OLENA KALINICHENKO

EXPLORING PERFORMANCE ASSESSMENT AND INCENTIVE REGULATION IN THE CONTEXT OF PRIMARY HEALTH CARE

UNIVERSIDADE DO ALGARVE

Faculdade de Economia

2016

OLENA KALINICHENKO

EXPLORING PERFORMANCE ASSESSMENT AND INCENTIVE REGULATION IN THE CONTEXT OF PRIMARY HEALTH CARE

Doutoramento em Ciências Económicas e Empresariais

(Especialidade em Gestão)

Trabalho efetuado sob a orientação de:

Prof. Doutora Carla Alexandra da Encarnação Filipe Amado,

Prof. Doutor Sérgio Pereira dos Santos

UNIVERSIDADE DO ALGARVE

Faculdade de Economia

2016

EXPLORING PERFORMANCE ASSESSMENT AND INCENTIVE REGULATION IN THE CONTEXT OF PRIMARY HEALTH CARE

Declaração de autoria de trabalho

Declaro ser a autora deste trabalho, que é original e inédito. Autores e trabalhos consultados estão devidamente citados no texto e constam da listagem de referências incluída.

Olena Kalinichenko

_____

Direitos de cópia ou Copyright

RESUMO

No contexto dos desafios globais nos cuidados de saúde, o desenvolvimento de sistemas de cuidados de saúde primários tem tido prioridade, enquanto a avaliação do desempenho na prestação de cuidados de saúde primários se assume como uma das direções centrais de investigação. Não existe, no entanto, uma forma padronizada sob a qual o desempenho dos prestadores de cuidados primários possa ser medido, tendo em conta a diversidade de definições para o termo "desempenho" em relação a esses mesmos prestadores. Além disso, não é claro que a investigação empírica nesta área tenha estado associada, com sucesso, à prática e política de cuidados de saúde.

A este respeito, o primeiro artigo incluído nesta tese pretende alcançar os dois objetivos seguintes: (1) explorar as dimensões e indicadores de desempenho principais das avaliações dos cuidados primários, ao efetuar uma revisão sistemática da literatura que adota um enquadramento concetual abrangente para a avaliação de desempenho dos prestadores de cuidados primários; e (2) examinar o potencial das avaliações dos cuidados primários em relação à melhoria de desempenho na prática e política de cuidados de saúde, através do prisma do envolvimento das partes interessadas (stakeholders) em três diferentes fases da avaliação de desempenho, tal como proposto por Smith (1996). Desenvolvemos um enquadramento concetual que inclui três fases de avaliação de desempenho (medição, análise, e ação) e que captura quatro dimensões de desempenho (acessibilidade, equidade, eficiência, e eficácia). Aplicámos este enquadramento para uma revisão sistemática de 166 artigos empíricos dedicados à avaliação de desempenho de prestadores de cuidados primários. Os resultados da revisão mostram que uma ampla variedade de medições de desempenho foi utilizada para avaliar prestadores de cuidados primários, as quais são sistematizadas em tabelas sumárias para auxiliar a seleção de um conjunto apropriado de indicadores de desempenho para futuras avaliações de cuidados primários. Também concluímos que as quatro dimensões têm sido maioritariamente exploradas de forma separada na literatura empírica analisada. Levando em conta as relações existem entre as diferentes dimensões de desempenho, pensamos que os futuros estudos de investigação devem aplicar maiores esforços para avaliar simultaneamente várias dimensões de desempenho e para explorar de forma mais profunda as relações existentes entre elas. Os resultados da

revisão mostram também um envolvimento insuficiente por parte dos interessados no processo de avaliação de desempenho, indicando o impacto potencialmente limitado da investigação analisada. Na maioria dos casos, as partes interessadas envolvidas nas avaliações de cuidados primários foram membros de staff, legisladores, e gestores, sendo que apenas uma pequena minoria de estudos relata o envolvimento de outras partes interessadas relevantes, como pacientes e cidadãos. Por conseguinte, aconselha-se que estudos futuros tenham o intento de envolver partes interessadas cruciais no processo de avaliação de cuidados primários, incluindo pacientes e cidadãos, entre outros, para garantir que os modelos e resultados têm validade aparente (*face validity*) e que consigam também fortalecer o impacto que a investigação empírica tem na prática e política de cuidados de saúde.

O segundo artigo incluído nesta tese explora o potencial da Análise Envoltória de Dados (*DEA*), uma técnica não paramétrica para a avaliação relativa de desempenho proposta por Charnes, Cooper e Rhodes (1978), a fim de melhorar delinear programas de pagamento por desempenho (*pay-for-performance* ou *P4P*) nos cuidados de saúde primários. Os programas de pagamento por desempenho são atualmente vistos como uma ferramenta promissora para melhorar a prestação de cuidados de saúde, e não existem dúvidas de que a conceção dos programas P4P tem um papel crucial na forma como os seus objetivos são cumpridos. Neste artigo pretendemos investigar um elemento fundamental na conceção destes programas que não tem recebido muita atenção na literatura P4P – o método para a avaliação de desempenho. Acreditamos que a utilização da DEA permite uma avaliação mais justa dos prestadores no contexto do P4P quando comparada com a análise de proporção (*ratio analysis*), que é mais tradicional nos programas P4P. Uma das principais vantagens da DEA, quando comparada com a análise de proporção, é que ela permite que cada unidade seja avaliada com uma maior ênfase nos indicadores de desempenho, nos quais a unidade avaliada teve um melhor desempenho, levando simultaneamente em conta as relações entre os diferentes indicadores e dimensões de desempenho. A DEA também compara cada unidade com um grupo de outras unidades que operam sob condições similares e com prioridades de desempenho iguais, identificando aqueles prestadores que são os melhores executantes e identificando metas para a melhoria de desempenho das unidades que obtêm um resultado inferior a 100% na avaliação da DEA.

A este respeito, este artigo tem dois objetivos de investigação: (1) desenvolver uma abordagem metodológica com base na DEA para a avaliação de desempenho de prestadores de cuidados de saúde primários em Portugal; e (2) propor uma metodologia alternativa para a contratualização de P4P por parte destes prestadores, com base nos resultados da avaliação da DEA. Para alcançar o primeiro objetivo, modificámos o modelo básico da DEA ao adotar a abordagem de benefício da dúvida (*the benefit-of-the-doubt approach*) e ao introduzir uma combinação única de restrições de pesos – um parâmetro que regula o nível de disparidade entre os pesos virtuais ideais que está combinado com restrições de pesos virtuais proporcionais e ordinais. Na nossa perspetiva, tal combinação de restrições de pesos faculta o melhor equilíbrio entre as preferências derivadas exogenamente a partir da relevância relativa de cada indicador de desempenho e de cada dimensão e o nível de disparidade acessível definido endogenamente na atribuição dos pesos virtuais. As restrições estabelecidas permitem ainda um grande grau de flexibilidade na atribuição dos pesos, em oposição a qualquer sistema fixo de pesos, preservando dessa forma uma das principais vantagens da DEA. A abordagem proposta com base na DEA também tem explicitamente em conta a influência de fatores ambientais no alcançar das metas estabelecidas pelo sistema P4P. Em particular, ao termos analisado os índices de Malmquist propostos por Camanho e Dyson (2006) e os coeficientes de correlação de Pearson, identificámos quatro fatores ambientais relevantes – o número de anos em operação das unidades prestadoras de serviços, a densidade populacional, a percentagem de utentes inscritos com 65 ou mais anos de idade, e a percentagem de utentes inscritos com diabetes – e efetuámos uma análise de *clusters* para classificar os prestadores em relação às suas condições ambientais. Como resultado desta análise de *clusters*, identificámos um pequeno grupo de prestadores que operam em condições menos favoráveis e um grande grupo de prestadores que operam em condições mais favoráveis. Depois, efetuámos uma avaliação da DEA, em que cada prestador foi comparado somente com os pares que operam nos mesmos ou em piores ambientes, como sugerido por Löber e Staat (2010).

Para alcançar o segundo objetivo deste estudo, propomos uma nova metodologia para a contratualização de P4P que liga os resultados da avaliação da DEA com as metas absolutas para as unidades que obtiveram um resultado inferior a 100% na avaliação da DEA e com um sistema de incentivos graduados para as unidades que obtiveram um resultado relativamente alto nesta avaliação. Também definimos uma

metodologia para a avaliação do alcançar absoluto das metas por parte das unidades que mostram um potencial para a melhoria de desempenho e desenvolvemos um sistema distinto de incentivos graduados associado a esta avaliação. Foi efetuada uma análise exploratória dos resultados atuais, das metas contratualizadas e das metas DEA sugeridas para algumas das unidades, sublinhando as vantagens da metodologia DEA proposta, quando comparada com a metodologia de avaliação atualmente implementada no programa P4P português. Também se sugere a implementação de um programa de *benchmarking* entre os prestadores com caraterísticas similares e prioridades de desempenho iguais, a fim de contribuir para a disseminação das melhores práticas.

Os estudos futuros são encorajados a explorar um leque mais vasto de indicadores relacionados com o desempenho dos prestadores de cuidados primários portugueses, a fim de permitir uma avaliação abrangente desses prestadores e também são encorajados a investigar a influência de outros fatores ambientais, como as caraterísticas socioeconómicas da população, na alcançabilidade das metas do P4P. Recomenda-se também a análise de possíveis efeitos não intencionais do programa P4P implementado. Além disso, propõe-se que um modelo de DEA baseado em indicadores de volume seja utilizado em futuras avaliações dos prestadores de cuidados primários portugueses, pois esse modelo tem em conta potenciais diferenças de recursos entre os prestadores.

Palavras-chaves: avaliação de desempenho, prestadores de cuidados primários, pagamento por desempenho, Análise Envoltória de Dados.

ABSTRACT

This thesis consists of two research articles. The first article addresses two research questions: (1) what have been the main performance dimensions and measures used for primary care assessment? And (2) to what extent have key stakeholders been involved in primary care assessment? To answer these questions, we developed a conceptual framework that includes three stages of performance assessment (measurement, analysis, and action) and captures four performance dimensions (accessibility, equity, efficiency, and effectiveness). We applied this framework for a systematic review of the 166 empirical articles devoted to performance assessment of primary care providers. The results of the review show that a wide variety of performance measures has been used to assess primary care providers, and also that the four dimensions have mostly been explored separately. Furthermore, the results of the review reveal an insufficient involvement of stakeholders in the performance assessment process, indicating the potentially limited impact of the reviewed research on health care policy and practice.

The second article aims two objectives: (1) to develop a DEA methodology for performance assessment of primary care providers in Portugal; and (2) to propose an alternative methodology for P4P contracting based on the DEA assessment results. To achieve the first objective, we modified the standard DEA model to account for the necessary relations between the weights attributed to each performance indicator and domain and to account for the effect of the relevant environmental variables. To achieve the second objective, we combined relative and absolute performance assessments in the elaboration of several bases for reward, and each of these two assessments is proposed to be linked to a system of graduated rewards. For underperforming units, we suggest establishing absolute targets derived from the relevant benchmarks, identified by DEA. We also recommend the creation of a benchmarking programme that will contribute to the dissemination of best practices among the providers.

Keywords: performance assessment, primary care providers, pay-for-performance, Data Envelopment Analysis.

CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ACES	Group of Health Centres (Agrupamento de Centros de Saúde)
ACSS	Central Health System Administration (Administração Central do Sistema de Saúde)
ANOVA	analysis of variance
ARS	Regional Health Administration (Administração Regional de Saúde)
BMI	Body Mass Index
CCR	Cooper, Charnes and Rhodes
CH index	Calinski-Harabasz index
CHC	community health centre
DB index	Davies-Bouldin index
DEA	Data Envelopment Analysis
DHA	District Health Authority
DMU	decision making unit
FTE	full-time equivalent
GDP	gross domestic product
GP	general practitioner
HbA1C	glycated haemoglobin
HIV	human immunodeficiency virus
HMO	health maintenance organisation
MI	Malmquist Index
IDG	Global Performance Index (Índice de Desempenho Global)
IE	within-group performance spread
IF	productivity gap between frontiers
NHS	National Health Service
OECD	Organisation for Economic Co-operation and Development

OPSS	Portuguese Health Systems' Observatory (Observatório Português dos Sistemas de Saúde)
P4P	pay-for-performance
PCAT-C	Primary Care Assessment Tool – Chinese version
PHCU	Primary Health Care Unit
PI	performance indicator
SFA	Stochastic Frontier Analysis
UCC	Comminity Health Unit (Unidade de Cuidados na Comunidade)
UCSP	Personal Health Unit (Unidade de Cuidados de Saúde Personalizados)
UK	United Kingdom
URAP	Shared Assistance Resource Unit (Unidade de Recursos Assistenciais Partilhados)
USA	United States of America
USF	Family Health Unit (Unidades de Saúde Familiar)
WHO	World Health Organisation

Chapter 1

INTRODUCTION

1.1 Introduction

In the context of global health care challenges, the development of primary care systems has been prioritised (WHO, 2008), while performance measurement in primary health care becomes one of the central research directions (Schäfer, Groenewegen, Hansen and Black 2011). This is due to the fact that primary health care has unique characteristics that may contribute to the improvement of the overall performance of the health systems worldwide (WHO, 2008). In particular, it has been recognised that a well-functioning primary care system has the potential to prevent the occurrence of common diseases up to 70% (WHO, 2008) and also to assist their early detection and management (e.g. Gorey, Luginaah, Holowaty, Fung and Hamm 2009; Campbell, Ramirez, Perez and Roetzheim, 2003; Roetzheim, Pal, Durme, Wathington, Ferrante, Gonzalez and Krischer, 2000). Primary care interventions, unlike specialist care services, are not usually limited to a particular disease or type of diseases, being more "generic" and "patient-oriented" rather than "disease-oriented". This property of primary care is particularly valuable in the context of an ageing population that struggles with increasing multiple morbidities (WHO, 2008).

Countries with strong primary care systems also report to have lower health care expenditure, possibly due to more effective disease prevention and lower hospitalisation rates (Starfield, Shi and Macinko, 2005; Starfield and Shi, 2002). This is a critical advantage of the health systems with an explicit focus on primary care, since most OECD countries are expected to spend more than 20% of their GDP on health care by 2050, if current trend in health care expenditure persists (Drouin, Hediger, Henke, Kanzler, Leydon and De Santis, 2008). To counteract the ongoing growth of health care costs, a series of fundamental health care reforms are urgently needed, including further

development of primary care systems that in many situations, as argued by Starfield et al. (2005), serve as a less expensive but equally effective alternative to specialist care.

Primary-care oriented health systems are also known to have fewer inequities in health service delivery (Starfield, 2009). For instance, it has been concluded that a higher supply of primary care physicians is associated with reductions in racial disparities related to referral-sensitive procedures (Basu and Clancy, 2001) and with decreased mortality among racial and ethnic minority groups (Shi, Macinko, Starfield, Politzer and Xu, 2005). The beneficial impact of primary health care is also confirmed in relation to the areas with high income inequality (Shi, Starfield, Kennedy and Kawachi, 1999).

In this context, the importance of performance measurement in primary health care should not be underestimated. In particular, performance measurement helps to set health policy goals and priorities and to design health care reforms; to ensure the providers are functioning efficiently and effectively and to control that money is being spent in line with taxpayers' expectations; to monitor whether appropriate health care services are accessible when needed and whether all patients are experiencing the same level of quality of health care delivery (Smith, Mossialos, Papanicolas and Leatherman 2009).

However, to date, there is no standardised way in which performance of primary care providers is measured due to the variation in the definition of "performance" of these providers. Furthermore, to our knowledge, there is no study that has explored the linkage between the empirical research on primary care assessments and health care policy and practice. In this respect, one of the broad aims of this thesis is to contribute to the understanding of the conceptual basis that underlies a comprehensive performance measurement of primary care providers and to explore the potential of the undertaken primary care assessments for improving health care policy and practice.

At a broader perspective, besides the above described objectives, performance assessment is crucial for promoting accountability relationships with multiple stakeholders, such as the government, health care managers and policy makers, taxpayers, purchaser organisations, providers, patients, and the general public (Smith et al., 2009). Within the context of these relationships, performance measurement is usually seen as a part of the management system, in which the embedded measurement

helps to trigger appropriate organisational responses and changes in behaviour (Smith et al., 2009; Smith, 1996). In order to provide a greater motivation for desired responses from the providers, the measurement of performance is frequently linked to incentive mechanisms, such as pay-for-performance (P4P) programmes. As pointed out by Cromwell, Trisolini, Pope, Mitchell and Greenwald (2011),

> "P4P is intended to bring incentives for improving quality of care directly into the payment system. By paying for specified standards of quality care, P4P may help equalize quality across different regions of the country and among different providers in the same region." (Cromwell et al., 2011: 12)

P4P progammes are now seen as a promising tool for improving quality of health care delivery. For this reason, methodological and practical issues related to the development and implementation of P4P programmes have attracted much attention in the research literature (e.g. Eijkenaar, 2011; Conrad, 2009). Nevertheless, there is still significant potential for improvement of P4P programme design, including the methodology for performance assessment and subsequent target setting among other aspects.

In particular, we believe that Data Envelopment Analysis (DEA), a non-parametric frontier technique for relative performance assessment proposed by Charnes, Rhodes and Cooper (1978), opens up new opportunities for P4P programme design in what concerns the methodology for performance assessment and subsequent target setting. The exploration of these opportunities presents the second broad aim of this thesis.

To date, the DEA applications in the context of P4P and incentive regulation have been scarce (e.g. Shwartz, Burgess and Zhu, 2016; Gök and Altındağ, 2015; Prior and Surroca, 2007; Agrell, Bogetoft and Tind, 2002) and, to our knowledge, none of the existing P4P programmes in health care has implemented DEA as a principal performance assessment methodology. However, if applied to P4P, DEA has a number of advantages, including the following:

1) DEA allows to assess multiple inputs (i.e. resources) and outputs (i.e. goods or services produced from the inputs) simultaneously, recognising the existence of trade-offs between different production factors, by contrast to a ratio analysis that is more traditionally used in P4P;

2) DEA suggests relative performance assessment, so that it stimulates continual performance improvement rather than encourages the achievement of determined targets;
3) DEA recognises the existence of providers' preferences in assigning relative importance to each of the explored performance indicators (PIs) by adopting a flexible weighting system;
4) DEA aims to show each provider in the best possible light by maximising its relative performance score under the incorporated model restrictions;
5) DEA can be effectively combined with value judgements and prior views regarding the relative importance of the explored PIs, as well as with the methods that allow to account for the effect of environmental factors (i.e. factors that are beyond the control of the providers);
6) DEA identifies realistic individual targets and benchmarks for each unit that shows the potential for performance improvement.

At the same time, DEA has some limitations, including the following:

1) DEA requires all units under assessment to be homogenous in order to produce realistic performance scores;
2) DEA has proved to be sensitive to the number of input and output variables in the presence of a small number of units, since a higher number of variables leads to a lower discriminating power of the technique;
3) DEA may provide biased results if data regarding inputs or outputs are missing for some providers;
4) DEA results are sensitive to model specification, since DEA assumes that the production process can be fully characterised by the inputs and outputs included into the analysis that in practice is not always possible.

We consider these limitations in our thesis and provide specific solutions and directions for future research in order to overcome them.

1.2 Research objectives

This thesis has four research objectives:

1. To explore the main performance dimensions and measures of primary care assessments by undertaking a systematic literature review that adopts a comprehensive conceptual framework for performance evaluation of primary care providers.

2. To examine the potential of primary care evaluations for the targeted improvement in health care policy and practice through the prism of stakeholders' involvement at three different stages of performance assessment– measurement, analysis, and action – as proposed by Smith (1996).

3. To develop a DEA-based methodological approach for performance assessment of primary care providers in Portugal that allows a fairer comparison of the providers in the context of P4P. This approach estimates the relative performance scores of the providers, considering the maximum possible performance level under the current technology identified for each provider individually, both providers' preferences and health care authorities' judgements regarding the relative importance of each PI, and the effect of environmental variables on P4P target attainment.

4. To propose an alternative methodology for P4P contracting with primary care providers in Portugal based on the DEA assessment results. This methodology allows us to effectively combine relative performance assessment with absolute performance assessment in the elaboration of several bases for reward for providers with different performance levels. Furthermore, the DEA assessment allows benchmarking programmes to be implemented between the providers with similar characteristics and strategic priorities in order to contribute to the dissemination of best practices among these providers.

1.3 Thesis outline

This introductory chapter provided an overview of the thesis context and set up research objectives. The remaining part of the thesis comprises two scientific papers, each of which is presented in the form of a chapter, followed by general conclusions.

In particularly, Chapter 2 investigates the conceptual basis for performance assessment of primary care providers and explores the potential of the undertaken primary care assessments for performance improvement in health care policy and practice. A conceptual framework for primary care assessments is developed. This framework is used for analysis of the empirical studies included into the systematic literature review to explore the main dimensions, sub-dimensions, and PIs of primary care assessments and to conclude about the level of stakeholders' involvement at different stages of organisational performance assessment. A functional classification of the explored PIs is also provided.

Chapter 3 explores the potential of DEA for enhancing performance assessment and subsequent target setting in the context of P4P. Firstly, we develop a DEA-based methodology for performance assessment of primary care providers in Portugal. In doing so, we modify the standard DEA model in order to account for the necessary relations between the relative weights attributed to each PI and each performance domain and in order to account for the effect of the relevant environmental factors. Secondly, we propose a new methodology for P4P contracting with the Portuguese primary care providers that links the DEA assessment results with absolute targets for the underperforming units and with a system of graduated rewards for the units that scored relatively high in this DEA assessment. We also set up a methodology for the assessment of absolute target attainment of the underperforming units and develop a distinct system of graduated rewards associated with this assessment. An exploratory analysis of the actual results, the contracted targets and the suggested DEA targets for some of the providers is performed, outlining the advantages of the proposed DEA methodology compared to the assessment methodology currently implemented in the Portuguese P4P programme.

Chapter 4 draws conclusions from the two papers and identifies the main suggestions for further research.

Chapter 2

ORGANISATIONAL PERFORMANCE ASSESSMENT IN PRIMARY HEALTH CARE: A SYSTEMATIC LITERATURE REVIEW AND A CONCEPTUAL FRAMEWORK

2.1 Introduction

Nowadays, more than ever, governments worldwide face increased pressure to ensure that their health systems deliver the best performance results. This is mainly due to the global demographic changes that are causing the average patient to be older and poorer in health (WHO, 2011), and also because of stringent resource constraints that contrast heavily with the persistent growth of health care expenditure (Drouin et al., 2008). Alerted by these tendencies, both researchers and policymakers have started to look for innovative solutions in organising the work of health systems in a more efficient, effective, equitable, and accessible manner. The search for new solutions has caused a fundamental shift in the health care research agenda by moving the focus away from hospitals and onto primary care providers (Schäfer et al., 2011). In particular, it has been recognised that a strong and well-established primary care system has great potential to cope with the challenges related to the quality of health care delivery. As a result, performance improvement in primary care provision has been acknowledged by several international entities, including the World Health Organisation (WHO) (2008), as one of the central priorities for future research.

