting in the influence of other socioeconomic factors, for instance education, professional status and marital status, like Gerdtham et al. (1997) did. In this analysis I found no evidence for other socioeconomic factors to increase or decrease inequity. This might be explained by the same reasoning as mentioned earlier. The sample in this study concentrates on a more vulnerable part of the Peruvian population. Another reason can be omitted variables, as the explanatory power of the residual was relatively large in the model using utilization of GP consults as outcome variable.

7.3 Limitations

The data and methodology I use in this research contain several limitations. First, it would have been interesting to examine the effect of health provider density on the level of utilization. This study concentrates on *whether* a person used health care or not. This provides no insight in the intensity of the utilization. Especially since the aggregated measure of health care services (*USEhc*) only registers if a person used one of the services, while it can be that someone used multiple services. Moreover, this study cannot completely explain how policy makers can increase health care utilization, because the research cannot answer questions about the quality of the supply side. Policy makers may enlarge the physician density, but if for instance absenteeism is a problem in Peruvian public health care institutions it may not have the desired effect.

Second, the data of the outcome variables on utilization and the control variables are derived from surveys. Although the INEI provides insight in their methodology, there is no possibility to control the way of collecting data or change the amount of observations per region. The outcome variables are completely dependent on the memory of individuals as they need to answer which health care services they used. Data obtained from health facilities might have been more reliable. Whether a person experienced any health problems is also undermined by subjectivity, since people assess their own health differently per person.
Furthermore, the data from the MINSA allowed to collect data on doctors, dentists, nurses and midwives. However, more health providers exist, also in the data from the MINSA, but a large group was labeled as ‘other health workers’. I excluded these health workers from the research as there is no information available on their professionalism and importance. It may however be that excluding this group has biased the results.
Another minor issue is that two geographical departments had to be deleted from the data. The reason was that the number of districts could not be matched during the years and the data were therefore unreliable. This has as a consequence that the results are not representative for the whole country.

Finally and probably most importantly, the model faces risks of endogeneity, caused by omitted variables and reverse causality. This matter has been explained before, but it would have been useful if the data had allowed to do a time-series research instead of cross-sectional analysis. By doing longitudinal research, time invariant characteristics can be been eliminated. This would have allowed for more causal inferences in the relation between provider density and health care utilization.

7.4 Policy Implications

This research is especially interesting for the Peruvian policymakers because the sample consists of only publicly insured or targeted people. The study has several implications addressing the health workforce in Peru, although the results should be treated carefully. From the main logistic regression results it can be interpreted that an increase in skilled health workers and especially doctors, increases the probability of visiting GP’s. Considering the lack of association between physician density and inpatient health care, it implies that GP visits have a preventive function. It would mean that if an increasing amount of people uses GP consults, use of inpatient care does not increase and costs can be contained. Therefore it may be useful to employ more primary physicians in the public health sector. Furthermore, the significant and positive associations between physician density and health care utilization should be an incentive for government to reduce migration of medicine graduates. Jiménez et al. (2015) indicates that 78% of the students would like to migrate because of better economic perspectives in other countries. I recommend the Peruvian government to address the underlying causes of outmigration.

The results on equity suggest that the utilization plotted by provider density is distributed quite equally. However, room for improvement exists concerning the distribution of physicians. The inequity analysis shows that health care utilization is unequal in regions with high provider density, due to the distribution of the physicians. When enlarging the health workforce it is probably most effective in terms of equality to distribute physicians to low density areas. Wealth is also an important factor determining inequality. The government already covers the expenses for the publicly and uninsured population, hence wealth should not be a problem. I suggest the government examines whether the reimbursement system is effective and otherwise enforces this.

7.5 Suggestions for Future Research

It would be interesting for future research to analyze the underlying factors of the results obtained in this study. The results indicate that provider density mainly correlates with visits to general practitioners, but provides no insight on why this occurs. As discussed, not only the number of providers is relevant, specifically the quality of the supply side is important as well (Kruk et al. 2009). Thus, analyzing the degree of absenteeism or patient satisfaction for instance on health care utilization would be interesting.

The results of this study also raise questions on the specific role of nurses and midwives in health care utilization. The associations are rather negative and it is interesting to know why in order to make well founded decisions. Is the relationship negative because nurses and midwives provide preventive care, therefore decreasing health care utilization in the long term? Or is the quality of services provided by nurses and midwives of insufficient quality, resulting in utilization in other facilities, for instance private clinics? Current literature has not found a similar association and so far no scientifically founded explanations exist for this phenomenon.

References

Anand, S. and Bärnighausen, T. (2004). Human resources and health outcomes: cross-country econometric study, *The Lancet*, 365, 1603 – 1609.

Andersen, R.M. and Newman, J.F. (1973). Societal and individual determinants of medical care utilization in the United States. *The Milbank Memorial Fund Quarterly: Health and Society,* 51, 95 – 124.

Andersen, R.M. (1995). Revisiting the behavioral model and access to medical care: does it matter? *Journal of Health and Social Behavior*, 36, 1 – 10.

Atun, R., Monteiro de Andrade, L.O., Almeida, G., Cotlear, D., Dmytraczenko, T., Frenz, P., … Wagstaff, A. (2014). Health-system reform and universal health coverage in Latin America, *The Lancet*, 1 – 15, doi: 10.1016/S0140-6736

Baker, J.B. and Liu, L. (2006). The determinants of primary health care utilization: a comparison of three rural clinics in Southern Honduras, *GeoJournal,* 66, 295 – 310.

Banerjee, A. and Duflo, E. (2006). Adressing Absence, *Journal of Economic Perspectives*¸ 20, 117 – 132.

Braveman, P. and Gruskin, S. (2002). Defining equity in health, *Community Health*, 57, 254 – 258.

Busato, A. and Künzi, B. (2008). Primary care physician supply and other key determinants of health care utilisation: the case of Switzerland, *BMC Health Services Research*, 8, 1 – 8, doi:10.1186/1472-6963-8-8

Chaudhury, N., Hammer, J., Kremer, M., Muralidharan, K. and Halsey, R. (2006). Missing in action: teacher and health worker absence in developing countries, *Journal of Economic Perspectives*, 20, 91 – 116.

Chen, L., Evans, T., Anand, S., Boufford, J.I., Brown, H., Chowdhury M., … and Wibulpolprasert, S. (2004) Human resources for health: overcoming the crisis, *The Lancet*, 364, 1984 – 1990.

Cometto G. and Witter, S. (2013). Tackling health workforce challenges to universal health coverage: setting targets and measuring progress, *Bulletin of World Health Organization*, 91, 881 – 885.

Continelli, T., McGinnis, S. and Holmes, T. (2010). The effect of local primary care physician supply on the utilization of preventive health services in the United States, *Health & Place*, 16, 942 – 951.

Doorslaer, van E., Masseria, C. and Koolman, X. (2006). Inequalities in access to medical care by income in developed countries, *CMAJ*, 174, 177 – 183.

Eibich, P. and Ziebarth, N.R. (2013). Analyzing Regional Variation in Health Care Utilization Using (Rich) Household Micro data, *Health Policy*, 114, 41 – 53.

Elo, I.T. (1992). Utilization of maternal health-care services in Peru: the role of woman's education, *Health Transformation Review*, 2, 49 – 69.

Francke, P. (2013). Peru's Comprehensive Health Insurance and New Challenges for Universal Coverage. *The World Bank Universal Health Coverage Studies Series*. 11, 1 – 27.

Gerdtham, U.G. (1997). Equity in health care utilization: further tests based on hurdle models and Swedish micro data, *Health Economics,* 6, 303 – 319.

Gosden, T., Forland, F., Kristiansen, I.S., Sutton, M., Leese, B., Giufridda, A., … , Pedersen, L. (2001) Impact of payment method on behaviour of primary care physicians: a systematic review. *Journal of Health Services Research & Policy,* 6, 44 – 55.

Grumbach, K., Vranizan, K. and Bindman, A.B. (1997). Physician supply and access to care in urban communities, *Health Affairs,* 16, 71 – 86.

Hosmer, D.W. and Lemeshow, S. (2000). *Applied Logistic Regression.* Retrieved from: http://media.hsph.edu.vn/sites/

Instituto Nacional de Estadistica e Informatica (2015). *Encuesta Nacional de Hogares* [Data file and code book]. Retrieved from: http://iinei.inei.gob.pe/microdatos/

Instituto Nacional de Estadistica e Informatica (2009). *Calidad de la Encuesta* (Encuesta Nacional de Hogares 2009). Retrieved from: http://iinei.inei.gob.pe/microdatos/

Jiménez, M., Mantilla, E., Huayanay, C., Mego, M. and Vermeersch, C. (2015). *Analysis of the health care labor market in Peru* (WB Discussion Paper No. 95115). Retrieved from https://openknowledge.worldbank.org/

Kravet, S.J., Shore, A.D., Miller, R., Green, G.B., Kolodner, K. and Wright, S.M. (2008). Health Care Utilization and the Proportion of Primary Care Physicians, *The American Journal of Medicine*, 121, 142 – 148.

Kruk, M.E., Prescott, M.R., Pinho, de H., and Galea, S. (2009). Are doctors and nurses associated with coverage of essential health services in developing countries? A cross-sectional study. *Human Resources for Health.*7, 1 – 9.

Labelle, R., Stoddart, G. and Rice, T. (1993). A Re-examination of the Meaning and Importance of Supplier-Induced Femand. *Journal of Health Economics,* 13, 347 – 368.