Performance assessment in the primary care context plays a vital role. It facilitates the control of the performance results achieved by primary care organisations and provides valuable information for performance improvement. However, no single framework for this purpose exists due to the variety of views on what is meant by

performance in primary health care and how to measure it in the most appropriate way. Moreover, it is unclear whether empirical research in this area has been successfully linked to health care policy and practice. In order to provide further and more conclusive evidence on these issues, our study aims to present a critical and updated review of the empirical research devoted to performance assessment of primary care providers. In doing so, we aim to answer two important research questions:

1. What have been the main performance dimensions and measures used for primary care assessment? By exploring this research question, we aim to systematise the main aspects of primary care performance that are crucial for the development of a performance measurement system in this context.
2. To what extent have key stakeholders been involved in the performance assessment process in primary health care? By answering this research question, we aim to reveal whether the level of stakeholders' involvement has been sufficient to ensure a formative evaluation of primary care providers in order to encourage subsequent performance improvement.

To date, several authors have performed literature reviews in the context of performance assessment in primary health care. Amado and Dyson (2008) reviewed the methods and measures that have been developed to compare primary care providers with particular reference to the use of the non-parametric technique known as Data Envelopment Analysis. Pelone, Kringos, Romaniello, Archibugi, Salsiri and Ricciardi (2015) reviewed the studies that have used DEA to compare primary care providers in developed countries. Kringos, Boerma, Bourgueil, Cartier, Hasvold, Hutchinson, Lember, Oleszczyk, Pavlic, Svab, Tedeschi, Wilson, Windak, Dedeu and Wilm (2010a) and Kringos, Boerma, Hutchinson, van der Zee and Groenewegen (2010b) explored performance dimensions and indicators related to this research topic, limiting the analysis of the literature within a five-year period. Hollingsworth, Dawson and Maniadakis (1999) and Hollingsworth (2008, 2003) conducted reviews of non-parametric and parametric applications in health care, including primary health care, focusing solely on efficiency measurement.

Whilst these studies have improved our understanding regarding different aspects related to the assessment of primary care providers, our review extends and complements this literature in several important ways. Firstly, this review offers a

conceptual framework for primary care evaluation and a functional classification of indicators within each performance dimension discussed. These theoretical developments may have important benefits for researchers, policymakers, and health care managers who are searching for effective tools in the design of a balanced performance measurement system. In particular, the proposed framework and classification can help to reveal the relationships between major conceptual elements related to primary care assessment. They can also facilitate the choice of appropriate dimensions and corresponding measures sensitive to specific aspects of primary care delivery. Secondly, this review provides one of the most exhaustive analyses of the literature in this research area until now, since in this study we do not limit our attention to any particular performance dimension, analytical technique, or time period. This allows us to provide a complete view on the state of the art in the referred area and also to reflect on how to best capture the performance of primary care providers. Thirdly, this review explores the level of stakeholders' involvement in the process of primary care assessment, an issue that has not been thoroughly studied before. In doing so, we aim to conclude whether the previous studies' research findings have been (or can be) used formatively to guide subsequent improvement in the performance of primary care providers. This is an important piece of knowledge, since it is recognised that formative evaluations are fundamental to successful translation of research evidence into relevant changes in policy and practice (Amado and Dyson, 2008). In the health care context, this is vital to enhance the health-related quality of life of the population.

The remainder of this paper is organised as follows. Section 2 sets out the conceptual framework developed for performance assessment of primary care providers. This framework provides the basis for analysing the studies included in this literature review. Section 3 discusses the protocol used for literature search and selection of studies. Section 4 presents a synthesis of the reviewed studies, including a functional classification of the indicators used, and offers suggestions for future research. Section 5 concludes by systematising the main findings of the review and discussing its main limitations.

2.2 Conceptual framework

Performance assessment in health care is a complex undertaking that involves several stages. According to Smith (1996), who investigated the peculiarities of performance assessment in the public sector, these stages can be broadly identified as the measurement stage, the analysis stage, and the action stage. Particularly, the measurement stage aims to create a methodological basis for the whole assessment process. In the pursuit of this objective, the most relevant dimensions of performance must be selected and the indicators that allow performance assessment must be defined. In addition, the most appropriate method to measure performance must also be chosen. Then, at the analysis stage, the results regarding the chosen sets of indicators are interpreted to provide an idea about the level of performance achieved in relation to each of the defined dimensions. The relationships between the results observed in the different performance dimensions should also be analysed at this stage. Overall, the main objective of the analysis stage is to derive learning regarding the performance of the system. Finally, the action stage suggests appropriate organisational responses to the results of the analysis of performance data, including the discussion and elaboration of plans, implementation of actions, and monitoring towards performance improvement among other activities. Given that performance assessment is a continuous process, the action stage also offers opportunities to reflect on the objectives of the system in order to return to the measurement stage.

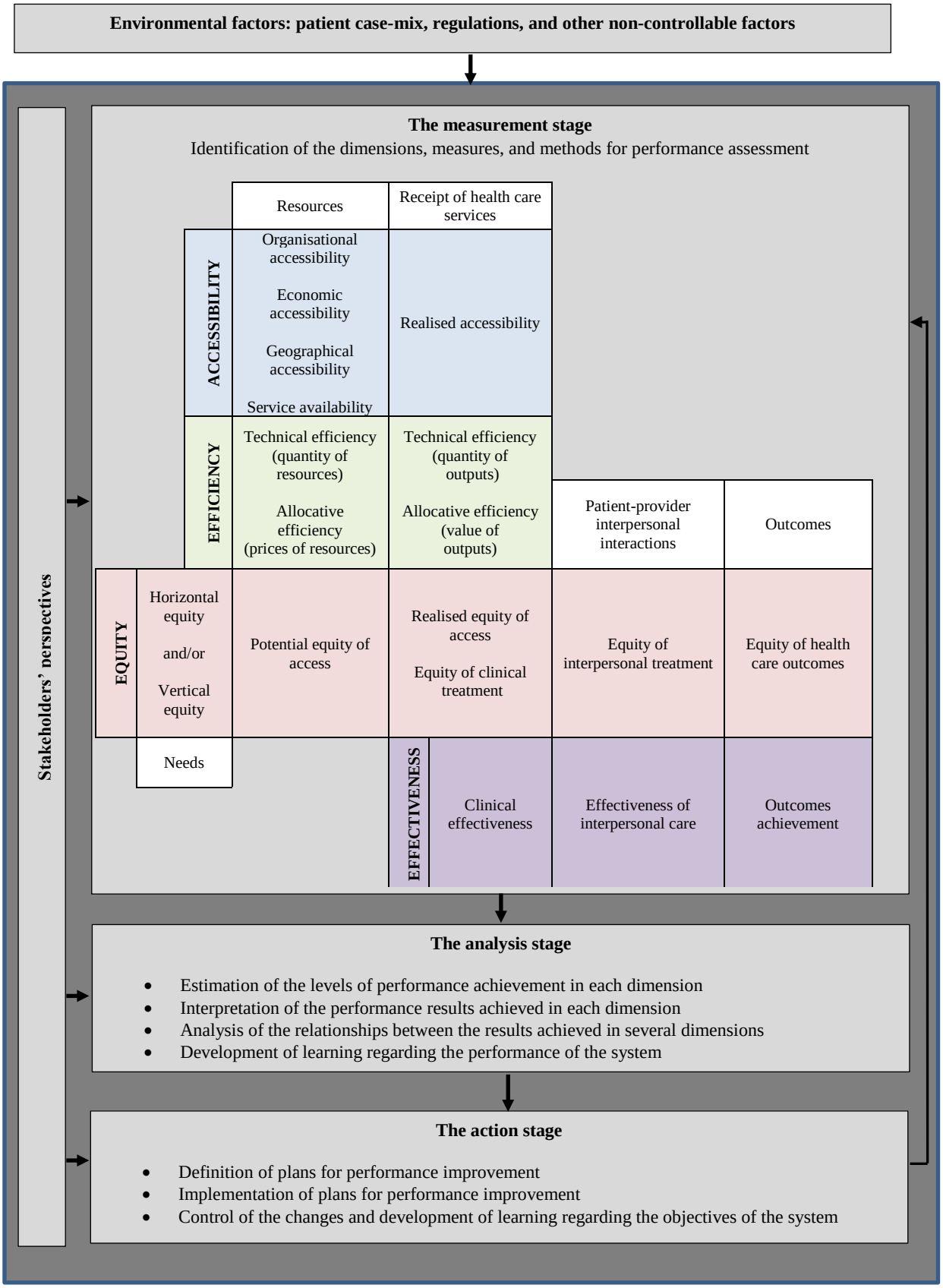
As we mentioned earlier, the development of a conceptual basis is fundamental to the measurement stage. The conceptual basis for health care assessment has been extensively explored and one of the most significant contributions in this respect belongs to Donabedian (1980). In particular, Donabedian (1980) proposed a framework for quality assessment consisting of three key elements: structure, process, and outcome. Structure refers to the organisational characteristics of health care providers, including human, physical, and financial resources and tools used in the delivery of health care services, representing the inputs in health care provision. Process refers to the activities that go on within and between health care practitioners and patients, focusing on conformity to technical and ethical norms of good care. And, finally, outcome refers to the impact of these activities on a patient's current and future health status. Most of the subsequently developed conceptual models in this area include the

components of the framework developed by Donabedian (1980). Among other models, Maxwell (1992) described how six dimensions of health care quality – effectiveness, acceptability, efficiency, access, equity, and relevance – can be assessed in relation to structure, process, and outcomes. Starfield (1998), in turn, proposed a model based on four structural elements and two process features to measure the most important attributes of quality in primary health care: first-contact care, longitudinality, comprehensiveness, and coordination. Afterwards, Campbell, Roland and Buetow (2000) analysed the most widely used approaches to measuring the quality of health care delivery, including the models developed by Maxwell (1992) and Starfield (1998), and concluded that a comprehensive health care assessment model incorporates the measurement of four performance dimensions: accessibility, equity, efficiency, and effectiveness.

Inspired by the works of Smith (1996) and Campbell et al. (2000), we propose a conceptual framework for performance assessment in primary health care. This framework involves the three stages of performance assessment (measurement, analysis, and action) and captures the four basic dimensions of performance (accessibility, equity, efficiency, and effectiveness). Figure 2.1 presents this conceptual framework.

There are two important aspects to be emphasised about this framework. Firstly, it recognises the influence of environmental factors in the performance assessment process. Control of these environmental factors can be made through, for example, case-mix adjustment. This type of adjustment aims to account for variations in the observed performance results of the providers that are caused by intrinsic patient characteristics (Iezzoni, 2009). Beyond patient attributes, there are other non-controllable factors that may impact the performance of primary care providers, such as, for example, health care regulations and the performance of other providers of health care services. Secondly, the framework acknowledges that performance has different meanings across multiple groups of stakeholders. In this respect, it is of paramount importance to identify the relevant stakeholders, whose objectives will underlie the whole performance assessment process. Furthermore, as defended by several authors (e.g. Amado and Dyson, 2008; Smith, 1996), it is essential to involve the key stakeholders in the three stages of the process. This involvement is critical for the performance measurement system to generate results that act as a trigger for discussion among

Figure 2.1 Conceptual framework for performance assessment of primary care providers



targeted stakeholder groups, fostering learning and promoting performance improvement. Stakeholders' involvement is also essential to guarantee face validity of the assessment results (that is, results that are recognised as valid by the stakeholders) and to contribute to subsequent implementation of the developed improvement strategies and plans.

In what concerns the measurement stage, our framework relates the four performance dimensions to five basic elements of primary health care delivery (needs, resources, receipt of services, patient-provider interpersonal interactions, and outcomes). Each of the performance dimensions has been divided into several sub-dimensions. Accessibility can take the form of organisational accessibility, economic accessibility, geographical accessibility, service availability, and realised accessibility. Equity can be analysed from two perspectives: horizontal equity and vertical equity. In each of these perspectives, equity can assume three forms: equity of access (potential and realised), equity of treatment (including both clinical treatment and interpersonal treatment), and equity of outcomes. Efficiency can take the form of technical efficiency and allocative efficiency. Finally, effectiveness can take the form of clinical effectiveness, effectiveness of interpersonal care, and outcomes achievement.

In order to clarify the meaning of the terms used in the conceptual framework, below we provide a discussion regarding each of the four performance dimensions and of their sub-dimensions.

Accessibility in the context of primary health care may have different interpretations. One of the most common definitions suggests that accessibility measures the possibility of potential or actual entry of an individual into the health care system (Andersen, McCutcheon, Aday, Chiu and Bell, 1983). Therefore, an accessible health care system removes the barriers to the initiation and maintenance of relationships between the individual and the provider (Gulliford, Figueroa-Munoz, Morgan, Hughes, Gibson, Beech and Hudson, 2002). Considering the existence of various barriers to access, Campbell et al. (2000) proposed the measurement of accessibility in terms of geographical (or physical) access, service and provider availability, and the affordability of health care expressed in monetary costs. We find this definition useful for analysing the accessibility dimension and thus we identified four corresponding sub-dimensions: organisational accessibility, economic accessibility,

geographical accessibility, and service availability. In particular, organisational accessibility measures the extent to which the organisation of health care delivery corresponds to patients' needs. Economic accessibility evaluates the acceptability of costs paid by the users of health care services. Geographical accessibility concerns the efforts needed to physically access a health care provider. Finally, service availability reflects whether a range of available services is adequate to satisfy patients' needs. Following the distinction between potential and realised access, first discussed by Andersen and Aday (1978), we decided to introduce an additional fifth sub-dimension entitled realised accessibility. This sub-dimension includes the performance indicators that show whether an individual or local population has succeeded in gaining access to primary care services.

Regarding equity, there are two different perspectives that underlie its definition. In particular, horizontal equity requires the same treatment of individuals with identical health care needs, regardless of their individual characteristics. Vertical equity supports the idea of preferential treatment of those who have greater health care needs. Given the existence of these two distinct ideological perspectives, it is essential to choose one of them for the evaluation of equity. Furthermore, the evaluation of equity also requires the concept of need to be defined. In the health care context, need has been most commonly viewed as the capacity to benefit from the consumption of health care (Allin, Hernández-Quevedo and Masseria, 2009).

Concerning the equity sub-dimensions, there are equity of access, equity of treatment, and equity of health care outcomes (Whitehead, 1990; Goddard and Smith, 2001). In the following discussion we will define the concepts using the perspective of horizontal equity, considering that this is the most commonly studied perspective. However, the definitions of the three forms of equity can be easily adjusted to refer to the perspective of vertical equity.

Equity of access implies equal entitlement to health care services for everyone, a fair distribution of resources, and the removal of all barriers to access (Whitehead, 1990). Equity of treatment, in turn, is frequently viewed as the absence of restricted use of health care services caused by any kind of patient disadvantage (Whitehead, 1990). However, we use this term in a broader sense to cover both clinical and interpersonal aspects in the delivery of health care, as proposed by Donabedian (1988). Finally,

equity of health care outcomes implies the same high standard in relation to outcomes achievement for everyone.

Efficiency is one of the four dimensions captured by the framework. The achievement of an efficient health care provision is targeted by most health care systems worldwide, as it ensures a balanced distribution of available resources among and within the systems. For the purpose of efficiency analysis, the economic literature considers that health care providers consume various inputs (human resources, capital, etc.) in order to produce valued outputs (consultations, home visits, etc.). The conceptual literature also distinguishes between several types of efficiency. In particular, the achievement of technical efficiency means that an increase in any output or input is not feasible without decreasing some other output(s) or input(s) (Cooper Seiford and Zhu, 2011). When price data for inputs are available, we can estimate the allocative efficiency of providers, which reflects the ability to mix resources in optimal proportions, given current input prices (Farrell, 1957). The achievement of both technical and allocative efficiency results in cost-efficiency, which reflects the ability to deliver a certain quantity of outputs at minimum cost. If only information on cost indicators is available, one may estimate cost-efficiency without being able to disentangle between technical and allocative efficiency.

Regarding efficiency measurement, it is common to evaluate efficiency in relative terms, by comparing the performance of several providers. One may compare the relative efficiency of providers using a ratio of a single output over a single input. In this case, this tends to result in a partial measure of efficiency. In order to estimate a comprehensive measure of efficiency, we must use a technique that allows the consideration of multiple inputs and outputs.

Stochastic Frontier Analysis (SFA) and Data Envelopment Analysis (DEA) are the two most commonly used analytic techniques for measuring relative efficiency in health care. In particular, SFA is a parametric technique that estimates the cost (or production) frontier based on a prior specification of its functional form. DEA, in turn, is a non-parametric technique that constructs a frontier by a linear combination of the best practice providers that “envelop” underperforming providers. In addition, as opposed to SFA, DEA does not require a priori assumptions with regards to the weights attached to input and output variables. The flexible weighting system adopted in DEA

aims to represent the provider under evaluation in the best possible light. An added advantage of DEA, when compared with SFA, is that it allows the simultaneous inclusion of multiple inputs and multiple outputs. A detailed description of SFA and DEA can be found elsewhere (Cooper et al., 2011; Kumbhakar and Lovell, 2003).

The fourth dimension included in our framework is effectiveness. According to Campbell et al. (2000), effectiveness in health care provision is defined as "the extent to which care delivers its intended outcome or results in a desired process, in response to need" (Campbell et al., 2000: 1616). This definition implies that effectiveness can be measured by process assessment and/or by outcomes assessment.

When a process assessment is undertaken, it is assumed that certain characteristics of health care delivery, measured by process indicators, are directly linked to targeted health care outcomes. Process assessment also distinguishes between the clinical effectiveness and effectiveness of interpersonal care that are associated with two key elements of the health care process – clinical management and management of interpersonal care (Donabedian, 1988). In particular, clinical effectiveness focuses on conformity to evidence-based practice or any health care standards and norms that are considered to be legitimate (Campbell et al., 2000). Effectiveness of interpersonal care reflects the quality of interpersonal interactions between the provider and the patient that is considered to have influence on patient health status and patient satisfaction (Beaulieu, Haggerty, Beaulieu, Bouharaoui, Lévesque, Pineault, Burge and Santor, 2011).

When an outcomes assessment is undertaken, data regarding the impact of health care delivery is collected (e.g. indicators regarding the improvement in health care status). Several authors have argued that outcomes assessment should play a central role in the investigation of effectiveness in primary care delivery (e.g. Amado and Dyson, 2008). However, the challenges involved in accurately measuring outcomes achievement in the health care context have also been acknowledged. The major challenges include the difficulty in attributing outcomes to specific actions carried out by health care providers, the existence of considerable time lags between certain health care processes and outcomes, and the interference of environmental factors (Lester and Roland, 2009). In this respect, it is essential that outcomes measurement accounts for all these aspects.

Having presented the conceptual framework that served as a basis for the analysis of the reviewed studies, we will now discuss the protocol used for the systematic review of the literature.

2.3 Methods

The article search and selection process adopted in this study followed the protocol of a systematic literature review. As stated by Petticrew and Roberts (2006), systematic reviews aim to provide a comprehensive and reliable overview of the evidence with regards to the specified research questions by following strict scientific protocols in the selection and analysis of the studies. For the article search we initially selected six keywords, namely, "performance", "*efficiency", "productivity", "*equity", "accessibility", and "*effectiveness" (an asterisk is used for searching related terms, e.g. "cost-efficiency" and "inefficiency"). These keywords were then successively combined with each of the following sets of terms: "measur*", "manag*", "assess*", "evaluat*", "Data Envelopment Analysis", "Stochastic Frontier Analysis" (an asterisk is used for searching a noun and a verb, e.g. "measurement" and "measure") and with "primary care", "primary health care", "health centre*", "health center*", "health care centre*", and "health care center*" (an asterisk is used for searching plural forms of the keywords) (Figure 2.2).

We searched for these keyword combinations in the title, abstract, and subject terms of the papers published between January 1968 and December 2014 and indexed in the EBSCOhost electronic databases, including *Academic Search Complete*, *CINAHL Plus*, and *Medline.*

As shown in Figure 2.3, we found 7527 papers that met the search criteria. After screening the titles, 2040 duplicates were removed, resulting in 5487 potentially relevant papers for further analysis. We then read the abstracts of these articles and excluded from further review those papers that did not meet the inclusion criteria indicated in Table 2.1. This resulted in 340 papers being selected for full-text screening. After reading the full text, we excluded those papers that did not meet the inclusion criteria. As a result, from the initial 7527 entries derived from the electronic database

search 130 papers were included in our review. Thirty-six additional papers were retrieved by means of reference mining. In total, 166 studies were included in this literature review.