Long, J.S. and Freese, J. (2001). *Regression Models for Categorical Dependent Variables Using Stata*. Retrieved from: https://is.muni.cz/el/

Makinen, M., Waters, H., Rauch, M., Almagambetova, N., Bitran, R., Gilson, L., … , and Ram, S. (2000). Inequalities in Health Care Use and Expenditures: Empirical Data from Eight Developing Countries and Countries in Transition. *Bulletin of the World Health Organization.* 78, 55 – 65.

Ministerio de Salud Peru (2015). *Informacion Estadística: Departmental y Distrital* [Data file and code book]. Retrieved from: http://www.minsa.gob.pe/index.asp?op=6

Ministerio de Salud Peru (2009). Peru: baselines 2007 – 2015 HRH goals. Retrieved from: http://www.observatoriorh.org/sites/default/files/webfiles/fulltext/baseline_peru_eng.pdf

O'Donnell, O., Doorslaer van E., Wagstaff, A. and Lindelow, M. (2008). Analyzing health equity using household survey data, *The World Bank Washington DC.*

Options Consultancy Services/Evidence for Action, Cambridge Economic Policy Associates , and the Partnership for Maternal, Newborn & Child Health. Success Factors for Women's and Children's Health: Country Specific Review of Data and Literature on 10 Fast-Track Countries' Progress Towards MDGs 4 and 5. 2013.

Ozegowski, S. and Schumacher, L. (2013). Understanding the gap between need and utilization in outpatient care – The effect of supply-side determinants on regional inequities, *Health Policy,* 114, 54 – 63, doi: 10.1016/j.healthpol.2013.08.005

Pan American Health Organization. (2012). *Health in the Americas. Edition: Country Volume.* (PAHO 2012). Retrieved from http://www.paho.org/saludenlasamericas/

Poel, van der E., O'Donnell, O. and Doorslaer, van E. (2008). Urbanization and the Spread of Diseases of Affluence in China. *Health Econometrics and Data Group Working Paper.* 08/25.

Stange, K. (2013). How does provider supply and regulation influence health care markets? Evidence from nurse practitioners and physician assistants, *Journal of Health Economics,* 33, 1 – 27.

Sundmacher, L. and Busse, R. (2011). The impact of physician supply on avoidable cancer deaths in Germany. A spatial analysis. *Health Policy,* 103, 53 – 62.

United Nations. (2005). *Investing in development: a practical plan to achieve the Millennium Development Goals*. New York. Retrieved from: http://www.unmillenniumproject.org/

Valdivia, M. (2002) Public health infrastructure and equity in the utilization of outpatient health care services in Peru, *Health Policy and Planning*, 17, 12 – 19.

Valdivia, M. (2004). Poverty, Health Infrastructure and the Nutrition of Peruvian Children, *Working paper of the Inter-American Development Bank,* R-498.

Wagstaff, A. (2011). The Concentration Index of a Binary Outcome Revisited. *Health Economics.* 20, 1155 – 1160.

Wagstaff, A., Doorslaer, van E. and Paci, P. (1990). On the measurement of horizontal inequity in the delivery of health care, *Journal of Health Economics*, 10, 169 – 205.

Wagstaff, A. and Doorslaer, van E. (1998). Equity in Health Care Finance and Delivery, *Handbook of Health Economics.*

Wennberg, J.E., Barnes, B.A. and Zubkoff, M. (1982). Professional uncertainty and the problem of supplier-induced demand, *Social Science & Medicine*, 16, 811 – 824.

Whitehead, M. (1991). The concepts and principles of equity and health, *Health Promotion International,* 6, 218 – 228.

Williams, R. (2015). Marginal Effects for Continuous Variables, *University of Notre Dame*. Retrieved from: https://www3.nd.edu/~rwilliam/stats3/Margins02.pdf

World Bank. (2011). *Peru Recurso Programmatic AAA – Phase IV – Improving health outcomes by strengthening users' entitlements and reinforcing public sector management* (WB Report No. 59218-PE). Retrieved from https://openknowledge.worldbank.org/

World Health Organization (2010). *World Health Statistics 2010.* (NLM Classification WA 900.1) Retrieved from: http://www.who.int/gho/publications/

Wright, D.B. and Ricketts, T.C. (2010). The road to efficiency? Re-examining the impact of the primary care physician workforce on health care utilization rates, *Social Science and Medicine,* 70, 2006 – 2010.