Figure 2.2 Search criteria tree

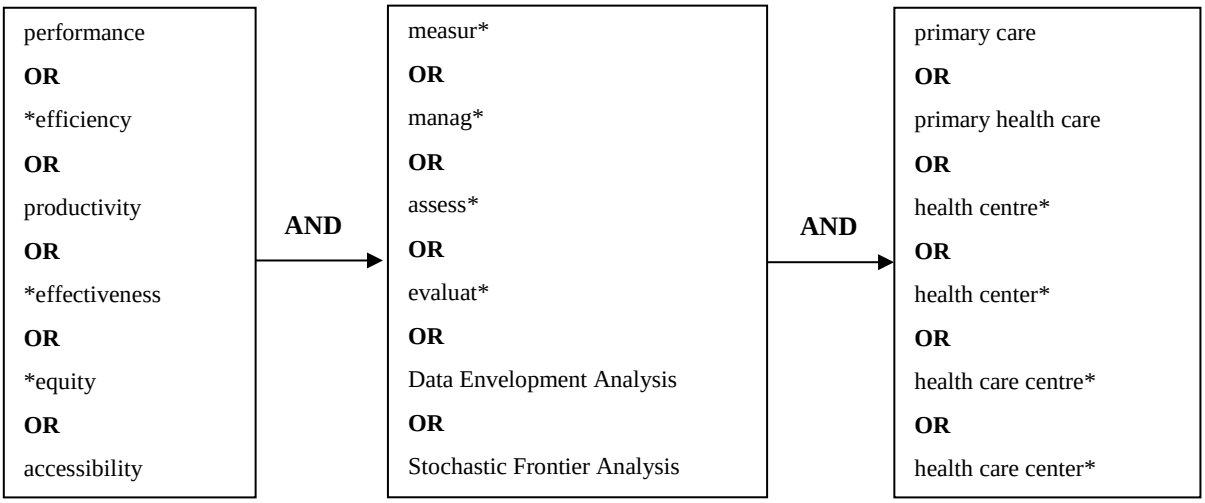


Figure 2.3 Flow chart of the systematic paper search and selection

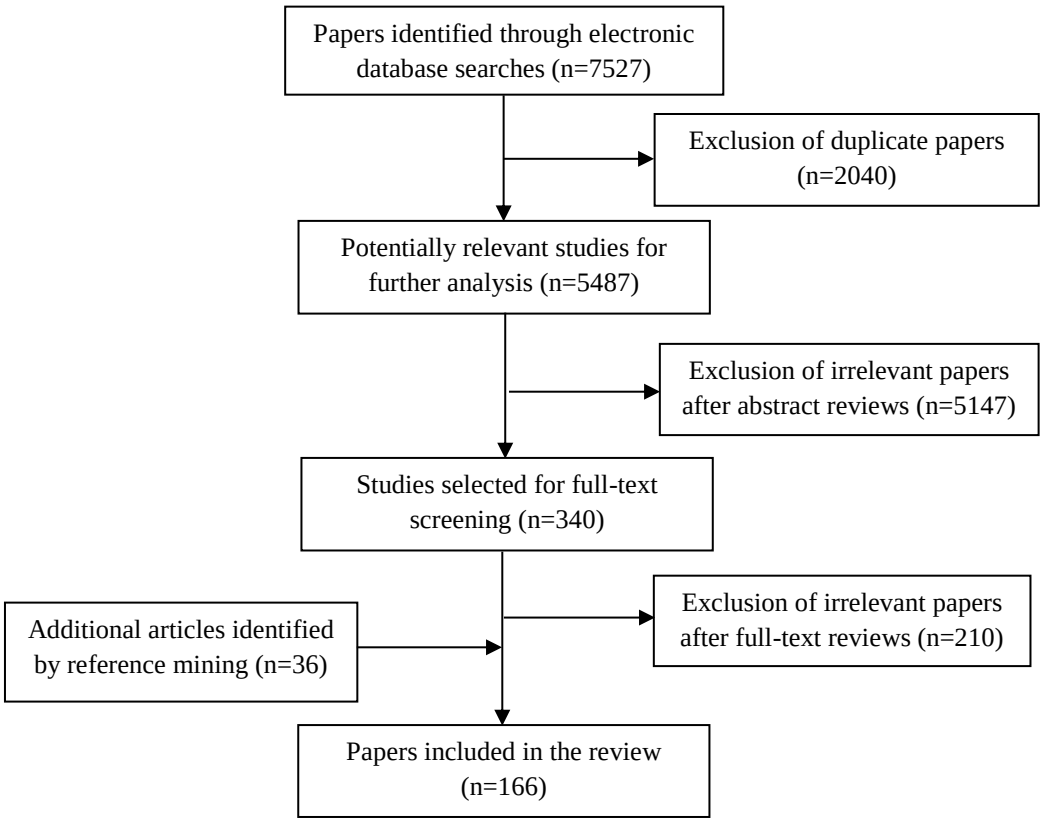


Table 2.1 Inclusion and exclusion criteria

Study characteristic	Inclusion criteria	Exclusion criteria
Study objective	Studies that aim to assess the performance of primary health care organisations	Studies that focus solely on performance assessment in secondary or tertiary health care Studies that report on national health strategy assessment; information health system evaluation; staff knowledge, attitudes, satisfaction, and workforce development; diagnostic and treatment innovations within the context of primary health care
Study population	The units of analysis include primary health care organisations of any type and dimension; they might be aggregated at a higher level (e.g. municipalities or district authorities) or decomposed into constituents (e.g. specific services provided, individual practitioners within a general practice)	The units of analysis only include hospitals and/or specialty care providers
Study design	Empirical studies with an application using real data	Observational and experimental studies reporting findings on the evaluation of medical interventions and programmes; validation studies that aim to test questionnaires or performance measurement systems without applying real data; theoretical and discussion papers, editorials, literature reviews, policy documents
Study publication	Peer-reviewed papers written in English and published between January 1968 and December 2014	Papers not written in English; papers not peer-reviewed

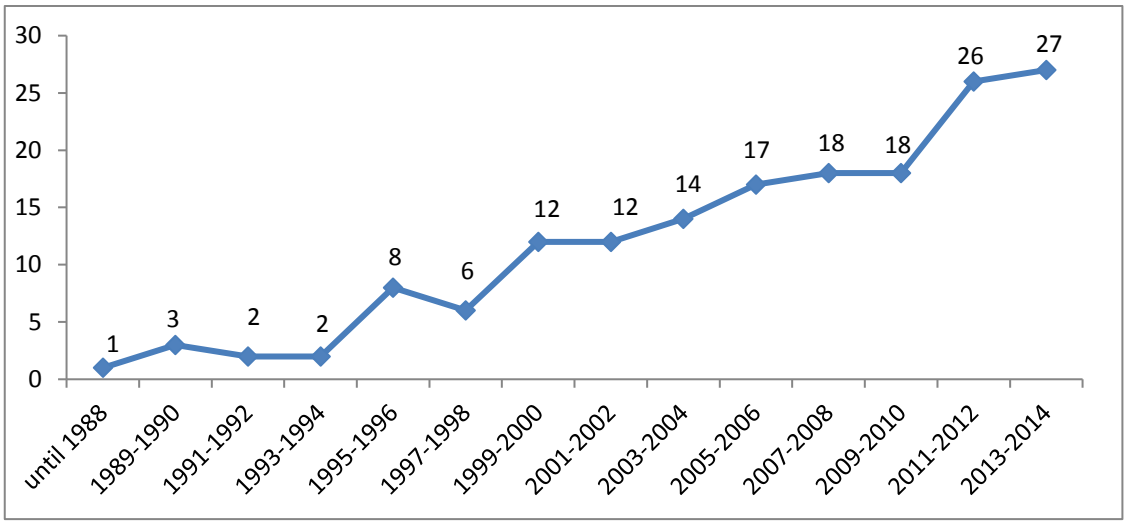
2.4 Results

2.4.1. Characteristics of the studies

Before discussing the results of the literature review, we will provide summary statistics of the sample of the articles selected for the study. A detailed description of each study reviewed is presented in Table A1 in Appendix A.

Figure 2.4 shows the distribution of the papers according to the year of publication. As can be observed, only four papers of the sample were published prior to 1990, which can be partly explained by the widespread political unacceptance of any kind of organisational performance assessment in health care until the nineties (Lester and Roland, 2009). After 1990, we can distinguish three chronological periods based on the average number of publications per year: from 1991 to 1998 (on average 2 papers published per year), from 1999 to 2010 (on average 8 papers published per year), and from 2011 to 2014 (on average 13 papers published per year). This generally increasing trend indicates the growing relevance of the research topic over the recent years, an issue that was also brought into sharp focus by one of the WHO reports (WHO, 2008).

Figure 2.4 Number of selected papers by year of publication



As for the geographical distribution of the studies, the 166 papers explored the performance of primary care providers from 42 countries across the world. As shown in Figure 2.5, more than half of the studies (96 studies) were conducted in the United States of America (USA), Canada, and the United Kingdom (UK), three countries that have highly developed health care systems. At the same time, we also found a significant number of papers focusing on primary care assessments in developing countries of Africa and Asia.

Finally, we looked at the medical conditions explored in the papers. Of the 166 articles analysed, 123 papers examined primary health care in general, without focusing on a particular disease or by studying the management of different medical conditions. Fourteen studies of the sample were devoted to the assessment of maternal and child care. Eleven studies focused on chronic disease management (diabetes, asthma, hypertension, coronary heart disease, and congestive heart failure). The remaining 18 papers were devoted to the examination of dental care (7 studies), acute/non-acute infectious disease management (sinusitis, otitis, tuberculosis, HIV) (6 studies), mental health care (3 studies), and preventive care (diagnostic screening and vaccination) (2 studies).

2.4.2 Performance dimensions and indicators used in primary care assessment

2.4.2.1 Accessibility

From the 166 reviewed papers, there are 59 that focused on the measurement of accessibility. As shown in Table 2.2, 50 studies have measured organisational accessibility. Access via telephone/ease of contact, ability to get an appointment when needed/waiting time, and working hours/out-of-hours service were the most commonly explored provider characteristics (e.g. Llanwarne, Abel, Elliott, Paddison, Lyratzopoulos, Campbell and Roland, 2013; Tourigny, Aubin, Haggerty, Bonin, Morin, Reinharz, Leduc, St-Pierre, Houle, Giguère, Benounissa and Carmichael, 2010; Montgomery, Irish, Wilson, Chang, Li, Rogers and Safran, 2004). In essence, these measures aim to reflect when and how a health care provider can be accessed. Staff

Figure 2.5 Number of selected papers by country studied

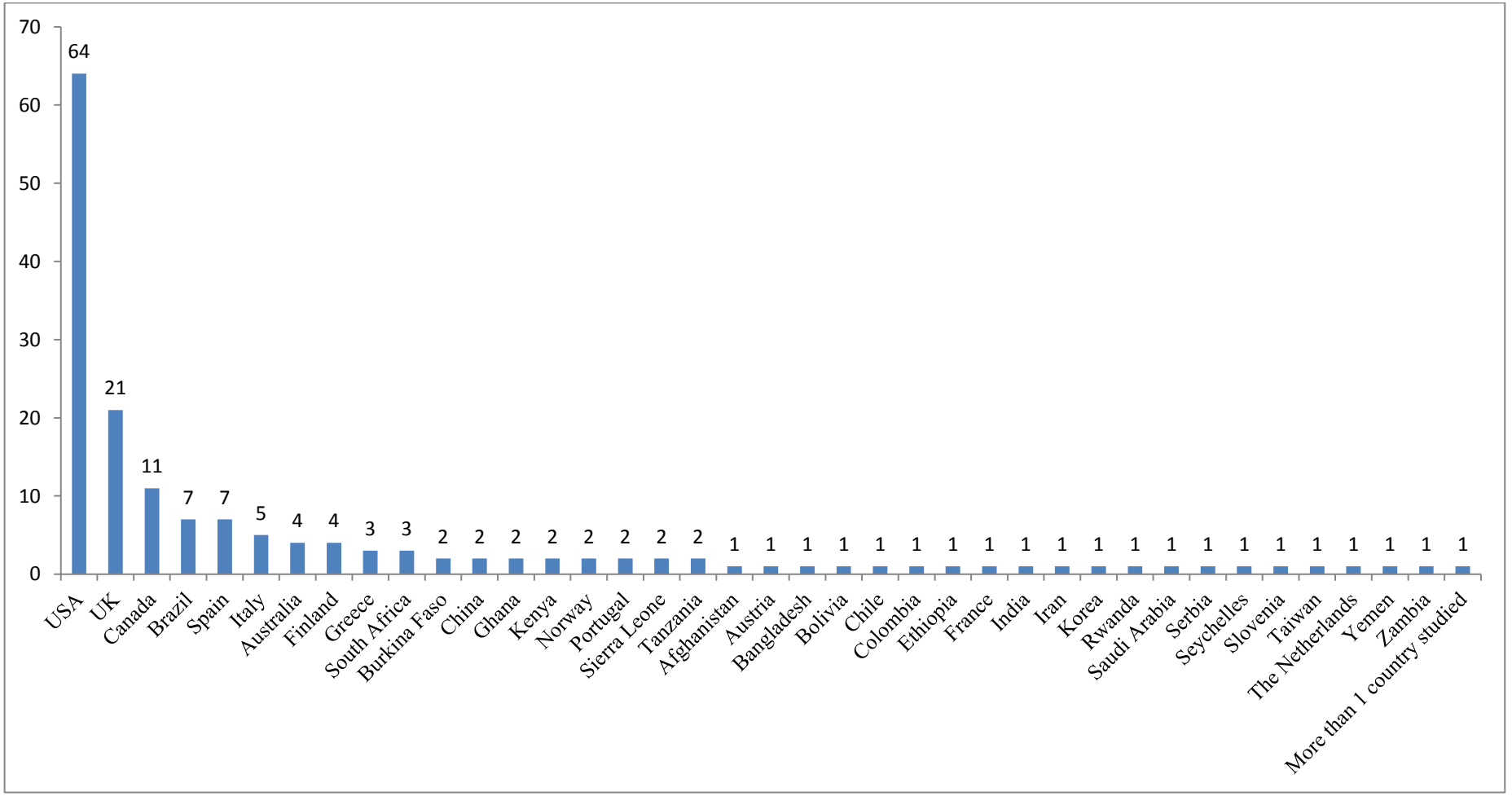


Table 2.2 Classification of accessibility measures

Sub-dimension	Classification of PIs	Aakvik & Holmås (2006)	Al-Taiar et al. (2010)	Arakawa et al. (2011)	Aysola et al. (2013)	Baker (1992)	Barbaro et al. (2014)	Bell et al. (2013)	Bissonnette et al. (2012)	Breyer (1977)	Edward et al. (2011)	Geboers et al. (2002)	Gilson et al. (1995)	Giuffrida (1999)	Giuffrida & Gravelle (2001)	Guagliardo et al. (2004)	Furtado et al. (2013)	Haggerty et al. (2007)	Haggerty et al. (2008)	Hall et al. (2008)	Heard et al. (2013)	Huerta Munoz & Källestål (2012)	Jacobs et al. (2006a)	Lee et al. (2009)	Lewis & Holcomb (2012)	Llanwarne et al. (2013)	Luo & Wang (2003)	Luo et al. (2004)	Luo & Qi (2009)	Macinko et al. (2007)	Manzoli et al. (2014)
Economic accessibility	Insurance coverage among patients																														
	Payment requirements and policies										X													X						X	
	Additional direct costs			X																											
	Opportunity costs			X																										X	
Geographical accessibility	Convenience of provider location			X														X													
	Transportation to/from medical consultations			X																											
	Travel distance/time to provider		X					X	X							X				X		X					X	X	X		
Organisational accessibility	Access via telephone/ease of contact					X	X										X	X	X	X			X	X		X				X	
	Adequacy and availability of premises, equipment, and medication			X								X	X	X	X						X									X	
	Ability to get an appointment when needed/ waiting time			X			X										X	X	X	X			X			X				X	
	Existence of language barriers							X	X																						
	Staff availability	X				X		X	X				X	X	X	X				X		X					X	X	X	X	
	Working hours/out-of-hours service					X	X			X		X					X	X	X	X						X				X	
Service availability (comprehensiveness)	Range of health care services					X	X			X		X	X				X	X			X			X	X					X	
	Range of social support services			X																											
Realised accessibility	Presence of unmet health care needs				X																										
	Rate of avoidable hospital admissions																														X
	Utilisation rate of health care services			X	X						X										X				X						

Table 2.2 Classification of accessibility measures *(continued)*

Sub-dimension	Classification of PIs	McGrail & Humphreys (2009)	McGrail (2012)	Mills et al. (2004)	Montgomery et al. (2004)	Mosquera et al. (2013)	Parchman & Culler (1994)	Pelone et al. (2012)	Perry & Gesler (2000)	Puig-Junoy & Ortún (2004)	Radford et al. (2007)	Ricketts et al. (2001)	Safran et al. (1998)	Safran et al. (2002)	Salinas-Jiménez & Smith (1996)	Seid & Stevens (2005)	Sequist et al. (2008)	Shi et al. (2003)	Stevens et al. (2014)	Teach et al. (2006)	Tilley & Chalkley (2005)	Tourigny et al. (2010)	Tsai et al. (2010)	Utriainen & Widström (1990)	Varatharajan et al. (2004)	Villiers et al. (2005)	Wang et al. (2014b)	Weiss et al. (2001)	Wensing et al. (2008)	Wong et al. (2010)	Frequency
Economic accessibility	Insurance coverage among patients										X					X												X			3
	Payment requirements and policies				X						X		X	X														X			8
	Additional direct costs																														1
	Opportunity costs																														2
Geographical accessibility	Convenience of provider location												X	X								X									5
	Transportation to/from medical consultations																											X			2
	Travel distance/time to provider	X	X						X																						12
Organisational accessibility	Access via telephone/ease of contact				X	X				X			X	X			X	X	X			X	X				X		X	X	23
	Adequacy and availability of premises, equipment, and medication			X											X											X	X				11
	Ability to get an appointment when needed/waiting time			X	X	X							X	X			X	X	X	X		X	X				X	X	X	X	24
	Existence of language barriers																	X	X				X					X			6
	Staff availability	X	X									X			X	X				X						X					21
	Working hours/out-of-hours service			X		X	X	X		X			X	X			X	X	X			X	X		X			X			24
Service availability (comprehensiveness)	Range of health care services			X		X												X	X				X				X	X		X	19
	Range of social support services																	X									X	X		X	5
Realised accessibility	Presence of unmet health care needs															X															2
	Rate of avoidable hospital admissions						X	X				X																			4
	Utilisation rate of health care services										X										X			X	X			X			10

availability, another parameter of organisational accessibility, was measured by population-to-provider ratios (e.g. Teach, Guagliardo, Crain, McCarter, Quint, Shao and Joseph, 2006), the presence of staff members at health facilities (e.g. Macinko, Almeida and de Sá, 2007), and the number of providers accepting new patients (e.g. Hall, Lemak, Steingraber and Schaffer, 2008). Less representative in this group were indicators related to the adequacy and availability of physical resources (facilities, equipment, and medication) (e.g. Geboers, Mokkink, van Montfort, van den Hoogen, van den Bosch and Grol 2002) and related to the existence of language barriers (e.g. Weiss, Haslanger and Cantor, 2001).

Service availability is another component of accessibility that has been frequently discussed in the reviewed literature. More specifically, it has been analysed in 20 reviewed papers. The importance of this sub-dimension and the reason why service availability is frequently seen as a part of accessibility derives from the assumption that a large supply of services makes the provider more accessible to people with different health care needs. In a number of studies, service availability was also used as an equivalent to comprehensiveness of health care (e.g. Tsai, Shi, Yu, Hung and Lebrun, 2010; Breyer, 1977). According to these studies, a comprehensive health care system provides a range of services that is wide enough to meet the common needs of the population.

Geographical accessibility has, in turn, been explored in 18 papers. This performance sub-dimension has frequently been measured in combination with organisational accessibility, given that a number of the reviewed studies used composite indicators that account for provider availability and travel distance/time to provider (e.g. Bell, Wilson, Bissonnette and Shah, 2013; Bissonette, Wilson, Bell and Shah, 2012; Luo and Wang, 2003). According to McGrail and Humphreys (2009), this approach allows researchers to extend the use of simple population-to-provider ratios to more sophisticated accessibility measures. Specifically, these advanced measures estimate provider availability within the population catchment areas defined by provider proximity to its potential users rather than on the basis of administrative division of the studied territories. However, the application of this approach becomes limited when individuals are not eligible to select primary care providers outside their residential areas, particularly, if these providers are practicing in the public sector.

A distinct subset of indicators has been used in 15 papers out of the 166 papers reviewed. These indicators aim to measure realised accessibility and they provide evidence of the extent to which access to primary care services has actually been realised. For example, Tilley and Chalkley (2005) estimated the number of patients that have received any kind of treatment in the UK General Dental Service within a six-year period as a measure of realised accessibility. This measure, along with other similar ones, such as the number of home visits (Arakawa, Arcêncio, Scatolin, Scatena, Ruffino-Netto and Scatena Villa, 2011; Varatharajan, Thankappan and Jayapalan, 2004) and the average number of visits per period (Lewis and Holcomb, 2012; Edward, Kumar, Kakar, Salehi, Burnham and Peters, 2011; Weiss et al., 2001), represent the utilisation rates of health care services and have commonly been used to evaluate realised accessibility. Other examples of indicators related to the measurement of realised accessibility include the presence of unmet health care needs (Aysola, Bitton, Zaslavsky and Ayanian, 2013) and the rate of avoidable hospital admissions (e.g. Manzoli, Flacco, De Vito, Arcà, Carle., Capasso, Marzuillo, Muraglia, Samani and Villari, 2014). Both indicators provide evidence of patients not having accessed a primary care provider despite needing to. In analysing unmet health care needs, this is done by direct investigation of unsatisfied demand in the receipt of health care services. The high rate of avoidable hospital admissions is believed to reflect poor access to high-quality primary care for patients with ambulatory sensitive conditions, given that these conditions could effectively be managed in the outpatient setting. However, the relationship between performance in primary care and hospital care should be explored further, as suggested by the results of some studies (e.g. Dusheiko, Gravelle, Martin, Rice and Smith, 2011).

Concerning economic accessibility, it mainly matters for those countries where the delivery of health care is not free at the point of use. We believe that this is the reason why its assessment is not very common among the reviewed studies. Specifically, only 10 papers have included this sub-dimension in their analysis. Without universal health coverage, the access to health care is secured by public and private insurance programmes, such as in the case of the USA. In this situation uninsured and low-income populations may experience difficulties in accessing health care systems, and the implementation of special payment requirements and policies aims to mitigate this problem (e.g. Radford, Pink and Ricketts, 2007; Montgomery et al., 2004). The

study of the influence of payer mix among patients also contributes to the understanding of economic accessibility among individuals covered by different health care plans and the uninsured population (e.g. Weiss et al., 2001). However, it is important to recognise that financial barriers may exist even if health care delivery is free of charge. These include costs directly associated with the access to a health care system (e.g. transportation expenses to arrive at a health care facility (Arakawa et al., 2011)) and opportunity costs due to losses in earnings (e.g. working days missed due to medical consultations (Arakawa et al., 2011; Macinko et al., 2007)).

To synthesise, the literature dealing with the measurement of accessibility to primary care services allows us to draw several conclusions. Firstly, it suggests that the measurement of accessibility can focus on the potential and/or on the realised access to health care providers. The former concept assesses the level of potential accessibility to services and corresponds to the prospective approach to measuring accessibility. The latter concept focuses on the actual use (or non-use) of health care services rather than on the possibility of such use, and it corresponds to the retrospective approach to measuring access to services. Secondly, although realised accessibility is highly correlated with potential accessibility, this relationship has not been captured by the indicators proposed in the reviewed studies. The conceptual economic literature suggests that potential access to a provider will be realised as long as the expected benefits from the consumption of health care services exceed the cost and time investments to be incurred by contact with the health system. This means that the decision to access a health care provider largely depends on the patients' personal valuations and priorities. In the particular case of primary health care, it is hypothesised that the influence of these factors may result in routine check-ups, preventive services, and health education sessions being placed in a less favourable position, given that these kinds of services do not normally produce immediate health improvements. However, this hypothesis should be carefully verified in future empirical research.