Appendix

Table 10. Pearson correlations

Pearson correlations	skilled	doctors	nurses	EDUlev	AgeSex	MARcon	URBdum	LnCON monpc	POVnonH	OCCsec	OCCsim	INStype	POV_INS	ILLany	INS_ILL	CRE	SVYyear	RESpro
skilled	1.0000																	
doctors	0.8613*	1.0000																
nurses	0.9586*	0.6810*	1.000															
EDUlev	0.0307*	0.1489*	-0.0393	1.000														
AgeSex	0.0535*	0.0491*	0.0495*	-0.0966*	1.000													
MARcon	-0.0247*	-0.0407*	-0.0128*	0.0768*	-0.1146*	1.000												
URBdum	-0.0537*	0.1422*	-0.1572*	0.4351*	-0.0050*	-0.0923	1.000											
LnCONmonpc	0.0660*	0.2486*	-0.0442*	0.4849*	0.0988*	-0.1392*	0.5664*	1.000										
POVnonH	-0.0875*	-0.1925*	-0.0182*	-0.3656*	-0.0928*	0.1117*	-0.3327*	-0.7722*	1.000									
OCCsec	0.0045	0.0528*	-0.0231*	0.1783*	0.3773*	-0.0620	0.2112*	0.2701*	-0.2178*	1.000								
OCCsim	0.0171*	-0.0132*	0.0320*	-0.0505*	0.4102*	-0.099*	-0.1162*	-0.0317*	0.0046	0.6333*	1.000							
INStype	-0.0797*	0.0586*	-0.1478*	0.2156*	0.2562*	-0.0723*	0.2803*	0.3148*	-0.2357*	0.2735*	0.1593*	1.000						
POV_INS	0.0217*	-0.1140*	0.0952*	-0.2939*	-0.2234*	0.0912*	-0.3222*	-0.5320*	0.6055*	-0.2679*	-0.1283	-0.7296*	1.000					
ILLany	0.0049*	-0.0067*	0.0108*	-0.0309*	0.1356*	-0.0397*	0.0069*	0.0284*	-0.0249*	0.0322*	0.0363*	-0.0209*	-0.0112*	1.000				
INS_ILL	0.0604*	-0.0340*	0.1062*	-0.1460*	-0.0850*	0.0260*	-0.1666*	-0.1720	0.1277*	-0.1492*	-0.0720*	-0.6582*	0.4506*	0.5194*	1.000			
CRE	-0.0171*	-0.2043*	0.0899*	-0.3589*	-0.0317*	0.0680*	-0.5543*	-0.4985	0.3394*	-0.1638*	0.0497*	-0.2712*	0.3419*	-0.0105*	0.1607*	1.000		
SVYyear	0.2165*	0.1164*	0.2468*	0.0259*	0.0302*	-0.0085*	-0.0168*	0.0983*	-0.1039*	0.0194*	0.0151*	-0.2695*	0.1393*	0.0336*	0.1986*	0.0112*	1.000	
RESpro	-0.2528*	-0.1591*	-0.2752*	0.1251*	-0.0096*	-0.0028	0.2301*	0.2104*	-0.1546*	0.0656*	-0.0283*	0.1222*	-0.1666*	0.0008	-0.0776*	-0.2813	-0.0026	1.000

Table 11. Descriptive statistics of the control variables

Variable	Coding	Value
AgeSex	Female 0 – 4	4.99%
	Male 0 – 4	5.08%
	Female 5 – 14	11.33%
	Male 5 – 14	11.50%
	Female 15 – 24	9.89%
	Male 15 – 24	10.37%
	Female 25 – 34	7.10%
	Male 25 – 34	6.48%
	Female 35 – 34	6.00%
	Male 35 – 44	5.54%
	Female 45 – 54	4.68%
	Male 45 – 54	4.37%
	Female 55 – 64	3.02%
	Male 55 – 64	2.97%
	Female 65 – 74	2.05%
	Male 65 – 74	1.86%
	Female 75+	1.59%
	Male 75+	1.17%
MARcon	1 = if married or has legal partnership	83.92%
	0 = if not married or has no legal partnership	16.08%
URBdum	1 = if urban area	52.51%
	0 = rural area	47.49%
EDUlev	0 = no adults in household	0.00%
	1 = none	1.80%
	2 = primary education incomplete	7.76%
	3 = primary education completed	10.68%
	4 = secondary education incomplete	20.10%
	5 = secondary education completed	29.37%
	6 = any tertiary education	30.29%
CONmonpc	Mean monthly non-health consumption per capita	233.41
	Standard deviation	179.56
	Minimum	8.35
	Maximum	5145.20
POVnonH	1 = non-health consumption below poverty line	54.75%
	0 = non-health consumption above poverty line	45.25%
OCCsec	1 = not working	45.42%
	2 = agriculture/fishing	25.44%
	3 = mining/manufacturing/construction	6.87%
	4 = services	22.43%

OCCsim	0 = not working	45.26%
	1 = professionals	2.52%
	2 = clerical support services	8.71%
	3 = skilled agricultural work	11.37%
	4 = craft and related trade workers	3.85%
	5 = plant and machine operators	3.76%
	6 = elementary	24.54%
INStype	1 = none	54.55%
	0 = SIS only	45.45%
ILLany	1 = if any health problem in the past 30 days	42.77%
	0 = if no health problem in the past 30 days	57.23%
*INS*ILL*	1 = if being insured and having a health problem	26.53%
	0 = if not insured and/or not having a health problem	73.47%
CRE	1 = if district has CRECER program	68.71%
	0 = if district has no CRECER program	31.39%
SVYyear	1 = 2007 observations	62,786
	2 = 2009 observations	59,722
	3 = 2010 observations	58,162
RESpro	Number of provinces	158
Total	Number of observations	180,670

Table 12. Results of the control variables for logistic regressions Model 1 with Skilled- average marginal effects

Multivariate analysis	USE01	USE03	USE04	USE05	USE07	USE13	USEhc	USE02
Skilled	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
AgeSex								
Male 0 – 4	-0.001	0.000	0.002**	0.000	-0.001	0.006***	0.004	0.007
	(0.005)	*(0.002)*	*(0.000)*	*(0.000)*	*(0.001)*	*(0.002)*	*(0.006)*	*(0.005)*
Female 5 – 14	-0.102***	-0.007***	0.000	-0.000	0.013***	-0.15***	-0.103***	-0.047***
	(0.005)	*(0.002)*	*(0.000)*	*(0.000)*	*(0.001)*	*(0.002)*	*(0.005)*	*(0.004)*
Male 5 – 14	-0.103***	-0.007***	0.001**	0.000	0.009***	-0.012***	-0.104***	-0.048***
	(0.005)	*(0.002)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.002)*	*(0.005)*	*(0.004)*
Female 15 – 24	-0.109***	0.001	0.007***	0.000	0.015***	0.046***	-0.052***	-0.061***
	(0.005)	*(0.002)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.002)*	*(0.005)*	*(0.005)*
Male 15 – 24	-0.116***	-0.005***	0.005***	0.000	0.012***	0.008***	-0.090***	-0.056***
	(0.005)	*0.002*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.002)*	*(0.006)*	*(0.005)*
Female 25 – 34	-0.085***	0.000	0.016***	0.001***	0.012***	0.070***	-0.021***	-0.036***
	(0.005)	*(0.002)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.006)*	*(0.005)*
Male 25 – 34	-0.090***	0.018***	0.011***	0.003**	0.011***	0.022***	-0.057***	-0.028***
	(0.006)	*(0.002)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.006)*	*(0.005)*
Female 35 – 34	-0.084***	0.004*	0.018***	0.001***	0.012***	0.052***	-0.035***	-0.032***
	(0.006)	*(0.003)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.006)*	*(0.005)*
Male 35 – 44	-0.101***	0.023***	0.013***	0.003**	0.011***	0.027***	-0.066***	-0.040***
	(0.006)	*(0.003)*	*(0.002)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.006)*	*(0.006)*
Female 45 – 54	-0.098***	0.022***	0.018***	0.005***	0.025***	0.036***	-0.052***	-0.037***
	(0.006)	*(0.003)*	*(0.002)*	*(0.001)*	*(0.002)*	*(0.003)*	*(0.006)*	*(0.005)*
Male 45 – 54	-0.109***	0.008***	0.013***	0.002*	0.023***	0.021***	-0.066***	-0.044***
	(0.006)	*(0.003)*	*(0.002)*	*(0.001)*	*(0.002)*	*(0.003)*	*(0.007)*	*(0.006)*
Female 55 – 64	-0.113***	0.025***	0.018***	0.004***	0.027***	0.031***	-0.071***	-0.044***
	(0.006)	*(0.003)*	*(0.002)*	*(0.001)*	*(0.002)*	*(0.004)*	*(0.007)*	*(0.006)*
Male 55 – 64	-0.110***	0.015***	0.020***	0.004**	0.021***	0.031***	-0.067***	-0.046***
	(0.007)	*(0.003)*	*(0.002)*	*(0.002)*	*(0.003)*	*(0.004)*	*(0.007)*	*(0.006)*
Female 65 – 74	-0.111***	0.026***	0.017***	0.004***	0.023***	0.035***	-0.069***	-0.034***
	(0.007)	*(0.004)*	*(0.002)*	*(0.001)*	*(0.003)*	*(0.005)*	*(0.007)*	*(0.006)*
Male 65 – 74	-0.099***	0.024***	0.021***	0.005**	0.030***	0.057***	-0.042***	-0.041***
	(0.008)	*(0.004)*	*(0.003)*	*(0.002)*	*(0.003)*	*(0.006)*	*(0.008)*	*(0.007)*
Female 75+	-0.132***	0.020***	0.019***	0.005***	0.028***	0.040***	-0.089***	-0.047***
	(0.007)	*(0.004)*	*(0.002)*	*(0.002)*	*(0.004)*	*(0.005)*	*(0.008)*	*(0.007)*
Male 75+	-0.105***	0.024***	0.026***	0.015***	0.039***	0.078***	-0.038***	-0.050***
	(0.008)	*(0.005)*	*(0.004)*	*(0.004)*	*(0.005)*	*(0.007)*	*(0.010)*	*(0.008)*
EDUlev								
1	-0.177	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	-0.157	-0.011
	(0.146)						*(0.142)*	*(0.145)*
2	-0.157	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	-0.133	0.025
	(0.146)						*(0.142)*	*(0.145)*
3	-0.151	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	-0.127	0.035
	(0.146)						*(0.142)*	*(0.145)*
4	-0.148	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	-0.124	0.039
	(0.146)						*(0.142)*	*(0.145)*
5	-0.144	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	-0.118	0.042
	(0.146)						*(0.142)*	*(0.145)*
6	-0.147	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	-0.114	0.039
	(0.146)						*(0.142)*	*(0.145)*