2.4.2.2 Equity

Equity has been assessed in 33 studies of the 166 studies included in the review. Concerning the equity perspective, most reviewed studies have concentrated on the measurement of horizontal equity. We believe that this can be explained by the fact that

the assessment of vertical equity in health care is difficult to perform accurately (Sutton, 2002). In our review only one out of the 33 studies that assessed equity has attempted to measure vertical equity. Specifically, Dahrouge, Hogg, Tuna, Russell, Devlin, Tugwell and Kristjansson (2010) explored vertical inequities in primary health care delivery for male and female patients in Canada. However, since the authors have not specified their judgements about the extent to which health care use should vary between men and women, the results of the vertical equity evaluation in this study should be treated with caution. Below we discuss each of the equity sub-dimensions analysed in the reviewed studies.

Equity of access has been analysed in 23 studies. As one can see by comparing Table 2.3 with Table 2.2, equity of access shares identical groups of indicators with accessibility. However, the main distinction of an equity analysis from an accessibility analysis is that the former considers the needs of specific patient groups rather than all individuals at once.

Similarly to the analysis of accessibility, the organisational characteristics of primary care providers has attracted the most attention in the study of equity of access, being explored in 17 studies (e.g. Toda, Opwora, Waweru, Noor, Edwards, Fegan, Molyneux and Goodman, 2012; Gesler, Sherman, Spencer, Preisser, Arcury and Perin, 2006). In particular, staff availability was the most explored provider characteristic, followed by the ability to get an appointment when needed/waiting time, working hours/out-of-ours service, access via telephone/ease of contact, and, finally, the adequacy and availability of physical resources. Disparities in staff availability raised the greatest interest in geographical context among the reviewed studies, since more deprived and rural areas are frequently known to have shortages of health care professionals.

Realised equity of access has also been commonly studied in the reviewed papers. This was done via the assessment of disparities in unmet health care needs (Aysola et al., 2013), avoidable hospital admissions (e.g. Gusmano, Weisz, Rodwin, Lang, Qian, Bocquier, Moysan and Verger, 2013), and health services' utilisation rates (e.g. Ridde, Haddad and Heinmüller, 2013). Overall, ten studies have analysed various patient-related parameters that may cause a disadvantage in gaining access to primary care providers. Most commonly, the evaluation of equity in realised access aimed to explore

Table 2.3 Classification of equity measures

Sub-dimension	Classification of PIs	Amado & Dyson (2009)	Amado & Santos (2009)	Aysola et al. (2013)	Buja et al. (2014)	Carrier et al. (2011)	Coughlin et al. (2008)	Dahrouge et al. (2010)	Dahrouge et al. (2013)	Ferreira et al. (2013)	Gesler et al. (2006)	Gray et al. (2007)	Gusmano et al. (2013)	Idzerda et al. (2011)	Kelaher et al. (2006)	Lamarche et al. (2010)	Lebrun et al. (2013)	Lemos et al. (2014)	Millett et al. (2008)	Millett et al. (2009a)	Millett et al. (2009b)	Morris & Landes (2006)	Morris et al. (2005)	Newton et al. (2008)	Ridde et al. (2013)	Roeger et al. (2010)	Ryvicker et al. (2012)	Sequist et al. (2011)	Shah et al. (2003)	Shi et al. (2013)	Stevens & Shi (2002)	Toda et al. (2012)	Wang & Luo (2005)	Wang (2011)	Frequency
Equity of access	Ability to get an appointment when needed/waiting time							X	X							X		X												X	X				6
	Access via telephone/ease of contact							X	X						X																X				4
	Additional direct costs																	X																	1
	Adequacy and availability of premises, equipment, and medication																															X			1
	Convenience of provider location														X	X		X												X					4
	Opportunity costs																	X																	1
	Payment requirements and policies													X	X																				2
	Presence of unmet health care needs			X																															1
	Range of health care services																														X	X			2
	Rate of avoidable hospital admissions												X																X						2
	Staff availability		X							X	X			X		X		X				X		X		X				X		X	X	X	13
	Travel distance/time to provider										X			X				X								X							X	X	6
	Utilisation rate of health care services			X					X		X					X						X	X		X		X								8
	Working hours/out-of-hours service							X	X						X	X															X				5
Equity of treatment	Management of chronic diseases	X			X	X		X				X						X	X		X							X							9
	Patient-provider interaction							X	X							X		X												X	X				6
	Receipt of preventive care and health promotion services			X		X	X	X	X																			X							6
Equity of outcomes	Health status parameters											X					X		X	X	X							X							6
	Health-related quality of life			X																															1

the influence of patients' race/ethnicity, socio-economic status of individuals and of residence areas on the utilisation rates of health care services.

The exploration of disparities in geographical access has been addressed in nine studies (e.g. Lemos, Alves, Oliveira, Rodrigues, Martins and Croda, 2014; Idzerda, Adams, Patrick, Schrecker and Tugwell, 2011). This characteristic was mainly measured through travel time/distance to providers and convenience of providers' location among patients of different race/ethnicity and patients residing in areas of differing degrees of rurality. The relevance of such focus can be explained by the fact that, on the one hand, racial/ethnic minorities frequently experience sociocultural barriers in accessing health care providers due to linguistic isolation and cultural differences. On the other hand, the rural population may also be placed in a disadvantaged position in comparison to urban citizens, as it is common to observe asymmetrical distribution of health care providers within the countries.

Disparities in economic access to primary care services have not received much attention from researchers, as only three studies have focused on this issue (Lemos et al., 2014; Idzerda et al., 2011; Kelaher, Dunt, Day and Feldman, 2006). At the same time, the ability to pay for the receipt of health care services is directly linked to the socio-economic status of individuals, a characteristic which has been commonly considered in the context of equity analysis.

In what concerns disparities in service availability, only two papers have assessed these types of disparities. Specifically, Stevens and Shi (2002) explored the differences in the range of services available for children of different race/ethnicity, whereas Toda et al. (2012) analysed the availability of services across areas with different socio-economic status.

Concerning equity of treatment, we have identified 11 studies that have made an assessment of this sub-dimension. Most equity studies belonging to this group have assessed racial and/or ethnic disparities in the treatment provided to patients. In particular, the reviewed studies have assessed the discrepancies in compliance with chronic disease management protocols (e.g. Lemos et al., 2014; Gray, Millett, Saxena, Netuveli, Khunti and Majeed, 2007) and in the delivery of preventive care and health promotion services (e.g. Aysola et al., 2013; Coughlin, Leadbetter, Richards and Sabatino, 2008) in relation to patients of different race/ethnicity. In both cases the

studies aimed to reveal whether racial/ethnic minorities tend to receive a different volume of health care services compared to non-minorities. However, beyond simple utilisation rates, which are frequently used to evaluate the level of realised accessibility, appropriateness and completeness of the treatment provided to patients has also been considered. In addition, equity of clinical treatment has also been assessed with regards to patients' age (Buja, Damiani, Gini, Visca, Federico, Donato, Francesconi, Marini, Donatini, Brugaletta, Baldo and Donata Bellentani, 2014), gender (Dahrouge et al., 2010), level of income and education (Dahrouge, Hogg, Ward, Tuna, Devlin, Kristjansson, Tugwell and Pottie, 2013; Carrier, Schneider, Pham and Bach, 2011), and residence (Amado and Dyson, 2009).

In the context of patient-provider interpersonal interactions, disparities in treatment may result from differences in the perceptions, valuations, and prejudices held by health care staff in relation to certain patient groups. The reviewed studies have addressed this issue for patients of different race/ethnicity (Lemos et al., 2014; Shi, Lebrun-Harris, Parasuraman, Zhu and Ngo-Metzger, 2013; Stevens and Shi, 2002), gender (Dahrouge et al., 2010), level of income and education (Dahrouge et al., 2013), insurance coverage (Shi et al., 2013), and for patients living in areas with differing degrees of rurality (Lamarche, Pineault, Haggerty, Hamel, Levesque and Gauthier, 2010).

Finally, equity of health care outcomes has been analysed in seven studies. The relatively lower number of the reviewed papers in this category can possibly be explained by the significant challenges involved in the measurement of this sub-dimension, most commonly associated with the need for rigorous case-mix control. All studies belonging to this group have investigated racial/ethnic disparities in health care outcomes. This was done via the assessment of racial/ethnic disparities in health status parameters for patients with chronic conditions (e.g. Gray et al., 2007) and for newborns (Lebrun, Shi, Zhu, Sharma, Sripipatana, Hayashi, Daly and Ngo-Metzger, 2013) and the assessment of these disparities in health-related quality of life indicators for children with asthma (Aysola et al., 2013).

In conclusion, several important issues must be taken into consideration in the analysis of equity. The first issue relates to the fact that it is essential to choose an appropriate perspective on equity measurement. The empirical studies undertaken until

now have mostly been concentrated on the assessment of horizontal inequities in primary health care. However, it might be the case that certain patient-related parameters should be considered from the vertical perspective. In particular, in the reviewed studies we found evidence of discrepancies in treatment and health care outcomes in relation to racial/ethnic minority groups in the UK and the USA, suggesting that some unobserved factors associated with racial/ethnic origin may affect the use and the outcomes of primary health care. If these factors are inherent, such as cultural differences related to health behaviours or genetic predisposition to certain diseases, then the application of the vertical perspective is more relevant for the analysis of racial/ethnic disparities. From this standpoint, patients' race/ethnicity can be viewed as a discriminating parameter for the purpose of equity analysis. Therefore, an appropriate distinction between discriminating and non-discriminating parameters that can (or cannot) affect the need for health care is another important issue to consider in future studies of equity in health care.

2.4.2.3 Efficiency

We identified 75 studies that reported the results of efficiency analysis (Table 2.4). To facilitate the analysis of these studies, we divided them into two broad categories depending on the approach used to measure the efficiency of health care providers.

The first category of studies has applied a single measure of input or output to develop partial ratio-based measures of efficiency. A total of 10 studies are included in this category (please refer to Table A1 in Appendix A). The partial measures of efficiency used relate to the number of consultations per patient (Lewis and Holcomb, 2012; Varatharajan et al., 2004; Abrams, Savela, Trinity, Falik, Tutunjian and Ulmer, 1995), total/average cost per visit (Heard, Nath and Loevinsohn, 2013; Radford et al., 2007; Venning, Durie, Roland, Roberts and Leese, 2000; Utriainen and Widström, 1990), and the number of patients seen per hour (Baker, 1992; Breyer, 1977) among other ratios.

The second category of studies has assessed relative efficiency of a sample of providers by incorporating simultaneously multiple inputs and multiple outputs in the

Table 2.4 Classification of efficiency measures

Input/output group	Classification of PIs	GROUP 1																														
		Amado & Santos (2009)	Andes et al. (2002)	Buck (2000)	Collier et al. (2006)	Draper et al. (2000)	Huang & McLaughlin (1989)	Ferreira et al. (2013)	Kirigia et al. (2001)	Kirigia et al. (2004)	Kirigia et al. (2008)	Kirigia et al. (2011)	Kontodimopoulos et al. (2006)	Kontodimopoulos et al. (2007)	Marathe et al. (2007) (M1)	Masiye et al. (2006) (M1)	Masiye et al. (2006) (M2)	Osei et al. (2005) (M1)	Osei et al. (2005) (M2)	Pelone et al. (2012)	Rahman & Capitman (2012)	Ramírez-Valdivia et al. (2011) (M1)	Ramírez-Valdivia et al. (2011) (M2)	Renner et al. (2005)	Rollins et al. (2001)	Rosenman & Friesner (2004) (M1)	Rosenman & Friesner (2004) (M2)	Schmacker (2008)	Sebastian & Lemma (2010)	Szczepura et al. (1993)	Thanassoulis et al. (1995)	Zavras et al. (2002) (M1,M2)
Labour resources (I)	Administrative, technical, support staff	P	P			C	P	P	P	P		P		P		P	C	P	P					P	C	P	C	P		P		P
	Medical staff	P	P	P	P	C	P	P	P	P	P	P	P	P	P	P	C	P	P	P	P	C	C	P	C	P	C	P	P	P	P,C	P
Health care services and ancillary resources (I)	Ambulatory procedures																															
	Diagnostic and laboratory tests																															
	Materials and equipment																															
	Medications/prescriptions																															
	Referrals to specialists																															
	Visits																															
Capital assets (I)	Square footage/depreciation/total assets		P																							C	P	P				
Total, operating, and other costs (I)	Operating costs																				C		C									
	Other costs					C				C															C	C						
	Total costs							C																								
Other inputs (I)	Hospital care related inputs					C				P			P						P						C							
	Atypical and specific inputs																															P
Health care services (O)	Ambulatory procedures																									P	P					
	Deliveries/abortions								P									P	P					P					P		P	
	Diagnostic and laboratory tests			P										P												P	P			P		P
	Health education sessions			P				P				P												P					P			
	Other health care services provided	P						P			P	P	P																P			
	Vaccinations and immunisations	P						P		P	P	P						P	P	P				P						P		
	Visits/episodes of care	P		P		P	P	P	P	P	P	P	P	P	P	P	P	P	P		P	P	P	P	P			P	P		P	P
	Total health care services provided		C		C																P											
Patients (O)	Registered patients																													P		
	Treated patients																												P		P	
Quality indicators (O)	Indicators on accessibility																			P												
	Composite quality indicators																															
	Indicators on effectiveness				R																										P	
Other outputs (O)	Hospital care related outputs					P							P						P	P					P							
	Atypical and specific outputs																													P	P	

I – input, O – output, P – physical units, C – costs/monetary units, R – percentage, S – index, M1-M5 – order number of a model specification in a study (if applicable)

Table 2.4 Classification of efficiency measures *(continued)*

Input/output group	Classification of PIs	GROUP 2														GROUP 3										GROUP 4												
		Akazili et al. (2008)	Cordero-Ferrera et al. (2011)	Cordero-Ferrera et al. (2014) (M1)	Cordero-Ferrera et al. (2014) (M2)	Defelice & Bradford (1997	Deidda et al. (2014)	García et al. (1999) (M1,M2)	Linna et al. (2003) (M2)	Marschall & Flessa (2011)	Murillo-Zamorano & Petraglia (2011)	Pina & Torres (1992)	Ramírez-Valdivia et al. (2011) (M3)	Ramírez-Valdivia et al. (2011) (M4)	Bryce et al. (2000)	Chilingerian & Sherman (1996)	Chilingerian & Sherman (1997)	Rosenman et al. (1997)	Siddharthan et al. (2000)	Staat (2003)	Wagner & Shimshak (2000) (M1, M2)	Wagner & Shimshak (2000) (M3, M4)	Wagner & Shimshak (2000) (M5)	Wagner et al. (2003) (M1-M4)	Amado & Dyson (2009) (M1)	Amado & Dyson (2009) (M2)	Halsteinli et al. (2010) (M1-M3)	Linna et al. (2003) (M1)	Ozcan (1998)	Ozcan et al. (2000)	Pai et al. (2000)	Schinnar et al. (1990) (M1)	Schinnar et al. (1990) (M2)	Testi et al. (2013)	Tyler et al. (1995)
Labour resources (I)	Administrative, technical, support staff	P	P	P	P	P	P		P	C		P	C			C				C								P				C	C		P
	Medical staff	P	P	P	P	P	P	P	P	C		P	C	C	C	C				C								P	C	P	P		C	C		P
Health care services and ancillary resources (I)	Ambulatory procedures															P	P			P																
	Diagnostic and laboratory tests					R		P								C	C	C											C	C	C					
	Materials and equipment								C																			C								
	Medications/prescriptions	C	P	P	P		P	C		C		P	C	C	C				C		C								C	C	C				P	
	Referrals to specialists							P								P	P	C		P										C						
	Visits														P	P	P		P		C	C	C	C					C	C	C				P	
Capital assets (I)	Square footage/depreciation/total assets									P,C																										
Total, operating, and other costs (I)	Operating costs													C																						C
	Other costs											C			C			C		C												C	C			
	Total costs																																			
Other inputs (I)	Hospital care related inputs	P													P	P	P		P				C	C					C				P		
	Atypical and specific inputs																																			
Health care services (O)	Ambulatory procedures																																			
	Deliveries/abortions	P																																		
	Diagnostic and laboratory tests																																			
	Health education sessions																																			
	Other health care services provided																																			
	Vaccinations and immunisations	P																																		
	Visits/episodes of care	P	P	P	P	P	P	P	P	P		P	P	P	P																					
	Total health care services provided																																			
Patients (O)	Registered patients														P	P	P	P	P	P	P	P	P	P	P											
	Treated patients																									P	P	P	P	P	P	P	P	P	P	P
Quality indicators (O)	Indicators on accessibility																																			
	Composite quality indicators				S			S			S											S	S	S												
	Indicators on effectiveness																																			
Other outputs (O)	Hospital care related outputs																						R	R												
	Atypical and specific outputs																														P					

Table 2.4 Classification of efficiency measures *(continued)*

Input/output group	Classification of PIs	GROUP 5														GROUP 6									Frequency (by study)
		Bates et al. (1996)	Giuffrida & Gravelle (2001) (M1)	Giuffrida & Gravelle (2001) (M2)	Linna et al. (2003) (M3)	Linna et al. (2003) (M4)	Luoma (1996)	Nuti et al. (2011)	Marathe et al. (2007) (M2)	Milliken et al. (2011) (M2)	Varela et al. (2010)	Pietilä et al. (1998)	Puig-Junoy & Ortún (2004)	Salinas-Jiménez & Smith (1996)	Cordero-Ferrera et al. (2014) (M3)	Giuffrida (1999) (M1, M2)	Giuffrida & Gravelle (2001) (M3)	García et al. (1999) (M3)	Milliken et al. (2011) (M1)	Jahangoshai Rezaee et al. (2012) (M1, M2)	Jahangoshai Rezaee et al. (2012) (M3)	Rabetti & Freitas (2011)	
Labour resources (I)	Administrative, technical, support staff														P				P					30
	Medical staff														P	P	P,C	P	P	P			P	48
Health care services and ancillary resources (I)	Ambulatory procedures																							3
	Diagnostic and laboratory tests																	P						8
	Materials and equipment																							1
	Medications/prescriptions	C													P			C						17
	Referrals to specialists																	P						6
	Visits																							10
Capital assets (I)	Square footage/depreciation/total assets																							4
Total, operating, and other costs (I)	Operating costs				C	C	C																	5
	Other costs																							9
	Total costs		C	C				C	C	C	C	C	C	C									C	10
Other inputs (I)	Hospital care related inputs																							14
	Atypical and specific inputs																			P	P			3
Health care services (O)	Ambulatory procedures																							1
	Deliveries/abortions																							6
	Diagnostic and laboratory tests																							5
	Health education sessions																							5
	Other health care services provided						P				P												P	9
	Vaccinations and immunisations												P											11
	Visits/episodes of care					P	P		P				P										P	35
	Total health care services provided							P																4
Patients (O)	Registered patients		P	P							P					P	P			P			P	14
	Treated patients	P			P							P												12
Quality indicators (O)	Indicators on accessibility		P	P									P	R		P	P							5
	Composite quality indicators									S			S		S			S	S					7
	Indicators on effectiveness		P	P										R		P	P							5
Other outputs (O)	Hospital care related outputs						P	P																9
	Atypical and specific outputs		P	P				P					R	R		P	P			P	P			8

development of efficiency models (Table 2.4). A total of 65 studies are included in this category. A significant number of studies in this category combine in a single model two types of inputs: inputs in terms of the quantities of resources used and inputs in terms of costs, making it difficult to identify which type of efficiency is analysed. Nevertheless, considering that the authors of these studies have used different combinations of inputs and outputs, we consider it helpful to cluster the studies into six main groups. In what follows, we discuss the main characteristics of the approaches used to measure efficiency in the studies belonging to each of the groups.

Group 1. The most common approach to measuring the efficiency of primary care providers involves relating the quantity (or cost) of human resources with the quantity (or value) of services delivered. The indicators related to the quantity of human resources are mainly expressed by the number of employed staff members, frequently separated by professional type into medical and non-medical categories. In some of the studies, full-time equivalents (FTEs) have been used to account for full and part-time employees (e.g. Rosenmann and Friesner, 2004; Andes, Metzger, Kralewski and Gans, 2002). Alternatively, total working time devoted to specific health care activities has been estimated (e.g. Thanassoulis, Boussofiane and Dyson, 1995). However, as noted by Thanassoulis et al. (1995), the distribution of working time between different activities is often difficult to measure accurately, since it is rarely recorded in routine practice. The costs of labour resources have usually been measured by the compensation paid to health care staff.

In addition to labour inputs, some authors have included capital inputs, expressed by square footage of practices (Schmacker and McKay, 2008; Rosenman and Friesner, 2004; Andes et al., 2002) and total assets (Rosenman, Siddharthan and Ahern, 1997). Depreciation of equipment has also been considered (Marschall and Flessa, 2011). From a theoretical perspective, this approach is in line with the classical definition of production factors.

The output measures used in the studies of this group refer to the different health care services provided. They have been most frequently measured by the number of visits (i.e. episodes of care), followed by the number of vaccinations and immunisations (e.g. Pelone, Kringos, Valerio, Romaniello, Lazzari, Ricciardi and Giulio de Belvis, 2012; Amado and Santos, 2009), deliveries/abortions (e.g. Renner, Kirigia, Zere, Barry,

Kirigia, Kamara and Muthuri, 2005; Thanassoulis et al., 1995), diagnostic and laboratory tests (e.g. Buck, 2000; Szczepura, Davies, Fletcher and Boussofiane, 1993), health education sessions (e.g. Kirigia, Sambo, Renner, Alemu, Seasa and Bah, 2011; Sebastian and Lemma, 2010), and ambulatory procedures (e.g. Rosenman and Friesner, 2004). Other indicators of health care services include indicators related to the provision of more specific actions undertaken by health care staff, for example, a measure of the assistance provided to patients with systematic arterial hypertension (Rabetti and Freitas, 2011) or the number of special examinations undertaken (Luoma, 1996).

Group 2. The studies that belong to this group have also used the total number of visits as the main output measure. However, these studies can be distinguished from those in group 1 by the types of inputs considered. The studies in group 2 complement labour inputs with other measures of inputs. Specifically, inputs related to the use of ancillary services, such as diagnostic and laboratory tests (García, Marcuello, Serrano and Urbina, 1999; Defelice and Bradford, 1997), inputs related to referrals to specialists (García et al., 1999), inputs related to materials and equipment (Linna, Nordblad and Koivu, 2003), and inputs related to medications (e.g. Cordero-Ferrera, Crespo-Cebada and Murillo-Zamorano, 2011; Pina and Torres, 1992). The inclusion of these input variables aims to make a fairer assessment of relative efficiency, considering that a higher number of visits may be easier to secure for some providers if more ancillary services and resources are available. However, it is important to point out that there is no unique pattern in the treatment of different health conditions that could justify, for example, identical needs in medications or diagnostic tests for all patients seen by a provider. For this reason, combining this type of inputs with outputs related to the total number of visits could be potentially misleading without an appropriate case-mix adjustment. This adjustment is essential to control for factors affecting the intensity of use of ancillary services and resources.

An aspect that is relevant for the majority of studies in groups 1 and 2 is that a larger volume of health care services does not necessarily indicate higher quality of care. To overcome this shortcoming, some of the studies that belong to group 1 and 2 include measures related to the quality of care provided, such as the rate of avoidable hospital admissions (Pelone et al., 2012), patient satisfaction indicators (Collier, Collier and Kelly, 2006; Thanassoulis et al., 1995), survival rate among newborns

(Thanassoulis et al., 1995), and composite quality indicators (Murillo-Zamorano and Petraglia, 2011; García et al., 1999).

Group 3. This group includes models that estimate the efficiency of health maintenance organisations (HMOs) and other managed care providers in the USA. Attempting to restrain health care costs, HMOs have implemented the capitation-based system, according to which contracted physicians receive a fixed payment for each registered patient (Zuvekas and Cohen, 2010). At the same time, HMO patients have gained access to comprehensive primary health care in exchange for a membership fee (Chilingerian and Sherman, 1997). Under such circumstances, registered patients generate revenues for HMOs, even if they never visit a physician, and an efficient HMO would use fewer health care services (inputs) to provide coverage for a certain number of enrolled patients (outputs) (Siddharthan, Ahern and Rosenman, 2000; Chilingerian and Sherman, 1996). Clearly, this approach to measuring efficiency does not fit other models of primary care delivery, in which a prepaid basis for the use of health care services is not adopted.

Group 4. The papers included in this group have considered a patient-related output variable, which is the number of patients treated. As implied by its name, this measure takes into account only those patients that have actually received treatment. In the reviewed studies, this variable has commonly been concentrated on specific conditions, such as diabetes (Amado and Dyson, 2009), dental health problems (Linna et al., 2003), otitis media (Ozcan, 1998), sinusitis (Ozcan, Jiang and Pai, 2000; Pai, Ozcan and Jiang, 2000), and mental illnesses (Tyler, Ozcan and Wogen, 1995; Schinnar, Kamis-Gould, Delucia and Rothbard, 1990). With regards to the inputs included in these studies, the most frequently used inputs were the costs of labour resources and/or the cost of health services provided and ancillary resources consumed.

Overall, the patient-related outputs used in groups 3 and 4 have frequently been controlled for case-mix factors, such as patients' age, gender, insurance type, residence status, residence in deprived/non-deprived areas, and disease severity.

Group 5. In this group of studies, the health care inputs have been reduced to a single measure of costs. In general, the use of total costs as a unique input measure is considered to be particularly useful whenever detailed information on the quantities of consumed resources and respective prices is not available. This justifies the fact that the

components of total costs have not been specified in some studies (e.g. Linna et al., 2003). In some of the reviewed studies, it is possible to find some variation in terms of the cost components included in the calculation of total costs. For instance, Giuffrida and Gravelle (2001) included the GPs' costs and the prescription costs, while Milliken, Devlin, Barham, Hogg, Dahrouge and Russell (2011) included medical and non-medical staff incomes, operating costs, and capital costs, expressed in rent and depreciation values. However, regardless of the nature of the cost components to be included, the aggregation of inputs into a single cost variable may also limit the potential of efficiency analysis. In particular, the use of a single input will not allow the analyst to disentangle between technical and allocative efficiency. The output side of the models used in this group of studies does not follow any particular pattern, presenting different combinations of the outputs discussed above.

Group 6. The sixth group of studies includes atypical efficiency models and models with specific inputs/outputs. In most cases, these models have been used due to specific study objectives. For example, Giuffrida (1999) and Giuffrida and Gravelle (2001) developed efficiency models which, among other variables, incorporated outputs on the achievement of quality targets by English Family Health Services Authorities in order to conclude about the effectiveness of the NHS organisational reforms. We also found inputs and outputs that do not fall into the categories discussed above. For example, a number of studies have combined the evaluation of primary health care with hospital care and, therefore, they have added hospital care related variables into the efficiency models (e.g. Bryce, Engberg and Douglas, 2000; Draper, Solti and Ozcan, 2000).

Overall, the specifications discussed above constitute the main basis for defining inputs and outputs in primary health care. Each of the specifications has its own advantages and drawbacks, which make their application dependant on the research context and aims. After identification of both inputs and outputs, the evaluation of efficiency requires the definition of appropriate weights to each of the selected variables and, according to the production theory, the establishment of a clear relationship between inputs and corresponding outputs. However, these tasks are far from straightforward in a complex setting such as primary health care. In this respect, SFA and DEA provide different solutions to both issues, as briefly discussed in Section 2. Despite the fact that each method has its own strengths and limitations, we found a

more frequent use of DEA in the reviewed efficiency studies in comparison with regression-based methods such as SFA. This can be partly explained by the fact that DEA does not require parametric assumptions regarding the relationship between the selected inputs and outputs. At the same time, as noted by Jacobs, Smith and Street (2006b), the efficiency scores derived from either of the techniques should not be treated as accurate efficiency estimates, given their high sensitivity to the choice of model specification.

2.4.2.4 Effectiveness

Effectiveness has been studied in 57 of the 166 papers reviewed (Table 2.5). Among all the effectiveness sub-dimensions, the investigation of clinical effectiveness has attracted the greatest attention in the reviewed literature, being explored in 39 papers.

The achievement of clinical effectiveness is commonly viewed as compliance with clinical practice guidelines that comprise a set of standardised algorithms in the management of common diseases. For example, an effective diabetes management has been defined as the accomplishment of the following activities: eye examination performed, foot and leg inspection undertaken, weight/BMI recorded, smoking status inquired, blood pressure checked, glycated haemoglobin level tested, cholesterol level recorded, kidney function checked, appropriate medications prescribed, treatment offered to improve diabetes control (if needed), self-management plan provided (if needed), and follow-up scheduled (e.g. Amado and Dyson, 2009; Campbell, Reeves, Kontopantelis, Middleton, Sibbald and Roland, 2007; Ulmer, Lewis-Idema, Von Worley, Rodgers, Berger, Darling and Lefkowitz, 2000).

Overall, there are five categories of health care activities that have mostly been explored in the management of different medical conditions: diagnostic and laboratory tests, health education and counselling, medication management, physical examination, and preventive screening, vaccination, and immunisation. Less frequently, clinical effectiveness has been measured through the indicators related to characteristics of medical history taking, follow-up visits, medical records, symptom relief, and referrals to specialists.

Table 2.5 Classification of effectiveness measures

Sub-dimension	Classification of PIs	Amado & Santos (2009)	Amado & Dyson (2009)	Aysola et al. (2013)	Barbaro et al. (2014)	Breyer (1977)	Campbell et al. (2007)	Collier et al. (2006)	Edward et al. (2011)	Ferreira et al. (2013)	Friedberg et al. (2007)	Furtado et al. (2013)	Geboers et al. (2002)	Giuffrida (1999)	Giuffrida & Gravelle (2001)	Habib et al. (2011)	Haggerty et al. (2007)	Haggerty et al. (2008)	Heard et al. (2013)	Hedeen et al. (2002)	Jaakkimainen et al. (2011)	Jacobs et al. (2006a)	Jaiveer et al. (2006)	Keller et al. (2014)	Kern et al. (2014)	Klemenc-Ketis et al. (2014)	Lee et al. (2009)	Lewis & Holcomb (2012)	Llanwarne et al. (2013)	Macinko et al. (2007)
Clinical effectiveness	Diagnostic and laboratory tests		X				X				X									X	X		X	X	X			X		
	Follow-up visits													X										X			X			
	Health education and counselling		X				X		X			X				X							X	X			X		X	
	History taking								X							X															
	Medication management		X				X				X	X								X	X		X	X	X			X		
	Physical examination		X			X	X		X		X					X				X	X		X	X	X	X				
	Preventive screening, vaccination, and immunisation			X							X			X	X					X	X		X	X	X	X	X	X		
	Record of health-related behaviours						X																X							
	Record of standard health status parameters		X				X																X							
	Record of symptoms						X																							
	Referral to specialists						X																							
	Symptom relief						X																			X				
	Other clinical activities						X						X																	
Effectiveness of interpersonal care	Continuity of care				X							X					X	X										X	X	X
	Quality of patient-provider communication											X	X													X	X		X	X
	Length of consultations											X														X				X
	Staff behaviour/attitude towards patients				X							X	X													X	X		X	X
	Patient attitude towasrds health care staff																										X		X	
	Psychosocial support/distress relief												X			X										X				
	Whole-person/family-focus orientation				X							X					X									X	X			X
Outcomes achievement	Health care needs met																													
	Health-related behaviours																					X								
	Health-related quality of life			X																										
	Health status parameters		X				X													X	X		X	X				X		
	Patient knowledge		X				X												X											
	Patient satisfaction	X	X					X	X	X																		X	X	

Table 2.5 Classification of effectiveness measures *(continued)*

Sub-dimension	Classification of PIs	Mills et al. (2004)	Montgomery et al. (2004)	Mosquera et al. (2013)	Pelletier et al. (2014)	Safran et al. (1998)	Safran et al. (2002)	Salinas-Jiménez & Smith (1996)	Schinnar et al. (1990)	Sequist et al. (2008)	Shahidzadeh-Mahani et al. (2009)	Shi et al. (2003)	Shi et al. (2012)	Stevens et al. (2014)	Teach et al. (2006)	Testi et al. (2013)	Thanassoulis et al. (1995)	Tourigny et al. (2010)	Tsai et al. (2010)	Ulmer et al. (2000)	Utriainen & Widström (1998)	Varatharajan et al. (2004)	Venning et al. (2000)	Villiers et al. (2005)	Wang et al. (2014a)	Wang et al. (2014b)	Wensing et al. (2008)	Wong et al. (2010)	Zallman et al. (2010)	Frequency
Clinical effectiveness	Diagnostic and laboratory tests	X			X					X		X		X		X				X					X			X	X	19
	Follow-up visits	X			X						X		X							X			X							9
	Health education and counselling					X	X			X	X	X		X	X			X	X						X			X	X	21
	History taking										X																			3
	Medication management				X							X		X	X				X	X			X		X			X	X	20
	Physical examination		X		X	X	X			X	X			X						X			X							21
	Preventive screening, vaccination, and immunisation							X		X			X		X			X												17
	Record of health-related behaviours															X				X					X					5
	Record of standard health status parameters	X									X									X					X				X	8
	Record of symptoms	X																		X										3
	Referral to specialists													X									X							3
	Symptom relief																													2
	Other clinical activities	X																												3
Effectiveness of interpersonal care	Continuity of care		X	X		X	X			X		X		X				X	X							X		X		18
	Quality of patient-provider communication		X	X		X	X			X	X	X		X					X				X					X		17
	Length of consultations	X	X	X		X	X			X		X		X								X						X		13
	Staff behaviour/attitude towards patients		X	X		X	X			X		X		X				X	X				X	X		X	X	X		21
	Patient attitude towards health care staff		X			X	X					X																		6
	Psychosocial support/distress relief																						X							4
	Whole-person/family-focus orientation		X	X		X	X			X		X		X				X	X							X		X		17
Outcomes achievement	Health care needs met								X																					1
	Health-related behaviours																													1
	Health-related quality of life																													1
	Health status parameters				X				X	X			X				X				X				X					14
	Patient knowledge										X																			4
	Patient satisfaction																X					X	X							10

The accomplishment of specific activities is not the unique parameter captured by these types of indicators. Since most health care activities have to be repeated regularly within the time frames defined by clinical guidelines, the measurement of clinical effectiveness also considers the regularity at which these activities should be undertaken. In general, timeliness, appropriateness, and completeness of health care delivery are the attributes of care that are essential to consider in the assessment of clinical effectiveness.

Effectiveness of interpersonal care has been studied in 28 of the 166 papers we have reviewed. There are five categories of measures that have been most commonly explored in the studies: staff behaviour/attitude towards patients, quality of doctor-patient communication, length of consultations, continuity of care, and whole-person/family-focus orientation (Table 2.5). Concerning staff behaviour and attitude towards patients, the reviewed studies have emphasised the importance of various characteristics that are crucial for establishing trustful patient-provider relationships. These characteristics include helpfulness, friendliness, caring, respect, patience, kindness, courtesy, and concern for patients demonstrated by health care staff (e.g. Sequist, Schneider, Anastario, Odigie, Marshall, Rogers and Safran, 2008; Geboers et al., 2002; Safran, Wilson, Rogers, Montgomery and Chang, 2002). The quality of patient-provider communication, in turn, has been assessed via the evaluation of the following characteristics: thoroughness of doctor's questions about symptoms and health-related concerns, attention to what the patient says, allowing time for questions, discussion of treatment options, clarity of explanations and instructions, and advice in making decisions about care (e.g. Klemenc-Ketis, Kravos, Poplas-Susič, Svab and Kersnik, 2014; Safran, Taira, Rogers, Kosinski, Ware and Tarlov, 1998). In addition, length of consultations has been used as a proxy measure to estimate the time potentially available for interpersonal interactions between the patient and the doctor (e.g. Macinko et al., 2007; Venning, Durie, Roland, Roberts and Leese, 2000). Continuity of care, another aspect in the measurement of effectiveness of interpersonal care, has been related to the patient's ability to see a designated doctor (e.g. Lewis and Holcomb, 2012; Macinko et al., 2007) and the length of patients' enrolment in the practice (e.g. Safran et al., 2002). Finally, whole-person and family-focused orientation in service provision has been defined as an important attribute of an effective health care delivery (e.g. Furtado, Braz, Pina, de Mello and de Lima, 2013; Montgomery et al.,

2004), as it ensures the consideration of a patient's personal and familial characteristics in the treatment provided.

Regarding outcomes achievement, this has been analysed in 25 papers. We found that control of health status parameters and patient satisfaction were the most commonly explored outcomes of primary care delivery, being assessed in 14 and 10 papers respectively. The measurement of health status parameters has often been limited to a restricted number of indicators in the reviewed studies. For example, outcomes of diabetes control were typically measured by the level of blood pressure, cholesterol level and glycated haemoglobin level (e.g. Campbell et al., 2007), despite the fact that diabetes management covers a broader range of health issues associated with this chronic disease. The measurement of patient satisfaction, in turn, has been used to assess the quality of health care delivery as perceived by patients (e.g. Amado and Santos, 2009; Collier et al., 2006). Other outcomes of primary care delivery include patient knowledge as a result of contacts with the provider (Heard et al., 2013; Amado and Dyson, 2009), changes (or their absence) in health-related quality of life (Aysola et al., 2013), changes (or their absence) in health-related behaviours (Jacobs, Martin, Goddard, Gravelle and Smith, 2006a), and the extent to which health care needs are met (Schinnar et al., 1990).

In summary, several concluding remarks regarding the measurement of effectiveness in health care can be provided. As we discussed in Section 2, effectiveness in health care can be measured in two distinct ways: directly, by observing the achieved outcomes of health care, and indirectly, by investigating the quality of the health care process. The selection of an appropriate approach is context-dependant and, despite the investigation of outcomes achievement being generally preferred for assessing the effectiveness of health care delivery and frequently being viewed as the gold standard, it is often impossible to perform or is limited to a restricted number of indicators. Moreover, whatever approach is selected, it is essential to guarantee validity of the observed performance results. Particularly, attributional validity is crucial to ensure that the observed outcomes can be attributed to the health care provided, while causal validity reflects the capacity of certain health care processes to produce specified outcomes under stipulated conditions (Donabedian, 1980).

2.4.2.5 Discussion of the main results regarding the different performance dimensions

Our review allows us to conclude that past studies of performance assessment in primary health care have focused on the four dimensions of performance included in the framework (accessibility, equity, efficiency, and effectiveness). Specifically, we have concluded that efficiency has been the dimension most commonly analysed (75 studies), followed by accessibility (59 studies), and effectiveness (57 studies). Despite its relevance, only 33 of the reviewed studies have focused on equity.

The majority of the studies have focused on a single dimension (please refer to Table A1 in Appendix A). Nevertheless, it is also possible to find some studies that have analysed several dimensions (e.g. Barbaro, Lettiere and Nakano, 2014; Aysola et al., 2013; Amado and Dyson, 2009; Amado and Santos, 2009). Given that trade-offs are likely to exist between the different performance dimensions, we consider that future research studies should make efforts to evaluate several performance dimensions within a single framework in order to explore the evidence of trade-offs between them.

Below we synthesise some of the key features to be considered in the assessment of each performance dimension and discuss other directions for future research.

Accessibility is a prerequisite for contacts to be established between the individual and the health care provider. In the health care context the measurement of accessibility can be performed from two different perspectives: by investigating the potential accessibility of health care providers and by measuring the level of realised access.

The potential accessibility of health care organisations reflects the presence or absence of geographical, economic, and organisational obstacles that may prevent potential users from entering into health systems, and also refers to the comprehensiveness of available health care services. Despite potential accessibility being relatively easy to assess by observing certain provider characteristics, the influence of different groups of accessibility factors on each other has not been well-studied. For example, it might be the case that a smaller range of health care services may result in a better organisation of their delivery and, consequently, this may lead to the reduction of organisational barriers to access, such as the shortening of waiting lists, at the expense of service availability. At the same time, the question about which

characteristics possess greater importance in forming an overall level of potential accessibility remains unanswered.

Realised accessibility, in turn, reflects the success or failure in gaining access to a primary care provider. In the measurement of realised accessibility it is important to consider whether the possibility to access a provider really exists and those environmental factors that influence an individual's decision not to access a provider, despite having such an opportunity and recognising the need for health care. We found that both issues are relevant for further analysis in order to develop more valid and reliable measures on realised accessibility.

Equity deals with a fair treatment of individuals at every stage and in every aspect of health care delivery. A comprehensive assessment of equity demands the analysis of all three sub-dimensions – equity of access, equity of treatment, and equity of health care outcomes. Furthermore, in order to investigate equity, it is essential to choose either a horizontal or vertical perspective on equity measurement as well as to define discriminating/non-discriminating patient-related parameters that should (or should not) influence the need for health care. The exploration of vertical inequities might be relevant in relation to some patient groups. However, its assessment is associated with certain challenges that should be explored in future research. In particular, it is important to establish criteria on how to appropriately differentiate the volume of health care services that is demanded by individuals with different levels of health care needs. The relationship between the three sub-dimensions of equity is another important issue to consider in the assessment of this dimension. In particular, disparities in access for certain patient groups result in hidden inequities in their treatment, since those patients that have not succeeded in gaining access to a health care provider automatically do not receive the necessary treatment. Similarly, disparities in treatment should, in principle, cause disparities in corresponding health care outcomes. If this correlation is not confirmed, a more careful control for case-mix factors should be carried out and the check of the attributional/causal validity between clinical processes and outcomes should be performed.

Efficiency aims to optimise the use of available resources for the production of targeted output levels. In this review we have distinguished six common specification patterns that are used for the evaluation of efficiency of primary care providers.

Independently of the pattern, the input side of the models is usually measured by the quantity (or costs) of labour resources, complemented (or not) with the amount (or costs) of ancillary resources, with measures of capital assets, and with other cost measures. Some models summarise all inputs in the form of total costs, which is considered to be particularly advantageous when the attribution of multiple inputs for the production of observed/expected outputs is particularly difficult to perform. Regarding the output variables of the models, these frequently include the volume of the health care services provided to patients or, alternatively, the number of registered/treated patients as an aggregate measure of an expected/performed volume of health care services. In some models these outputs are complemented with quality measures to account for those provider characteristics that are considered to be important attributes of a high-quality primary health care. This is essential to prevent activities being performed whenever they can adversely impact the quality of the health care provided. Overall, regardless of the choice of the efficiency measurement models to adopt, there is still no consensus on how to define the relative importance of the multiple outputs produced. It is also unclear how to associate the specific contribution of each input with the production of each output. The DEA technique provides one of the most commonly adopted solutions to both issues and its application has become more frequent for the assessment of efficiency of primary care providers. Nevertheless, questions on how to select the most appropriate models and on how to derive robust efficiency estimates should be further explored.

Effectiveness is concerned, in turn, with the accomplishment of the objectives of health care, which is usually captured by outcomes achievement. At the same time, the assessment of clinical effectiveness and the effectiveness of interpersonal care aim to present alternative approaches of effectiveness measurement in health care by exploring the quality of two core aspects of the health care process. In particular, clinical effectiveness assumes that a targeted impact of health care is likely to occur if health care processes correspond to the requirements of evidence-based practice or any legitimate health care standards and norms. Effective interpersonal care, in turn, is believed to be a good predictor of improved health care outcomes. Yet, the correspondence between outcomes and processes of care has to be carefully verified, since predicted outcomes may significantly differ from observed consequences of the health care provided. For this reason, outcomes measurement remains an essential

element in assessing the effectiveness of health care provision, which directs attention towards the impact of health care services rather than their attributes and volume. Indeed, future empirical research on the evaluation of effectiveness of primary care providers should pay more attention to the development of patient-reported outcome measures, such as indicators on health-related quality of life, as their measurement can provide evidence of the impact of health care on patients' life beyond standard clinical parameters. This will contribute to balancing the functioning of health care systems to equally target the achievement of both clinical and non-clinical outcomes.

2.4.3 Exploring stakeholders' involvement at different stages of organisational performance assessment

As discussed in Section 2, in order to promote performance improvement as a response to assessment results, it is fundamental that targeted stakeholders are involved during the three stages of the performance assessment process: measurement, analysis, and action. Supporting this idea, we analysed in this review how common it has been to involve stakeholders in primary care assessment.

We have summarised the results of our analysis in Table 2.6. Only 17 out of the 166 studies reviewed provided clear evidence of stakeholders' involvement at the measurement and/or analysis stages. This involvement occurred through discussions (Gusmano et al., 2013), focused interviews (Abrams et al., 1995), workshops (Amado and Dyson, 2009), expert consultations (Ramírez-Valdivia, Maturana and Salvo-Garrido, 2011; García et al., 1999) and other forms of collaboration. Judging by the evidence provided in the reviewed papers, one may conclude that the development of conceptual frameworks and/or the interpretation of the observed performance results have rarely been done with the participation of stakeholders. We also found a limited amount of evidence demonstrating that the research results have been distributed to the targeted groups of stakeholders in order to link the findings to specific actions in practice. This information was presented only in 11 studies. It is still possible that some studies may have involved the targeted stakeholders in the assessment process, but have not reported this fact in the paper. However, we believe that this is the exception rather than a common feature of the majority of the reviewed studies.