MARcon	0.008***	0.003***	0.001**	-0.000	0.002***	0.005***	0.014***	0.007***
	(0.002)	*(0.001)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.002)*
URBdum	0.007***	0.003***	-0.000	0.001**	0.007***	0.008***	0.013***	0.018***
	(0.003)	*(0.001)*	*(0.001)*	*(0.000)*	*(0.001)*	*0.001*	*(0.003)*	*(0.003)*
lnCONmonpc	0.046***	0.013***	0.006***	0.001***	0.015***	0.008***	0.060***	0.062***
	(0.003)	*(0.001)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.001)*	*(0.002)*	*(0.002)*
OCCsec								
2	-0.033***	-0.014***	-0.010***	-0.004***	-0.002*	-0.025***	-0.054***	-0.034***
	(0.003)	*(0.001)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.002)*	*(0.003)*	*(0.003)*
3	-0.014***	-0.014***	-0.009***	-0.003***	-0.000	-0.023***	-0.034***	-0.009**
	(0.004)	*(0.002)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.002)*	*(0.004)*	*(0.004)*
4	-0.014***	-0.010***	-0.007***	-0.003***	0.000	0.022***	-0.029***	-0.003
	(0.003)	*(0.001)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.003)*
INStype	0.079***	0.008	-0.012*	-0.001	-0.000	-0.030***	-0.018***	0.075***
	(0.021)	*(0.008)*	*(0.007)*	*(0.001)*	*(0.001)*	*(0.002)*	*(0.005)*	*(0.010)*
ILLany	0.370***	0.038***	0.019***	0.003***	0.004***	0.010***	0.367***	0.604***
	(0.006)	*(0.001)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.004)*
*INS*ILL*	0.155***	0.026***	-0.006	0.000	0.001	0.001	0.087***	0.115***
	(0.022)	*(0.010)*	*(0.004)*	*(0.000)*	*(0.002)*	*(0.002)*	*(0.006)*	*(0.009)*
CRE	0.004	-0.002	-0.000	0.000	0.000	-0.006***	0.001	-0.006**
	(0.003)	*(0.002)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.002)*	*(0.003)*	*(0.003)*
SVYyear								
2009	0.017***	0.000	0.001	0.000	0.000	-0.004***	0.014***	-0.009***
	(0.002)	*(0.001)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.001)*	*(0.002)*	*(0.002)*
2010	0.016***	-0.004***	-0.000	0.001*	-0.000	-0.004***	0.011***	0.003
	(0.003)	*(0.001)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.002)*
RESpro	x	x	x	x	x	x	x	x
No. Observations	180,670	178,744	172,102	111,612	176,842	180,662	180,670	180,670

*** coefficient is significant at α<=0.001 ** coefficient is significant at α<=0.005 * coefficient is significant at α<=0.10. Standard errors are robust standard errors. 158 provinces are included in all models, but to save space these results are not presented in this table.

Tabel 13. Results of the control variables for logistic regressions Model 2 with Doctors & Nurses- average marginal effects

Multivariate analysis	USE01	USE03	USE04	USE05	USE07	USE13	USEhc	USE02
Doctors & Nurses	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient	Coefficient
AgeSex								
Male 0 – 4	-0.001	0.000	0.002**	0.000	-0.001	0.007***	0.004	0.007
	(0.005)	*(0.002)*	*(0.000)*	*(0.000)*	*(0.001)*	*(0.002)*	*(0.006)*	*(0.005)*
Female 5 – 14	-0.101***	-0.007***	0.000	-0.000	0.013***	-0.15***	-0.102***	-0.047***
	(0.005)	*(0.002)*	*(0.000)*	*(0.000)*	*(0.001)*	*(0.002)*	*(0.005)*	*(0.004)*
Male 5 – 14	-0.102***	-0.007***	0.001**	0.000	0.009***	-0.012***	-0.104***	-0.048***
	(0.005)	*(0.002)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.002)*	*(0.005)*	*(0.004)*
Female 15 – 24	-0.107***	0.001	0.007***	0.000	0.015***	0.046***	-0.050***	-0.060***
	(0.005)	*(0.002)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.002)*	*(0.005)*	*(0.005)*
Male 15 – 24	-0.102***	-0.004**	0.005***	0.000	0.012***	0.008***	-0.089***	-0.055***
	(0.005)	*0.002*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.002)*	*(0.006)*	*(0.005)*
Female 25 – 34	-0.108***	0.012***	0.015***	0.001***	0.012***	0.070***	-0.020***	-0.035***
	(0.005)	*(0.002)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.006)*	*(0.005)*
Male 25 – 34	-0.116***	0.001	0.011***	0.003***	0.011***	0.022***	-0.059***	-0.030***
	(0.006)	*(0.002)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.006)*	*(0.005)*
Female 35 – 34	-0.083***	0.018***	0.018***	0.001**	0.012***	0.051***	-0.034***	-0.031***
	(0.006)	*(0.003)*	*(0.001)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.006)*	*(0.005)*
Male 35 – 44	-0.104***	0.005**	0.013***	0.004**	0.011***	0.027***	-0.069***	-0.042***
	(0.006)	*(0.003)*	*(0.002)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.006)*	*(0.006)*
Female 45 – 54	-0.096***	0.023***	0.019***	0.005***	0.026***	0.036***	-0.050***	-0.036***
	(0.006)	*(0.003)*	*(0.002)*	*(0.001)*	*(0.002)*	*(0.003)*	*(0.006)*	*(0.005)*
Male 45 – 54	-0.112***	0.008***	0.013***	0.002*	0.022***	0.021***	-0.069***	-0.047***
	(0.006)	*(0.003)*	*(0.002)*	*(0.001)*	*(0.002)*	*(0.003)*	*(0.007)*	*(0.006)*
Female 55 – 64	-0.112***	0.026***	0.018***	0.004***	0.027***	0.030***	-0.070***	-0.043***
	(0.006)	*(0.003)*	*(0.002)*	*(0.001)*	*(0.002)*	*(0.004)*	*(0.007)*	*(0.006)*
Male 55 – 64	-0.113***	0.015***	0.019***	0.004**	0.021***	0.030***	-0.070***	-0.049***
	(0.007)	*(0.003)*	*(0.002)*	*(0.002)*	*(0.003)*	*(0.004)*	*(0.007)*	*(0.006)*
Female 65 – 74	-0.110***	0.026***	0.018***	0.004***	0.024***	0.035***	-0.068***	-0.033***
	(0.007)	*(0.004)*	*(0.002)*	*(0.001)*	*(0.003)*	*(0.005)*	*(0.007)*	*(0.006)*
Male 65 – 74	-0.102***	0.024***	0.020***	0.005**	0.029***	0.056***	-0.045***	-0.044***
	(0.008)	*(0.004)*	*(0.003)*	*(0.002)*	*(0.003)*	*(0.006)*	*(0.008)*	*(0.007)*
Female 75+	-0.132***	0.020***	0.019***	0.005***	0.029***	0.040***	-0.088***	-0.046***
	(0.007)	*(0.004)*	*(0.002)*	*(0.002)*	*(0.004)*	*(0.005)*	*(0.008)*	*(0.007)*
Male 75+	-0.106***	0.023***	0.025***	0.015***	0.038***	0.078***	-0.040***	-0.052***
	(0.008)	*(0.005)*	*(0.004)*	*(0.004)*	*(0.005)*	*(0.007)*	*(0.010)*	*(0.008)*
EDUlev								
1	-0.184	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	-0.165	-0.016
	(0.143)						*(0.141)*	*(0.144)*
2	-0.163	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	-0.139	0.021
	(0.143)						*(0.141)*	*(0.144)*
3	-0.156	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	-0.133	0.031
	(0.143)						*(0.141)*	*(0.144)*
4	-0.152	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	-0.129	0.035
	(0.143)						*(0.141)*	*(0.144)*
5	-0.149	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	-0.123	0.038
	(0.143)						*(0.141)*	*(0.144)*
6	-0.153	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	-0.121	0.035
	(0.143)						*(0.141)*	*(0.144)*