Table 2.6 Evidence of stakeholders' involvement

Paper's reference	Evidence of stakeholders' involvement at the measurement and/or analysis stages	Evidence of stakeholders' involvement at the action stage
Abrams et al. (1995)	Yes. Focused interviews were undertaken with health care executives to address the impact of community health centres (CHCs) on accessibility in primary care and to discuss CHCs' performance	No
Amado and Santos (2009)	Yes. Comments were received from nine doctors of the Portuguese national health service regarding the conceptual framework used and the results obtained	Yes. The article was sent to several professionals involved in the management and delivery of primary care in Portugal, having the potential to have impact in practice
Amado and Dyson (2009)	Yes. Workshops were undertaken with some of the Board members of the participating Primary Care Groups and Trusts to develop the conceptual framework for analysis and to validate the results	Yes. Meetings were held with eight general practices, with the performance manager and three Board members of one of the participating Primary Care Groups in order to discuss the results and to potentiate their use in practice
Breyer (1977)	Yes. On-site interviews were held with the health centres' managers and staff members to determine the attributes of the selected performance indicators	No
Collier et al. (2006)	Yes. Interviews were held with three part-time physicians to identify the importance of an accurate part-time-equivalent estimate	No
García et al. (1999)	Yes. The choice of the output variables was made by a group of experts (managers in primary care units) among other methods employed in the study	No
Geboers et al. (2002)	Yes. "A panel of five experts defined 27 indicators on quality for six dimensions of the quality of care." (Geboers et al., 2002: 52)	Yes. "After the assessment took place, 20 practices participated in an intervention study in which they set priorities and performed concrete improvement projects." (Geboers et al., 2002: 52)

Table 2.6 Evidence of stakeholders' involvement *(continued)*

Paper's reference	Evidence of stakeholders' involvement at the measurement and/or analysis stages	Evidence of stakeholders' involvement at the action stage
Gusmano et al. (2013)	Yes. Four meetings were held with the key stakeholders to present the developed indicators and methods and to discuss the local factors that may explain variations in the performance results	Yes. Preliminary findings of the study were discussed with the key stakeholders
Lee et al. (2009)	Yes. "A nine expert panel was assembled for [conceptual] tool development." (Lee et al., 2009: 103)	No
Macinko et al. (2007)	Yes. "The project described here was carried out in close collaboration with municipal health authorities." (Macinko et al., 2007: 168)	Yes. "Discussions with municipal managers in Petrópolis revealed that the results of the provider and user surveys were easily understood and, perhaps more importantly, readily linked to specific actions to improve the organization and delivery of care at both the clinic and district levels." (Macinko et al., 2007: 174)
Mosquera et al. (2013)	No	Yes. The findings of this study were reported to the District Health Secretariat
Nuti et al. (2011)	Yes. "Chief executive officers from the 12 Local Health Authorities in the Tuscany Region worked with the research team to select the variables to be used in this study." (Nuti et al., 2011: 325)	Yes. "Although data envelopment analysis has been accepted by the chief executive officers as a valid tool to measure efficiency, it proved to be a complicated tool to manage efficiency. Indeed, the chief executive officers found it problematic to translate the efficiency scores into specific actions to be taken in order to improve their performance." (Nuti et al., 2011: 328)
Pina andTorres (1992)	No	Yes. The managers of health centres were informed about how to improve the allocation of available resources on the basis of the study results
Radford et al. (2007)	Yes. Participants of the study (representatives of CHCs) selected the measures to be included into a customised balanced scorecard for each center	Yes. "Participants [of the study] indicated their scorecards would be shared with managers, department heads, boards of directors, finance committees, and clinical providers. They planned to use the information for board education, planning, benchmarking, quality improvement, and modifying operations," (Radford et al., 2007: 28)

Table 2.6 Evidence of stakeholders' involvement *(continued)*

Paper's reference	Evidence of stakeholders' involvement at the measurement and/or analysis stages	Evidence of stakeholders' involvement at the action stage
Ramírez-Valdivia et al. (2011)	Yes. "The variables included in the DEA model are selected (…) by consensus with the Ministry of Health Primary Care Department experts." (Ramírez-Valdivia et al., 2011: 1024)	No
Schinnar et al. (1990)	Yes. Three expert panels (state officials and planners, partial care program directors, university researchers) were asked to develop specifications for each measure selected	No
Sebastian and Lemma (2010)	No	Yes. The study results were discussed with the Tigray Health Bureau in order to develop an improvement strategy
Thanassoulis et al. (1995)	Yes. "The resources used by DHAs [District Health Authorities] to provide perinatal care and ways of reflecting corresponding activity levels were carefully examined in consultation with specialists in the managerial and medical aspects of perinatal care provision." (Thanassoulis et al., 1995: 590)	No
Varatharajan et al. (2004)	No	Yes. Key informant interviews were held to identify the directions for performance improvement in primary care delivery in the province of Kerala, India, based on the study results
Villiers et al. (2005)	Yes. Focus group discussions with health care worker and patients were conducted during the study	No
Wang et al. (2014)	Yes. "The expert review identified key modifications to the PCAT-C version [Primary Care Assessment Tool – Chinese version] to reflect the Tibetan context." (Wang et al., 2014: 5)	No

Overall, the findings of our review suggest that the involvement of stakeholders at different stages of organisational performance assessment has been rare. Therefore, it remains questionable whether the studies' research findings possess face validity and whether these findings can be successfully used for the development of specific improvement strategies and plans. Furthermore, we found that the majority of studies that have reported an involvement of stakeholders were privileged with the involvement of managers and health care professionals. Notwithstanding the relevance of involving staff members, there are other stakeholders that should also be involved in the performance assessment process. In particular, the greater involvement of patients and citizens can contribute to highlight different perspectives, doing justice to the complexities related to performance assessment in health care provision.

2.5 Conclusions

The review undertaken fills a gap in the health care literature by analysing and systematising a very broad range of empirical studies dedicated to performance assessment in primary health care. By answering two relevant research questions, it extends and complements previous reviews in several important ways, offering valuable information for researchers, practitioners, and policymakers.

In order to answer the research questions addressed in this study, we developed a conceptual framework for the performance assessment of primary care providers that was inspired by previous theoretical works. This framework served as a basis to analyse the selected empirical studies and also constitutes a valuable tool for researchers, policymakers, and health care managers to overview the process of performance measurement in the health care context.

The first research question aimed to identify the main performance dimensions and measures that have been used in primary care assessments. We concluded that the past studies of performance in this context have focused on the four dimensions of performance included in the framework (accessibility, equity, efficiency, and effectiveness) and that a wide array of performance measures has been developed to assess these dimensions. We then systematised these measures in summary tables to assist the selection of an appropriate set of performance indicators for the evaluation of health care providers. We also concluded that the majority of

the studies have focused on the investigation of a single dimension. Given that trade-offs are likely to exist between the different performance dimensions, we believe that future research studies should make more effort to evaluate several performance dimensions and to discuss the evidence of trade-offs between them.

The second research question of our review aimed to explore the extent of stakeholders' involvement in the performance assessment process in primary health care. We concluded that only a minority of the reviewed studies have reported the involvement of targeted stakeholders during the three stages of performance assessment. This finding allowed us to infer about a weak connection between the reviewed empirical studies and health care policy and practice, which diminishes the potential impact of the studies' research findings on performance improvement of primary care providers. Furthermore, the stakeholders involved in primary care assessments have mostly been health care staff members, policymakers, and managers, with only a small minority of studies reporting the involvement of other relevant stakeholders, such as patients and citizens. Therefore, it is advisable that future studies aim to involve the key stakeholders in the process of primary care assessment, including patients and citizens among others, to ensure that the models and results have face validity and also to strengthen the impact of empirical research on health care policy and practice.

Despite the contributions discussed above, there are several limitations associated with this review. The first limitation relates to the fact that we only reviewed papers written in English. Indeed, there is a chance that including studies written in other languages would have identified additional performance measures or could have changed the proportion of studies that report having actively involved stakeholders in the assessment process. At the same time, by providing a review of the literature written in different languages we could expect the sample of papers to be more representative and better reflect the state of the art in this research area. It would therefore be valuable if future studies could cover articles written in other languages. The second limitation relates to the fact that data extraction involved judgements by the authors, mainly in respect to the classification of the measures used for performance assessment. Despite our great care in this process, this can be a potential source of bias. The third limitation relates to the absence of a quality assessment of the studies. We carefully selected criteria for exclusion and inclusion of the papers based on the common recommendations for systematic reviews in health care. However, to our knowledge, there is no established criterion for conducting quality appraisal of the empirical studies devoted to health care performance assessments. Therefore, this issue must be addressed in future

research. Notwithstanding these limitations, it is our conviction that this review adds to our understanding of performance assessment in primary health care and opens up new avenues for future research.

Chapter 3

EXPLORING THE POTENTIAL OF DATA ENVELOPMENT ANALYSIS FOR ENHANCING P4P PROGRAMME DESIGN IN PRIMARY HEALTH CARE: AN APPLICATION TO THE PORTUGUESE CONTEXT

3.1 Introduction

For the last several decades, the implementation of pay-for-performance (P4P) programmes in health care has become a worldwide initiative that now involves nearly two-thirds of the member countries of the Organisation for Economic Co-Operation and Development (OECD) (OECD, 2016). Such programmes aim to encourage quality improvements in the delivery of health care services by rewarding providers (primary care staff, specialists, or hospitals) for the achievement of desirable performance results (Cashin, Chi, Smith, Borowitz and Thomson, 2014). The idea behind P4P is that by making explicit the connection between the results of the work and respective bonus, one can motivate the staff to perform better in a problematic or highly important area of health care delivery (Eijkenaar, 2011). Considering that different areas of health care performance may require improvement, the objectives of P4P programmes significantly vary, from increasing uptake of IT services to lowering health care expenditure, reducing health disparities, or improving patient satisfaction.

Although the present popularity of P4P programmes is not supported by conclusive evidence regarding their efficacy in practice (Cashin et al., 2014; Eijkenaar, Emmert, Scheppach and Schöffski, 2013; Scott Sivey, Ouakrim, Willenberg, Naccarella, Furler and Young, 2011; Van Herck, De Smedt, Annemans, Remmen, Rosenthal and Sermeus, 2010), there is no doubt that P4P programme design plays a central role on how well a programme

will accomplish its objectives (Eijkenaar, 2011). All P4P programmes share a core set of design elements, including performance domains and measures; target entity; basis for reward or penalty; nature, size, and frequency of the reward or penalty; data reporting and verification among others, and each of these elements is associated with a wide variety of choices (Cashin et al., 2014; Conrad, 2009). Several studies have provided overviews on the state of the art on P4P programme design and important determinants of a programme's success have been highlighted, such as: (1) defining performance broadly, via a comprehensible set of process and outcome measures; (2) active monitoring of side-effects, such as "patient selection" and "treating to the test" among others; (3) applying sophisticated risk adjustment techniques; (4) involving key stakeholders in programme design; (5) favouring group incentives over individual incentives, especially for targets that require cooperation and coordination; (6) either rewarding or penalising the incentivised units, depending on the context; (7) targeting frequent and low-powered incentive payments; (8) favouring absolute targets over relative targets; (9) focusing on multiple targets rather than on single targets; and (10) P4P should be a permanent element of overall provider payment systems (Eijkenaar, 2011; Conrad, 2009).

While recognising the importance of the suggested design choices, in this paper we aim to investigate a fundamental design element that has not attracted much attention in the P4P literature – the method for performance assessment. We can generally say that most P4P evaluations assume the assessment of a unit's performance against a number of performance indicators (PIs) and these PIs are usually assessed separately from each other. Such evaluations neglect the existence of trade-offs between different PIs. They usually involve a simple ratio analysis, where the numerator represents the unit's actual performance results on a particular PI and the denominator represents a defined benchmark: an absolute threshold, if an assessment of absolute performance takes place, or the top-ranked result on the PI among peer units, in case of relative performance assessment, or the unit's previous results on the same measure, if the change in measure is being assessed. Afterwards, regardless the method chosen for the assessment of individual PIs, the results on multiple indicators are aggregated into a composite measure of performance, and for its calculation a fixed weighting system is commonly applied. These weighting systems aim to reflect value judgements regarding the relative importance of each PI. However, due to their inflexibility, these weighting systems frequently fail to account for possible variations in priorities held by different stakeholders.

As an alternative to the above described methodology for performance evaluation, in this paper we aim to explore the potential of Data Envelopment Analysis (DEA) for enhancing P4P programmes in health care. DEA is a non-parametric method, developed by Charnes et al. (1978) for measuring the relative performance of peer Decision Making Units (DMUs), for example, health care providers. According to the DEA methodology, a DMU's relative performance score is defined as the weighted sum of outputs (i.e. goods or services produced from the inputs) to the weighted sum of inputs (i.e. resources), engaged in the production process. This means that, unlike a ratio analysis, DEA treats all inputs and outputs at once and considers the existence of trade-offs between different production factors. Furthermore, with an application of the standard DEA model, each DMU is free to select the weights that maximise its relative performance score, and this score ranges from 0 to 100%. As a result of DEA, DMUs with the maximum score (100%) are considered to be top performers and they form an empirical best practice frontier, while DMUs with a relative score less than 100% are seen as underperforming units that are "enveloped" by this frontier.

To date, DEA has been used extensively in a wide range of industries. The majority of DEA applications can be found in banking, health care, agriculture, and transport systems (Liu, Lu, Lu and Lin, 2013). Within the health care systems, DEA has been commonly used for constructing composite performance measures for hospitals and less frequently for the performance assessment of primary care providers (Kalinichenko, Amado and Santos, 2013; Amado and Dyson, 2008; Hollingsworth, 2008). However, in the context of incentive regulation, DEA has received limited attention both within and outside the health care sector. The main contribution in this respect belongs to Bogetoft and colleagues (i.e. Bogetoft, 1994; Agrell et al., 2002) who have explored the use of DEA in minimising the expected payments to incentivised units. Following these studies, Felder and Schmitt (2004) developed a DEA-based reimbursement scheme to incentivise German hospitals that have realised cost savings in comparison to the individual reference levels, identified by DEA. Later, Prior and Surroca (2007) proposed a common "reasonable frontier", based on the concept of super-efficiency (see, for example, Andersen and Petersen, 1993), to set incentive-related targets for public hospitals in Spain. More recent studies include the paper by Perronnin, Pichetti and Sermet (2014) that ranked 693 French general practitioners according to their DEA scores and the study by Gök and Altındağ (2015) that performed a DEA-based assessment of the P4P programme's effect on efficiency in Turkish public and private hospitals. Both studies used the standard DEA model for the calculation of composite quality measures, while the latter

study also examined the patterns of efficiency change over time by estimating the Malmquist Productivity Index. Likewise, Bastian, Kang, Griffin and Fulton (2016) measured the effect of the incentive programme on hospital efficiency and outcomes in the US military health system, using DEA. This study incorporated a DEA model with time windows and difference-in-differences estimation analysis. Finally, the study by Shwartz et al. (2016) developed a DEA model for estimating composite quality measures in P4P health care programmes with an empirical application to the US nursing homes. The proposed DEA model includes the adjusted opportunity-based weights that reflect the actual results on the explored outcomes to the predicted probability of these outcomes in the US nursing homes.

We believe that our study extends and complements the previous research in several important ways. Firstly, we step back from setting absolute weight restrictions derived from expert opinion, as in the study by Shwartz et al. (2016), or using unrestricted DEA models, as in the study by Perronin et al. (2014) and the study by Gök and Altındağ (2015). Instead of this, we develop a DEA model that has a greater degree of flexibility in choosing optimal virtual weights than any model with fixed numerical thresholds in weight assignment, but that still does not allow the assignment of close to zero weights to any of the inputs or outputs, as opposed to unrestricted DEA models. This is done through the imposition of ordinal and proportional virtual weight restrictions (Sarrico and Dyson, 2004; Wong and Beasley, 1990) and by introducing a parameter that regulates the level of dissimilarity between optimal virtuals (Ramón, Ruiz and Sirvent, 2010).

Secondly, unlike previous research, in this study we explicitly account for the influence of environmental factors in the estimation of DEA scores related to P4P evaluations. We integrate our methodological approach with the composite index proposed by Camanho and Dyson (2006), to verify the relevance of the existing clustering of primary care providers in Portugal in relation to specific structural and populational characteristics. We also perform a cluster analysis to classify these providers according to their environmental characteristics, and then we adopt the procedure by Löber and Staat (2010) to compare each unit only with those peers that operate in the same or worse environmental conditions.

Finally, this study also outlines how to operationalise the proposed DEA methodology in the context of P4P. The suggested DEA methodology opens up new possibilities for the elaboration of several bases for reward with an application of DEA, including rewards that are

linked to absolute and relative performance assessments, and also contributes to setting up learning networks among the providers.

Therefore, the objectives of this study are twofold:

1. To develop a DEA methodology for performance assessment of primary care providers in Portugal. This approach provides a fairer basis for providers' comparison, as it allows each unit to be evaluated with a greater emphasis on the PIs on which the unit under assessment performs better, whilst accounting for the trade-offs between the different PIs and domains. In this respect, the proposed approach compares each unit with a group of other units that operate under similar conditions and with similar performance priorities, identifying those providers that are best performers and identifying targets for performance improvement for underperforming units. This is achieved by modifying the standard DEA method in order to account for the necessary relations between the weights attributed to each PI and each performance domain and in order to account for the effect of the relevant environmental variables.

2. To propose an alternative methodology for P4P contracting based on the DEA assessment results. This methodology includes a combination of relative and absolute performance assessments in order to reinforce the effect of P4P on the Portuguese primary care providers with different performance levels, and each of these two assessments is proposed to be rewarded distinctly. Besides the P4P target setting, we propose how the DEA assessment results can be used in specific measures within the context of P4P in order to better stimulate the improvement of USFs' performance.

The remainder of the paper is organised as follows. Section 2 provides background to the P4P programme in the Portuguese primary care system and discusses the potential for improvement of P4P contracting in Portugal. Section 3 proposes an alternative DEA-based approach to assessing the performance of the Portuguese primary care providers and a new methodology for P4P contracting. Section 4 reports the results of the DEA assessment of USFs and compares the actual results, the contracted targets, and the suggested DEA targets for some USFs. Finally, section 5 draws the main conclusions from the paper, identifies the main limitations and suggestions for further research.

3.2 The P4P programme in the Portuguese primary care system

3.2.1 Organisational structure of the primary care system in Portugal

The primary care system is an entering point into the Portuguese National Health Service (NHS). It aims to provide health care services to patients with non-emergency health concerns, patients with chronic conditions, and also plays a central role in disease prevention and health promotion for local populations. The majority of primary care services in the Portuguese NHS are delivered by general practitioners (GPs)/family doctors who are assisted by nurses and other auxiliary health care staff.

The organisational structure of the primary care system in Portugal has been undergoing a major reform, initiated by the Ministry of Health in 2005. Following this reform, Family Health Units (Unidades de Saúde Familiar, USFs) have been introduced into the Portuguese NHS as local providers of primary care services. At the moment, both newly established USFs and earlier established Primary Health Care Units (PHCUs) are operating in the health care system in Portugal, and there is a roughly equal split between the two models with regards to the population covered by each model (OECD, 2015).

PHCUs and USFs have different organisational forms, level of autonomy, and payment mechanisms. PHCUs are clinic settings that comprise varying number of GPs, who are paid a fixed salary and equally service their patient lists or patients not registered with them. USFs, in turn, are voluntary organised multiprofessional teams that consist of two to twelve family doctors, the same number of family nurses, and a varying number of administrative staff members. Unlike PHCUs, USFs have functional and technical autonomy and a payment system that adopts P4P, and USFs can only provide services to patients from their registered patient lists. Also, there are two operational models of USFs – Model A and Model B. All USFs start as Model A, as it corresponds to the initial phase of USF's development. Once greater organisational maturity has been achieved, USFs Model A can progress to Model B (OECD, 2015). By the end of 2015, there were 449 USFs, of which 241 USFs were classified as Model A and 208 USFs were classified as Model B (ACSS, 2015).

One of the key distinctions between USFs Model A and USFs Model B is in their payment systems. In USFs Model A, the staff payment is mostly a fixed salary regulated by public administration legislation, but a financial incentive is also provided to the whole USF

for accomplishing specific incentive targets. In this model, there is no reward for individual staff members for performance on PIs. In USFs Model B, the staff payment consists of a small fixed component and a series of supplements that represent the largest part of the payment. These supplements include a capitation based payment with a defined maximum threshold, a fee-for-service payment for home visits and other contracted services, and a P4P component that is based on the attainment of individual staff members' targets and practice targets (OECD, 2016).

Another organisational change that was introduced by the primary care reform is the creation of Groups of Health Centres (Agrupamentos de Centros de Saúde, ACES). These centres work as local health authorities with administrative and financial autonomy and they aim to guarantee the provision of primary care services to the population in a determined geographic area. ACES encompass different units, including USFs, Personal Health Units (Unidades de Cuidados de Saúde Personalizados, UCSPs), Community Health Units (Unidades de Cuidados na Comunidade, UCCs), and Shared Assistance Resource Units (Unidades de Recursos Assistenciais Partilhados, URAPs) to provide services in the area of primary health care, social and public care, complementing the actions of each other (OECD, 2015; ACSS, 2016). According to the official statistics, 55 ACES have been operating since 2012 (ACSS, 2015).

Finally, at the highest levels of the organisational hierarchy, there are five Regional Health Administrations (Administrações Regionais de Saúde, ARS) that are responsible for the provision of primary care services in five regions of Portugal: North, Lisbon and Tagus River, Centre, Alentejo, and Algarve. The work of the ARS, in turn, is coordinated by the Central Health System Administration (Administração Central do Sistema de Saúde, ACSS).

The P4P programme that was introduced as a part of the primary care reform links three levels of organisational structure of the Portuguese primary care system through several types of contracting arrangements. In particular, there is an external contracting process that occurs between the ARS and the respective ACES and an internal contracting process that takes place between ACES and local primary care providers (ACSS, 2016). In this paper we will focus on the internal P4P contracts with the USFs.

3.2.2 Conceptual and methodological basis of P4P contracts with USFs

The P4P contracts with USFs are designed to improve quality, accessibility, and efficiency of primary care services in the Portuguese NHS. In essence, they present annual arrangements between an ACES and a USF to be fulfilled by the respective USF in the prioritised areas of primary care delivery.

The P4P contracts in 2013 covered four performance domains - accessibility, clinical performance, efficiency, and patient satisfaction – with a total of 32 PIs. These PIs are organised into two subsets – 15 PIs related to institutional incentives (hereinafter referred to as "institutional PIs") (Table 3.1) and 17 PIs related to financial incentives (hereinafter referred to as "financial PIs") (Table 3.2) (ACSS, 2012).

Table 3.1 Description of the institutional PIs

Domain	**Indicator(s)**
Accessibility	• Proportion of consultations provided by a designated family doctor • Global utilisation rate of medical consultations • Rate of home visits by the doctor (per 1000 patients) • Rate of home visits by the nurse (per 1000 patients)
Clinical performance	• Percentage of patients with hypertension with at least one record of blood pressure in each semester • Percentage of women aged 25-64 who have had a colpocytology in the past 3 years • Percentage of diabetics with at least 2 HbA1C tests registered during the past 12 months and covering 2 semesters • Percentage of first consultations in the first 28 days of the newborn's life • Percentage of fully immunised 2-year-old children in accordance with the National Vaccine Programme • Percentage of first consultations in the first trimester of pregnancy
Patient satisfaction	• Percentage of satisfied and very satisfied patients
Efficiency	• Average cost of medications prescribed per user • Average cost of complementary diagnostic and therapeutic means prescribed per user
Accessibility/Clinical performance	• 2 PIs identified by the ARS

Table 3.2 Description of the financial PIs

Domain	**Indicator**
Clinical performance	• Proportion of family planning consultations provided by the nurse • Percentage of women aged 25-49 who have had a colpocytology in the past 3 years • Percentage of pregnant women with 6 and more maternal health consultations provided by the nurse • Percentage of postpartum home visits by the nurse • Percentage of pregnant women who have received a postpartum examination • Percentage of neonatal heel pricks (Guthrie test) performed by the 7th day of the newborn's life

Table 3.2 Description of the financial PIs *(continued)*

Domain	Indicator
	• Percentage of newborns who have received a home visit by the nurse by the 15th day of postnatal life • Percentage of babies with at least 6 infant health consultations provided between 0 and 11 months of the child's life • Percentage of children with at least 3 infant health medical consultations during the second year of the child's life • Percentage of 2-year-old children with at least one weight and height record in the past 12 months • Percentage of fully immunised 2-year old children in accordance with the National Vaccine Programme • Percentage of diabetics who have received a nurse consultation • Percentage of diabetics who have had at least one foot examination in the year • Percentage of diabetics with a record of therapeutic regime management (three items) in the past year • Percentage of patients with hypertension with at least one record of blood pressure in each semester • Percentage of patients with hypertension with a record of Body Mass Index in the past 12 months • Percentage of patients with hypertension aged 25 and more who have received a vaccination against tetanus

The two subsets of PIs have several distinctions. In particular, the institutional incentives are designed to fulfill specific organisational objectives, such as infrastructure development or completion of specified training by staff, and both USFs Model A and Model B are eligible for this incentive payment. The financial incentives, in turn, are attributed to the nurse and administrative staff exclusively, and only USFs Model B can benefit from this type of incentive. Another distinction is that the financial PIs are uniform across all USFs and they explicitly focus on the assessment of clinical performance (Table 3.2). By contrast, the institutional PIs cover all four domains and they included a fixed subset of 13 nationwide PIs and a variable subset of two PIs that were defined by the respective ARS in 2013 (Table 3.1).

The P4P contracts with USFs stipulate absolute targets on the institutional and financial PIs. These targets are revised annually through a negotiation process between the ACES and each USF. For the target setting purpose, the following information is considered:

- the reference value on a PI from the local, regional, and national documents on operational and strategic planning in health care delivery;
- the history of performance results on a PI of the respective USF, and also at regional and national levels;
- the resources available in the respective USF and the influence of context variables on health care delivery;

- best practices in health care delivery, such as the improvement of accessibility, patient satisfaction, clinical performance, and efficiency (ACSS, 2016).

In order to facilitate the process of target setting, the USFs' past performance results are clustered in relation to three structural characteristics (the number of doctors, the number of years in operation (only for USFs Model A), and the model type) and three populational characteristics (the proportion of patients aged 65 and more, the population density of the respective municipality, and the proportion of diabetics in the patient list). As a result of this analysis, six different sets of cluster groups are formed to classify USFs against the above mentioned characteristics. Then, for each cluster group, the range of target values on PIs is derived to be used in further negotiation of the exact targets, established for every USF individually.

When the contracting year is finished, the assessment of target attainment is executed separately for the institutional and financial PIs. For the financial incentives, the USF's actual performance results on each PI are compared with three absolute thresholds. These thresholds are labeled as "achieved" (more than 90% of the target value is attained), "almost achieved" (from 80% to 90% of the target value is attained), and "not achieved" (less than 80% of the target value is attained). Then, a 2-point scoring system is applied for the calculation of a total achievement score against the financial PIs. In particular, 2 points are assigned for each financial PI that has been fully achieved, 1 point is assigned for each financial PI that has been partly achieved, 0 points are assigned if the target has not been achieved. A total of 30 to 34 points makes a USF eligible to 100% of the financial incentive payment, while a total of 25 to 29 points is sufficient in order to receive 50% of the payment. There is no financial incentive for USFs that score less than 25 points.

For the institutional incentives, a composite measure of performance, known as Global Performance Index (Índice de Desempenho Global, IDG), is calculated. At the first stage of analysis, similarly to the financial incentives, the USF's results on each institutional PI are compared with the contracted targets. Then, depending on a level of target attainment, the result on a PI is substituted with 110% (if more than 110% of the target value is achieved), or it remains the same (if 80% to 110% of the target value is achieved), or the adjusted result equals to 0 (if less than 80% of the target value is achieved). At the second stage of analysis, the adjusted results on PIs are assigned with fixed relative weights for the calculation of the IDG. Finally, the value of the IDG is calculated and it determines whether a USF will receive

the institutional incentive or not. In particular, an IDG lower than 75% does not allow a USF to receive the incentive, an IDG that is ranged from 75% to 100% multiplies the maximum bonus payment on its value, and an IDG of 100% makes a USF eligible to the maximum size of institutional incentive payment (ACSS, 2016).

3.2.3 Potential for improvement of P4P contracting with USFs

As described in the previous section, the P4P contracts with USFs implemented in the Portuguese NHS have the following characteristics:

- they include absolute targets with several absolute thresholds for the evaluation of target achievement that do not reward any improvement "above the target";
- P4P targets are established exogenously, on the basis of a complex system of the reference values and through a negotiation process with the respective ACES;
- the evaluation of target achievement is performed for each PI separately and a fixed weighting system is applied for the calculation of composite performance scores for the institutional and financial incentives.

Each of these characteristics presents the potential for improvement that we discuss below in detail.

Absolute performance assessment and its alternatives. The use of absolute targets for measuring and rewarding performance of health care providers may be considered to have advantages and disadvantages. On the one hand, absolute targets are generally more favoured in P4P programmes due to their transparency and less uncertainty, as they establish, in advance, a direct standard of expected results and set up clear goals to work toward (Cromwell et al., 2011; Eijkennar, 2011). On the other hand, absolute targets are frequently exposed to the effect of goal gradient, when little effort for improving performance can be expected if an established goal is unattainable or it has been already achieved (Mehrotra, Sorbero and Damberg, 2010). If we look again at the methodology for P4P evaluations of USFs, we can see that there is no additional reward for achieving more than 110% (or 90%) of the target value on an institutional (or financial) PI. For this reason, we might expect the existence of the effect of goal gradient at these levels of target attainment.

To avoid the effect of goal gradient, relative performance assessment may be implemented. Unlike the use of absolute targets, relative assessment of performance does not rely on a priori target values that are known in advance. By contrast, this type of assessment constructs a ranking of the incentivised units and rewards only top performers among those participating in a P4P programme. However, although relative rankings stimulate continual improvement, it has been argued that they may also encourage competition, reduce collaboration, and dissemination of best practices among the providers (Cromwell et al., 2011; Eijkenaar, 2011). Furthermore, when a relative ranking is applied, there might be a situation when high levels of target achievement are not rewarded, if the distribution of performance scores is narrow among top performers, corresponding to a situation in which the difference in performance achievement of top providers is not statistically significant (Doran, Kontopantelis, Reeves, Sutton and Ryan, 2014).

The third approach to measuring and rewarding performance involves the measurement of change in PIs over time. This is done by comparing actual performance results with past performance results on the same measures. Similarly to relative targets, this approach encourages continual improvement by avoiding the all-or-nothing property of absolute thresholds. This approach has the added advantage of eliminating the need for complex case-mix adjustments. However, improvement measurement can also favour those providers that have the worst performance results at the beginning of a P4P programme, as they possess the largest scope for improvement (Cashin et al., 2014; Cromwell et al., 2011).

As one can see, each of the approaches has its own advantages and limitations. In practice, as concluded by Cashin et al. (2014), 50% of the examined P4P programmes in OECD countries favour absolute targets, 17% of the studied programmes use relative rankings, and the remaining 33% of the studied programmes use a combination of several approaches.

Despite the limited use of relative rankings in the present P4P programmes, the methodology for P4P evaluations proposed in this study assumes relative performance assessment. As opposed to a ratio analysis, our DEA-based methodology suggests the assessment of all PIs within a single optimization procedure and takes into account the best practice achievements of other providers with similar characteristics and similar strategic priorities. In this way, it guarantees a fairer and holistic comparison of health care providers. Moreover, the DEA approach does not encourage competition or reduce cooperation between

health care providers, since each unit with potential for improvement has its unique combination of performance targets determined by peer DMUs, by contrast to relative rankings that assume the assessment of performance against a valuation system that is common to all providers. Another advantage of our approach is that it does limit the number of top performers, whilst relative rankings based on a ratio analysis usually define a single best performing unit. In this way, we can ensure that all providers that have no potential for improvement are assigned with 100% performance scores, avoiding the problems associated with P4P in cases of a narrow distribution of top performance scores in relative rankings, mentioned by Doran et al. (2014).

Target setting. The involvement of health care professionals in P4P target setting is recognised to bring certain benefits. In particular, this involvement helps to elicit necessary information about contextual factors that influence the attainability of the proposed targets and also enhances providers' motivation for target achievement in P4P programmes. However, when key users are involved in such discussions, biased decisions can be made due to an existing conflict of interests. This is particularly true if there is no formalised and systematic process for the negotiation of P4P targets (Kirschner et al., 2012; Doran et al., 2014).

To mitigate the problem of potential bias, a system of target reference values was designed to support the negotiation of P4P targets between the ACES and the USFs. As a result of cluster analysis, there are different sets of cluster groups formed for each structural and populational characteristic, and each of these clusters has its own range of target values. We do not have access to information about the cluster method used or the way how these ranges of target values are derived. However, according to the ACSS methodology (ACSS, 2016), there is no clear hierarchy that ranges the importance of the obtained clusters of USFs or the respective ranges of target reference values, which significantly limits the functionality of these clusters for the purpose of target setting. In order to overcome this problem, in this study we aim to derive a single set of target reference values that accounts for all relevant contextual factors.

The current methodology used in the Portuguese NHS also suggests that USFs' past performance results have direct influence on subsequent targets on PIs. In particular, the identification of target reference values in 2016 was based on the following principle: the potential for improvement in 2016 for each USF is greater if a low performance level was

achieved by the corresponding USF in 2015, and this potential is identified by cluster membership of the unit under assessment (ACSS, 2016). This principle has been developed in order to avoid situations characterised by USFs which are not willing to achieve the maximum possible level of performance in one year in order to have less demanding targets in the following year. However, the present methodology does not provide a quantitative method to verify if the targeted levels of performance are realistic and feasible for USFs. The feasibility of the proposed targets is determined through negotiation between ACES and USFs, and in such situations, the bargaining power of the incentivised units may become the principal factor for deciding on individual P4P targets.

By contrast, our DEA-based approach suggests that P4P targets are derived endogenously, from the output levels of the identified benchmarks (if performance improvement is possible). The proposed targets present the maximum level of output production, considering current technology, available resources, and environmental factors. In this way, we can avoid the subjectivity of establishing targets through a negotiation process and confirm their feasibility based on the observed performance of benchmarks with similar characteristics.

Composite measures of performance and related weighting systems. Composite measures are commonly used in P4P programmes to aggregate the results on multiple PIs into a single metric of performance. However, the design of meaningful and reliable composite measures is still challenging, as it requires careful consideration of a number of theoretical and methodological aspects (Nardo et al., 2008; Goddard and Jacobs, 2009).

One of the aspects that considerably affects the resulting score on the composite measure is the weighting system used to combine the PIs into a single measure. At a broad view, we can distinguish two approaches for weightings: equal weighting and differential weighting. Equal weighting of PIs is an option that is easy to interpret, as it suggests that all PIs have equal importance in the composite. Differential weighting of PIs, in turn, aims to capture the underlying system of values associated with the measurement of different aspects of performance. Differential weighting of PIs is essential if greater organisational effort is needed in particular areas of performance, although in practice it may be difficult to reach a consensus regarding the relative importance of each PI (Jacobs, Smith and Goddard, 2004).

Additionally, we can distinguish between fixed and flexible weighting systems. A fixed weighting system implies the assignment of weights that are uniform across all units under

assessment, while a flexible weighting system includes the possibility of modifying weights of PIs for each provider individually.

The calculation of the IDG, as proposed by the ACSS methodology (ACSS, 2016), suggests the application of a differential fixed weighting system. Relative weights of the institutional PIs are defined by the ACSS in collaboration with health care experts, and these weights are fixed for all USFs. In particular, among the 10 institutional PIs on accessibility and clinical performance that are common for all USFs, and that will be further explored in this study, the greatest weight is assigned to the global utilisation rate of medical consultations (21.88% of the total weight assigned to the 10 PIs), followed by the weight assigned to the percentage of women aged 25-64 who have had a colpocytology in the past 3 years (18.75% of the total weight assigned to the 10 PIs). The rates of home visits by the doctor and by the nurse are weighted equally, each assigned 12.5% of the total weight. Similarly, an equal weighting of 6.25% each was assigned to the following five PIs: the proportion of consultations provided by a designated family doctor; the percentage of diabetics with at least 2 HbA1C tests registered during the past 12 months and covering 2 semesters; the percentage of fully immunised 2-year-old children in accordance with the National Vaccine Programme; percentage of first consultations in the first 28 days of the newborn's life; and the percentage of first consultations in the first trimester of pregnancy. Finally, the least valued PI, receiving only 3.13% of the total weight attributted to the 10 PIs, is the percentage of patients with hypertension with at least one record of blood pressure in each semester.

Although this fixed weighting system aims to reflect the varying complexity, associated priorities and benefits of the undertaken activities, it is likely to be inherently restrictive due to its inflexibility. The assignment of fixed weights to PIs does not respect the existing diversity in performance preferences held by policy makers, health care managers and providers, and the general public among other stakeholders (Smith, 2002).

By contrast, the standard DEA model allows for flexibility in choosing weights assigned to each input and output of the DMU under assessment in such a way that it maximises the performance score of the respective DMU. The benefit-of-the-doubt approach to weighting, rooted in DEA, as discussed by Cherchye, Moesen, Rogge and Puyenbroeck (2007), suggests that more (or less) weight should be attributed to those PIs in which the unit performed better (or worse), assuming that better performance results were achieved on PIs that are more

important for the DMU under assessment. This means that the weights that are applied to the data are actually based on the observed performance data, or determined endogenously. However, an unrestricted benefit-of-the-doubt approach to weighting may be an undesirable feature of DEA, when a close to zero weight is assigned to some PIs or even to the majority of PIs. In this situation, all PIs with a close to zero weight will be almost ignored for the calculation of composite measures, and, therefore, the obtained results can be considered as unrealistic.

There are different approaches to overcoming this problem. The most common solution described in the DEA literature involves the imposition of weight restrictions in DEA models, such as absolute multiplier restrictions (Roll, Cook and Golany, 1991), cone ratio restrictions (Charnes, Cooper, Huang and Sun, 1990), assurance regions (Thompson, Langemeier, Lee, Lee and Thrall, 1990; Thompson, Singleton, Thrall and Smith, 1986), and facet models (Green, Doyle and Cook, 1996; Portela and Thanassoulis, 2006), among other methods. In this study, we will use a DEA model that incorporates several types of weight restrictions, such as ordinal and proportional virtual weight restrictions (Sarrico and Dyson, 2004; Wong and Beasley, 1990) and introducing a parameter that regulates the level of dissimilarity between optimal virtual weights (Ramón et al., 2010), in order to best adopt the DEA methodology for the use in the context of P4P evaluations of USFs. A detailed description of the proposed approach is presented in the next section.

3.3 Methods

3.3.1 A DEA-based approach to assessing the performance of USFs

DEA is a non-parametric method that derives a single relative performance score for each unit when compared with a set of homogeneous DMUs. DEA estimates the relative performance of a DMU by the extent to which the unit under assessment matches or falls short of the expected performance level identified by the production frontier.

There are a number of the considerations involved in the development of a DEA model, including the most appropriate orientation towards the frontier and regarding the most appropraite returns to scale assumption (Jacobs et al., 2006b). With respect to the orientation

to the frontier, an input-oriented DEA model aims to minimise the quantity (or value) of the consumed inputs, while an output-oriented DEA model aims to maximise the quantity (or value) of the produced outputs. Regarding the returns to scale assumption, constant returns to scale assume that changes in the inputs result in proportional changes in the outputs, while variable returns to scale assume that an increase (or decrease) in the inputs does not necessarily lead to proportional changes.

The mathematical formulation of the output-oriented DEA model, developed by Charnes et al. (1978), in its multiplier version and assuming constant returns to scale, is presented below:

$$E_o = \min \sum_{i=1}^{m} v_i x_{i0} \qquad (1)$$

$$\text{s.t. } \sum_{r=1}^{s} \mu_r y_{r0} = 1,$$

$$\sum_{i=1}^{m} v_i x_{ij} - \sum_{r=1}^{s} \mu_r y_{rj} \geq 0, \quad j=1,\ldots n,$$

$$v_i, \mu_r \geq \varepsilon > 0, \quad i=1,\ldots m, \; r=1,\ldots s$$

In this formulation, we assume that each DMU_j consumes m inputs (x_{ij}, i=1,…,m) to produce s outputs (y_{rj}, r=1,….,s), while v_i and μ_r are the optimal pure weights attributed to the inputs and outputs respectively. Besides the optimal pure weights, we can also distinguish the optimal virtual weights that represent the product of the inputs (or outputs) by the respective optimal pure weights. The objective of the programme is to maximise the relative performance score of a given DMU_0, calculated as $\frac{1}{E_o}$, by finding the positive optimal weights v_i and μ_r for each input and output included into the model.

The calculation of composite measures under the benefit-of-the-doubt approach suggests the use of a DEA model, which includes a dummy input (or output) equal to one and multiple outputs (or inputs) to represent individual PIs to be combined into the composite (Cherchye et al., 2007). To date, there have been a number of DEA applications in different fields that have used a benefit-of-the-doubt weighting approach to constructing composite measures. Examples of such applications are the estimation of urban quality of life (Morais and Camanho, 2001), the measurement of human development of countries (Despotis, 2005),

the measurement of the level of technological achievement and capabilities of countries (Filipetti and Peyrache, 2011; Cherchye, Moesen, Rogge, Van Puyenbroeck, Saisana, Saltelli, Liska and Tarantola, 2008), the measurement of countries' environmental performance (Zanella, Camanho and Dias, 2013), and the measurement of the level of active ageing of countries (Amado, São José and Santos, 2016), among others. By adopting the benefit-of-the-doubt approach, with a single dummy input, the formulation in (1) reduces to model (2):

$$E_o = \min v \qquad (2)$$

$$\text{s.t. } \sum_{r=1}^{s} \mu_r y_{r0} = 1,$$

$$v - \sum_{r=1}^{s} \mu_r y_{rj} \geq 0, \quad j=1,\ldots n,$$

$$v, \mu_r \geq \varepsilon > 0, \quad r=1,\ldots s$$

In this study, we will use a DEA model with a constant input equal to one and 10 outputs that correspond to the fixed set of 10 institutional PIs on accessibility and clinical performance, as presented in Table 3.1. All outputs are formulated as "more is better" and are measured as a percentage. In this respect, only a minimal transformation of two variables was performed (converting the rate of home visits by the doctor and the rate of home visits by the nurse from percentiles into per cent terms). In consistency with previous works by Zanella et al. (2013) and Amado et al. (2016), the proposed model has an output orientation in order to estimate the maximum possible level of target attainment on all PIs, identified by the production frontier.

In this model, we have imposed several types of weight restrictions to diminish the total flexibility in the assignment of relative output weights. Whilst aiming to achieve a fairer distribution of weights by allowing some level of flexibility, our weight restrictions aimed to respect existing value judgements used in the calculation of the IDG.

The approach we have developed to restrict the weights involves imposing three types of weight restrictions aimed at: 1) restricting the dissimilarity in the weights assigned by different units; 2) respecting the order of weighting established in the current NHS system; and 3) balancing the distribution of weights between accessibility measures and clinical performance measures. To achieve the first objective, we adopted a multiplier bound

approach, proposed by Ramón et al. (2010), in order to deal with close to zero weights for the optimal multipliers. Previous DEA studies on constructing composite measures frequently relied on weight restrictions, derived from an exploratory analysis with a DEA model (e.g. Zanella et al., 2013). However, the presence of alternative optimal weight solutions, as previously discussed by Cooper, Ruiz and Sirvent (2007), makes it difficult to justify any specific weighting scheme obtained in such a way.

The approach by Ramón et al. (2010) noticeably differs from more traditional weight restriction approaches and facet models, mentioned earlier, as it does not require any external information or related value judgements for the definition of weight boundaries or the existence of full dimensional efficient facets on the frontier. Moreover, the developed model does not lead to infeasibility problems or problems in choosing between the alternative optimal solutions. Finally, as argued by Ramón et al. (2010), it not only avoids the problem of close to zero weights, but also avoids extreme dissimilarity in the optimal weighting schemes to the maximum possible extent. Considering these desirable properties, we believe that the approach developed by Ramón et al. (2010) is the most suitable to achieve our purpose to restrict the total flexibility in weight assignment in the context of P4P evaluation of USFs.

For an application of the approach by Ramón et al. (2010), it is important to identify the extreme efficient units (i.e. units, for which there is at least one optimal solution where none of the variables was assigned a weight approximately equal to zero) by running the standard CCR model. After that, at the first step of analysis, we maximise the minimum of the ratios between virtuals of the extreme efficient units, so that we look for the least dissimilar optimal virtual weights that allow the extreme efficient units (i.e. DMUs in E) to remain as such. At this stage, the following problem has to be solved:

$$\max \varphi_{jo}=\frac{z}{h} \qquad (3)$$

$$\text{s.t. } v=1,$$

$$\sum_{r=1}^{s} \mu_r y_{r0}=1,$$

$$v-\sum_{r=1}^{s} \mu_r y_{rj} \geq 0, \quad j \epsilon E$$

$$z \leq \mu_r y_{r0} \leq h, \; r=1,\ldots,s$$

$$\mu_r, z, h \geq \varepsilon >0, \; r=1,\ldots,s$$

In this formulation, the level of dissimilarity between the virtuals of the extreme efficient units is measured by a parameter φ_{jo}. If $\varphi_{jo}=1$, then all virtuals have the same value, which is equivalent to a system of equal weighting. If $\varphi_{jo}=0$, this means that the virtuals are free to be assigned with any value and most probably, close to zero weights will be attributed to some of the variables. However, in case of the extreme efficient units, there is at least one solution with non-zero optimal weights, so there is no possibility to get $\varphi_{jo}=0$ by solving model (3).