MARcon	0.008***	0.003***	0.001**	-0.000	0.002***	0.005***	0.014***	0.007***
	(0.002)	*(0.001)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.002)*
URBdum	0.009***	0.004***	-0.000	0.001**	0.007***	0.008***	0.014***	0.018***
	(0.003)	*(0.001)*	*(0.001)*	*(0.000)*	*(0.001)*	*0.001*	*(0.003)*	*(0.003)*
lnCONmonpc	0.039***	0.010***	0.005***	0.001***	0.014***	0.009***	0.054***	0.057***
	(0.003)	*(0.001)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.002)*
POVnonH	-0.018***	-0.003**	-0.001	-0.001	-0.002**	0.004**	-0.017***	-0.012***
	(0.003)	*(0.002)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.002)*	*(0.004)*	*(0.003)*
OCCsec	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)
OCCsim	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)	(not estimable)
INStype	0.086***	0.009	-0.013*	-0.001	-0.001	-0.033***	-0.011**	0.080***
	(0.021)	*(0.008)*	*(0.007)*	*(0.001)*	*(0.001)*	*(0.002)*	*(0.006)*	*(0.010)*
ILLany	0.369***	0.038***	0.019***	0.003***	0.004***	0.010***	0.367***	0.605***
	(0.006)	*(0.001)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.004)*
*INS*ILL*	0.154***	0.026***	-0.006	0.000	0.001	0.001	0.086***	0.114***
	(0.022)	*(0.009)*	*(0.004)*	*(0.000)*	*(0.002)*	*(0.002)*	*(0.006)*	*(0.009)*
CRE	0.004	-0.002	-0.000	0.000	0.000	-0.006***	0.001	-0.006**
	(0.003)	*(0.002)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.002)*	*(0.003)*	*(0.003)*
SVYyear								
2009	0.016***	0.000	0.001	0.000	0.000	-0.004***	0.013***	-0.009***
	(0.002)	*(0.001)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.001)*	*(0.002)*	*(0.002)*
2010	0.017***	-0.003***	-0.000	0.001*	-0.000	-0.003**	0.012***	0.002
	(0.003)	*(0.001)*	*(0.001)*	*(0.000)*	*(0.001)*	*(0.001)*	*(0.003)*	*(0.003)*
RESpro	x	x	x	x	x	x	x	x
No. Observations	180,670	178,744	172,102	111,612	176,842	180,662	180,670	180,670

*** coefficient is significant at α<=0.001 ** coefficient is significant at α<=0.005 * coefficient is significant at α<=0.10. Standard errors are robust standard errors. 158 provinces are included in all models, but to save space these results are not presented in this table.

Table 14. Fitstat measures probit and logit for *USEhc* (left: models with skilled, right: models with doctors & nurses)

```
Measures of Fit for logit of USEhc

Log-Lik Intercept Only:    -107280.015   Log-Lik Full Model:          -81392.947
D(180469):                  162785.895   LR(200):                      51774.135
                                         Prob > LR:                        0.000
McFadden's R2:                   0.241   McFadden's Adj R2:                0.239
ML (Cox-Snell) R2:               0.249   Cragg-Uhler(Nagelkerke) R2:       0.358
McKelvey & Zavoina's R2:         0.424   Efron's R2:                       0.259
Variance of y*:                  5.707   Variance of error:                3.290
Count R2:                        0.763   Adj Count R2:                     0.157
AIC:                             0.903   AIC*n:                       163187.895
BIC:                        -2.022e+06   BIC':                        -49353.249
BIC used by Stata:          165218.885   AIC used by Stata:           163187.895
```

```
Measures of Fit for probit of USEhc

Log-Lik Intercept Only:    -107280.015   Log-Lik Full Model:          -81426.770
D(180469):                  162853.540   LR(200):                      51706.489
                                         Prob > LR:                        0.000
McFadden's R2:                   0.241   McFadden's Adj R2:                0.239
ML (Cox-Snell) R2:               0.249   Cragg-Uhler(Nagelkerke) R2:       0.358
McKelvey & Zavoina's R2:         0.432   Efron's R2:                       0.258
Variance of y*:                  1.762   Variance of error:                1.000
Count R2:                        0.763   Adj Count R2:                     0.156
AIC:                             0.904   AIC*n:                       163255.540
BIC:                        -2.022e+06   BIC':                        -49285.603
BIC used by Stata:          165286.530   AIC used by Stata:           163255.540
```

```
Measures of Fit for logit of USEhc

Log-Lik Intercept Only:    -107280.015   Log-Lik Full Model:          -81383.288
D(180468):                  162766.576   LR(201):                      51793.453
                                         Prob > LR:                        0.000
McFadden's R2:                   0.241   McFadden's Adj R2:                0.240
ML (Cox-Snell) R2:               0.249   Cragg-Uhler(Nagelkerke) R2:       0.359
McKelvey & Zavoina's R2:         0.424   Efron's R2:                       0.259
Variance of y*:                  5.709   Variance of error:                3.290
Count R2:                        0.763   Adj Count R2:                     0.157
AIC:                             0.903   AIC*n:                       163170.576
BIC:                        -2.022e+06   BIC':                        -49360.463
BIC used by Stata:          165211.671   AIC used by Stata:           163170.576
```

```
Measures of Fit for probit of USEhc

Log-Lik Intercept Only:    -107280.015   Log-Lik Full Model:          -81418.320
D(180468):                  162836.640   LR(201):                      51723.389
                                         Prob > LR:                        0.000
McFadden's R2:                   0.241   McFadden's Adj R2:                0.239
ML (Cox-Snell) R2:               0.249   Cragg-Uhler(Nagelkerke) R2:       0.358
McKelvey & Zavoina's R2:         0.433   Efron's R2:                       0.258
Variance of y*:                  1.763   Variance of error:                1.000
Count R2:                        0.763   Adj Count R2:                     0.155
AIC:                             0.904   AIC*n:                       163240.640
BIC:                        -2.022e+06   BIC':                        -49290.399
BIC used by Stata:          165281.734   AIC used by Stata:           163240.640
```