At the second stage of analysis, we apply φ^*, which is the minimum across all φ_j, $j \epsilon E$, to underperforming DMUs (all those that are not extreme efficient) by solving the following problem:

$$E_o = \min v \qquad (4)$$

$$\text{s.t. } \sum_{r=1}^{s} \mu_r y_{r0}=1,$$

$$v - \sum_{r=1}^{s} \mu_r y_{rj} \geq 0, \; j=1,\ldots n,$$

$$z \leq \mu_r y_{r0} \leq h, \; r=1,\ldots,s$$

$$\frac{z}{h} \geq \varphi^*$$

$$v, \mu_r, z, h \geq \varepsilon > 0, \; r=1,\ldots,s$$

By identifying φ^*, we preserve the original distribution between underperforming and extreme efficient units, identified by the standard CCR model and we also regulate the dissimilarity between the optimal virtual weights of all DMUs. It is important to note that although the φ^* value is fixed for a given set of DMUs, each DMU is free to select the optimal virtual weights on the condition that the minimum ratio between them is not lower than φ^*. This is a clear advantage of this approach, as we step back from imposing absolute boundaries, between which the virtuals are forced to lie, and only set weighting limits in form of φ^*, which is derived endogenously, from the dataset explored.

To achieve our second objective for the weighting system, we developed a set of restrictions to account for the existing value judgements regarding the relative importance of the 10 institutional PIs. Although the current ACSS methodology (ACSS, 2016) provides the exact weighting of each PI in the calculation of the IDG, it also clearly sets out that "we

cannot say that an indicator weighted with "4" is twice as complex or requires two times more work than an indicator weighted with "2". We can only say that the indicator weighted with "4" is probably more complex or the activity, which it reflects, requires more resources or it is more effective in producing health gains than the one weighted with "2" [translated from Portuguese]" (ACSS, 2016: 47). In this respect, in order to capture an ordering of importance of the included PIs, rather than to refer to the exact values of the weights, we formulated ordinal virtual weight restrictions, or virtual assurance regions of type I (virtual ARI restrictions), as proposed by Sarrico and Dyson (2004). Particularly, we ordered the outputs from the most valued to the least valued PI, as suggested by the weights proposed in the ACSS methodology. In this way, we were able to identify five groups of virtuals with differing degrees of relative importance, given that PIs with equal weighting were assigned to the same group. The formulation of the set of restrictions that was applied when evaluating DMU_0 is presented below:

$$\mu_2 y_{2o} \geq \mu_6 y_{6o} \geq \mu_3 y_{3o}, \mu_4 y_{4o} \geq \mu_1 y_{1o}, \mu_7 y_{7o}, \ \mu_8 y_{8o}, \mu_9 y_{9o}, \ \mu_{10} y_{10o} \geq \mu_5 y_{5o} \qquad (5)$$

As one can see, the formulated ordinal virtual weight restrictions have only been applied to the DMU under assessment to obtain a reasonable weighting system specific for each USF. Sarrico and Dyson (2004) recognised such approach to be a better alternative to applying restrictions to all DMUs, as it helps to avoid the infeasibility problem. However, one needs to be aware of the limitations of imposing virtual restrictions only to the DMU under assessment – each DMU will be assessed with a different set of restrictions, which means that rankings cannot be elaborated based on the results.

In order to achieve our third objective for the weighting system, we developed another virtual weight restriction. Considering that the explored PIs belong to different dimensions – accessibility and clinical performance – we aimed to regulate the magnitude of relative contribution of each of the performance domains to the resulting score. We believe that target achievement on both accessibility and clinical performance is equally important, as we could not find scientific evidence that would support the opposite view. For this reason, we imposed proportional virtual weight restrictions, as proposed by Wong and Beasley (1990), to ensure that the sum of the virtual weights attributed to the four PIs on accessibility equals the sum of the virtual weights assigned to the six PIs on clinical performance. In fact, if we look at the distribution of relative weights between the 10 PIs defined by the ACSS methodology, we can

see that the sum of the weights in both domains is approximately equal (53.13% of the total weight is attributed to the PIs on accessibility and 46.87% of the total weight is attributed to the PIs on clinical performance). The formulation of the restriction applied when evaluating DMU_0 is presented below:

$$\sum_{r=1}^{4} \mu_r y_{r0} = \sum_{r=5}^{10} \mu_r y_{r0} \qquad (6)$$

Following the approach used in the previous case, this restriction was only applied to the DMU under assessment.

3.3.2 Accounting for the influence of environmental factors

In performance assessment it is important to consider the effect of environmental factors, given that the attainment of performance targets is more difficult for the units that operate in adverse environmental conditions. However, as pointed out by Jacobs, Smith and Goddard (2004), "there is no generally accepted method for taking into account environmental variables at the level of the composite scores, or for testing whether an environmental variable has a significant influence on the production process and the resultant performance of the unit"(Jacobs et al., 2004: 59).

In order to test the influence of the earlier mentioned structural and populational characteristics of USFs on DEA scores, we applied the approach proposed by Camanho and Dyson (2006). This approach suggests the computation of a Malmquist index that comprises two components: first component allows assessing performance differences between groups of DMUs, and second component measures the gap between the frontiers identified for these groups. In this way, it is possible to verify if different groups of DMUs have similar performance levels, measured to their group frontiers, and also to test whether these groups of DMUs share the same frontier or not. This means that the effect of environmental variables can be identified by a significant gap between the frontiers of different groups.

Assume that we have N_A DMUs belonging to Group A that consume inputs X^A to produce outputs Y^A , and N_B DMUs belonging to Group B that consume inputs X^B to produce outputs Y^B. Each DMU_j from Group A can be then represented by its input-output vector (X_j^A,

Y_j^A) that in case of the benefit-of-the-doubt model transforms into (1, Y_j^A). A similar notation (1, Y_j^B) can be used for DMUs in Group B. In this case, $\frac{1}{E^A}(1, Y_j^B)$ represents the measure of relative performance for a DMU belonging to Group B with regards to the frontier of Group A. The Malmquist Index (MI) can be formulated as the following:

$$MI^{AB} = \underbrace{\frac{\left[\prod_{j=1}^{N^A} \frac{1}{E^A}(1, Y_j^A)\right]^{\frac{1}{N^A}}}{\left[\prod_{j=1}^{N^B} \frac{1}{E^B}(1, Y_j^B)\right]^{\frac{1}{N^B}}}}_{\text{Within-group performance spread (IE}^{AB})} \times \underbrace{\left[\frac{\left(\prod_{j=1}^{N^A} \frac{1}{E^B}(1,Y_j^A)\right)^{\frac{1}{N^A}}}{\left(\prod_{j=1}^{N^A} \frac{1}{E^A}(1,Y_j^A)\right)^{\frac{1}{N^A}}} \times \frac{\left(\prod_{j=1}^{N^B} \frac{1}{E^B}(1,Y_j^B)\right)^{\frac{1}{N^B}}}{\left(\prod_{j=1}^{N^B} \frac{1}{E^A}(1,Y_j^B)\right)^{\frac{1}{N^B}}}\right]^{\frac{1}{2}}}_{\text{Productivity gap between frontiers (IF}^{AB})} \quad (7)$$

It is important to note that in our formulation we use Farrell's measure of performance (Farrell, 1957) that is the reciprocal to the input-oriented distance measure, developed by Färe and Lovell (1978) originally included into the index by Camanho and Dyson (2006). Also, the index can be applied to any number of groups of DMUs, as discussed by Camanho and Dyson (2006). In our study, we tested the difference between two, three, and four groups of USFs in compliance with the groupings suggested by the ACSS methodology (ACSS, 2016).

In this formulation of the index, the first ratio reflects the within-group performance spread by comparing the geometric means of the performance estimates of DMUs, identified by their group frontiers. If this ratio is more than one, then greater consistency in performance levels is observed among DMUs of Group A compared to DMUs of Group B, while the opposite is true if the ratio is less than one.

The second component of the index estimates the productivity gap between the frontier of Group A and the frontier of Group B. It includes four possible combinations of estimating the performance of DMUs of each group to both of the group frontiers. A resulting score on this component of more than one indicates a higher performance level of the frontier of Group A in comparison to the one of Group B, while the opposite is true if the ratio is less than one. In addition, if first (or the second) component of the index equals to one, than we can say that there is no difference in the distances to the respective frontiers (or between the performance frontiers) of the two groups.

In this study, the values obtained for the frontier gaps between different groups was used to identify the factors that appear to have a negative impact on the level of productivity (identified by an IF smaller than one). Afterwards, these factors were included in a cluster analysis aimed at forming clusters that are subject to similar conditions.

In the process of verification of the existing clustering of USFs for the purpose of DEA-based assessment, we have identified two important issues that may impede effective target setting for future periods. Firstly, as we have mentioned earlier, USFs' past performance results are directly accounted for in the present clustering of USFs, so that USFs with sub-optimal performance results can be potentially assigned to groups with less demanding targets for the following years. Secondly, each of the six environmental characteristics has been treated separately in the performed cluster analysis, resulting in six different sets of clusters without a clear hierarchical system that relates these clusters for the purpose of target setting.

In this respect, in order to achieve an effective target setting, we decided to run a cluster analysis that will not consider USFs' past performance results for assigning cluster membership to USFs and will also produce a unique set of clusters by considering all relevant environmental factors simultaneously.

Cluster analysis can be performed using various methodologies. In this study, we have tested different methods of cluster analysis, such as two-step clustering, k-means clustering, and several methods of hierarchical clustering (centroid method, Ward's method, and the average linkage within groups). For all methods, where it was relevant, we have opted to choose the squared Euclidean distance as a measure of distance between cases. In addition, we performed validation of the obtained results via two internal indices of cluster evaluation – the Davies-Bouldin index (DB index) (Davies and Bouldin, 1979) and the Calinski-Harabasz index (CH index) (Calinski and Harabasz, 1974). In particular, by minimising DB index, we obtain clusters that are the most distinct from each other, while by maximising CH index, we achieve the maximum separation of the clusters calculated via the average between- and within-cluster sum of squares (Liu, Li, Xiong, Gao and Wu, 2010).

Considering the way the clustering variables were defined, we were able to define an hierarchy of clusters based on how favourable is the environment with regards to performance achievement. In this way, in order to evaluate USFs with the consideration of the differences in their environments, we applied the methodology by Banker and Morey (1986) for incorporating categorical variables in DEA models, using a simplified algorithm of analysis,

suggested by Löber and Staat (2010). The categories used are the clusters that resulted from the cluster analysis. As argued by Löber and Staat (2010), instead of introducing categorical variables, we can run separate DEA models for the following datasets: one run based on observations from a category with the most adverse environmental conditions, followed by another run for obtaining results for DMUs of a category with the second most adverse environmental conditions by including observations from this and previous categories, followed by another run for obtaining results for DMUs of a category with the third most adverse environmental conditions by including observations from the three described categories, and so forth. In this way, we will compare each DMU only with those DMUs that operate in the same or worse environmental conditions.

3.3.3 P4P contracting with USFs based on the DEA assessment results

As we have mentioned earlier, there are three different approaches to incentivising health care providers, which are based on the measurement of absolute performance, relative performance, or improved performance respectively (OECD, 2016; Cromwell et al., 2011). These three approaches tend to reward and stimulate different types of performers. For example, the measurement of absolute performance mostly rewards existing top performers, regardless of the fact of whether they have exceeded the contracted targets or not. Relative targets, in turn, provide the greatest incentive to improve for those providers who are ranked close to the top performers, whilst top and poor performers are motivated to a lesser extent to enhance their performance results. Finally, if improved performance is rewarded, any type of performers can be entitled to a bonus payment, on the condition that these providers have improved their performance compared to the previous results, although low-performing providers may find it easier to earn a reward in a long term perspective, as they possess the largest scope for improvement (OECD, 2016; Cromwell et al., 2011).

Besides the use of a single approach to rewarding providers' performance, it is possible to create various combinations of these approaches to reinforce their effect on different types of performers. Furthermore, instead of using a single performance benchmark, many of the existing P4P programmes adopt a system of tiered benchmarks in order to evaluate a unit's performance against a broader performance spectrum. One of such systems is actually used in the present P4P contracts with USFs, as USFs' target attainment on institutional and financial PIs is assessed against three absolute thresholds. At the same time, these contracts rely

exclusively on absolute performance assessment for assigning bonus payments to USFs. As a result, such incentive system may not encourage performance improvement among USFs due to the effect of goal gradient and may disguise the maximum performance level of each USF.

As an alternative to the existing approach to rewarding USFs, we suggest adopting the earlier described DEA-based methodology in order to effectively combine absolute measurement with relative measurement in the elaboration of several bases for reward. We believe that a distinct system of graduated bonus payments should be linked to each of the two approaches to performance measurement along with several supplementary rewards for the accomplishment of certain conditions that are detailed below. In particular, it is our conviction that the following methodology can be applied for P4P contracting with USFs:

1. The evaluation of USFs under the suggested DEA-based methodology to determine the best performing units and underperforming units and also to provide bonus payments for a particular year, associated with relative performance assessment, to those USFs that received high DEA scores. In particular, we recommend implementing a system of graduated rewards with a differentiation in the size of a bonus payment according to the achieved performance levels. For example, it is possible to introduce a reward system that provides the full bonus to the best performing USFs, a smaller incentive (for instance, 80% of the maximum bonus) to USFs that scored 80% and more (but not 100%), and no incentive to USFs that scored below 80%.

2. For the underperforming USFs, absolute targets for the following year will be derived from the respective benchmarks, identified by the DEA model results. For the best performing USFs, we suggest the creation of a benchmarking programme that will contribute to transferring knowledge from the best practices to the units that did not score 100%. This programme may include trainings and workshops led by the best performing USFs as a form of establishing learning networks between providers with similar strategic priorities and environmental characteristics. For the development of these trainings and workshops, the best performing USFs should receive an additional reward.

3. The evaluation of USFs' performance results in the following year should include two components. The first component would involve measuring absolute target attainment of the USFs that were previously classified as

underperforming. In particular, it would be possible to adopt the system of tiered thresholds and the 2-point scoring systems that are currently used for performance assessment against financial PIs and described earlier in this study. In particular, for each PI, the attainment of more than 90% of a DEA target could be scored with 2 points, the attainment from 80% to 90% of a DEA target could be scored with 1 point, and no points would be assigned for the attainment of less than 80% of a DEA target. In this case, considering that there are 10 PIs, a total of 20-18 points, 17-15 points, and 14 points and less entitle a USF to 100%, 50%, and 0% of the bonus payment, associated with absolute performance measurement. Furthermore, we can link the performance results demonstrated by underperforming units in the following year with a supplementary bonus attributed to the respective best performing USFs that acted as benchmarks and developed the targeted benchmarking events, if all absolute targets established for the underperforming units have been achieved. The second component of this evaluation repeats the first step of the algorithm (i.e. the evaluation of USFs with an application of the DEA model, definition of the best performing and underperforming USFs, attribution of the bonus payment associated with relative performance assessment to the entitled USFs) and, therefore, makes the whole process recurrent.

The suggested approach to P4P target setting and performance evaluation of USFs has several valuable advantages. In particular, it provides clear and direct standards of the expected performance determined for each underperforming USF individually in the context of absolute performance measurement, so that there is less uncertainty regarding the attainability of the respective reward compared to a reward system that adopts relative rankings. These absolute targets are derived endogenously, on the basis of DEA, and their feasibility is confirmed by the performance levels of the respective best performing USFs. In addition, by implementing relative performance assessment through DEA, USFs are encouraged to show their best possible performance results instead of focusing on the achievement of absolute targets (for USFs classified as underperforming in the previous period) or in the light of the threat of being overtaken (for USFs classified as best performing in the previous period). This means that if the production technology of primary health care adopted by USFs improves over time, the respective benchmarks will automatically increase. Finally, the suggested approach contributes to an effective dissemination of best practices

through the implementation of learning networks between the best performing USFs and underperforming USFs. This is a clear advantage of the approach in the context of P4P, as we would ultimately aim for an overall improvement of USFs' performance rather than merely rewarding current high performers. Furthermore, by providing a reward to the benchmarks related with the performance improvement of their respective underperforming units, we cultivate a culture of collaboration rather than a culture of competition.

3.4 Results

3.4.1 Summary statistics of the data

The data used for this study are the values obtained in each of the PI used for the contracting year of 2013. We have selected 356 USFs that complied with the following criteria:

1. The USFs under selection started their operation before 1 January 2013;
2. The USFs under selection did not cease their activities before 1 January 2014;
3. The USFs' performance data are available with respect to the 10 specified institutional PIs for the contracting year of 2013.

The first two criteria allowed us to choose only those USFs that were functioning 12 months in the year of assessment. According to the ACSS methodology (ACSS, 2016), all USFs that operate more than six months in the year of assessment are eligible to participate in the P4P programme and, therefore, they can be potentially included into the DEA assessment. However, for USFs operating less than one year performance targets are reduced proportionally to a period in operation, and direct comparison of performance results of these units with the remaining USFs would be unfair.

The third criterion was formulated to refine the initial set of PIs for analysis and to deal with missing data. Of the 15 institutional PIs that were identified for the contracting year of 2013, we have selected a set of 10 PIs, as presented in Table 3.3. In particular, we have chosen those PIs that were common for all USFs, eliminating in this way two PIs identified by the ARS for each USF individually for the contracting year of 2013. In addition, we have

excluded the three PIs from the initial set, for which no data were available – one PI on patient satisfaction and two PIs on efficiency, described in Table 3.1.

Table 3.3 Descriptive statistics on the 10 institutional PIs

Domain	Indicator	Ref. value[1]	Mean	SD	Min	Max
Accessibility	• Proportion of consultations provided by a designated family doctor (y1)	85.00%	85.68%	5.54%	62.53%	96.16%
	• Global utilisation rate of medical consultations (y2)	75.00%	71.60%	5.61%	47.39%	83.35%
	• Rate of home visits by the doctor (per 100 patients) (y3)	3.00%	2.72%	1.01%	0.26%	5.99%
	• Rate of home visits by the nurse (per 100 patients) (y4)	14.50%	13.37%	4.14%	1.15%	28.76%
Clinical performance	• Percentage of patients with hypertension with at least one record of blood pressure in each semester (y5)	95.00%	81.77%	14.05%	19.37%	98.24%
	• Percentage of women aged 25-64 who have had a colpocytology in the past 3 years (y6)	60.00%	81.45%	14.51%	18.70%	97.73%
	• Percentage of diabetics with at least 2 HbA1C tests registered during the past 12 months and covering 2 semesters (y7)	85.00%	59.94%	11.84%	20.48%	85.81%
	• Percentage of first consultations in the first 28 days of the newborn's life (y8)	75.00%	96.14%	3.16%	84.98%	100.00%
	• Percentage of fully immunised 2-year-old children in accordance with the National Vaccine Programme (y9)	98.00%	89.59%	8.82%	57.49%	100.00%
	• Percentage of first consultations in the first trimester of pregnancy (y10)	80.00%	88.38%	6.99%	60.62%	100.00%

SD – Standard deviation; Min – Minimum; Max - Maximum

According to official statistics (ACSS, 2015), there were 357 USFs operating by the end of 2012, so we assume that only one USF is missing from our sample to present a complete list of USFs that comply with all three criteria. No other missing data entries were identified.

Additionally, prior to running DEA, the dataset was verified for the presence of extreme values with an application of the outlier labeling rule, or Tukey's method (Hoaglin, Iglewicz and Tukey, 1986; Tukey, 1977). This method is based on multiplying the interquartile range (i.e. the difference between the third and first quartile of the distribution) by a factor of 2.2 and subtracting the resulting value from the first quartile value (or adding it to the third

[1] Reference values used for target setting in 2012 (ACSS, 2012).

quartile value) in order to establish a lower (or upper) boundary, below (or beyond) which any value is considered to be extreme. The detected extreme values were then replaced with the lower bounding value, if the extreme values were located below the lower boundary, or with the upper bounding value, if they the extreme values were located beyond the upper boundary.

Descriptive statistics on target attainment against the 10 institutional PIs for the contracting year of 2013 are presented in Table 3.3. Along with standard summary statistics, such as the means, standard deviations, minimum and maximum values of the indicators, Table 3.3 also includes reference values used for target setting in 2012[2], as we have no access to data on the exact targets established for USFs for the year of 2013.

As one can see from Table 3.3, a substantial amount of variation is observed across the indicators. For instance, there is an evident variability with regards to the average performance results, since some indicators present low average values (such as the rate of home visits by the doctor - 2.72%), while the average for others is close to 100% (such as percentage of first consultations in the first 28 days of the newborn's life - 96.14%). It should be noted, however, that not all accessibility indicators are feasible and desirable at 100% level, such as the case of the rate of home visits by the doctor and the rate of home visits by the nurse, as the relevant reference values were 3% and 14% accordingly in 2012. And the same time, we recognise that, except for these two cases, the majority of institutional PIs, presented in Table 3.3, are designed in a way to cover as many eligible patients as possible. In fact, there was no reference target value for the remaining 8 PIs that was lower than 60% for the year of 2012.

3.4.2 Identifying factors for classifying USFs into clusters with similar environmental conditions

In order to support the process of P4P target setting and evaluation, the present ACSS methodology (ACSS, 2016) assumes the clustering of USFs in accordance with the following characteristics:

[2] These reference values were used for setting exact targets for USFs for the contracting year of 2012; in 2013, they were substituted by other reference values, obtained after a cluster analysis. The targets used in 2013 were not published for all USFs.

1. Structural characteristics:

1a. Number of family doctors working in USF:

- Group A - [0; 5[doctors;
- Group B - [5; 9[doctors;
- Group C - [9; +∞] doctors;

1b. Number of years in operation as USF (only for Model A):

- Group D - [0; 2[years;
- Group E - [2; +∞] years;

1c. Model type:

- Group F- USF Model A;
- Group G - USF Model B;

2. Populational characteristics[3]:

2a. Percentage of patients aged 65 and more in the patient list:

- Group H - [0; 3.87[% of patients aged 65 and more;
- Group I - [3.87; 4.62[% of patients aged 65 and more
- Group G - [4.62; 6.78[% of patients aged 65 and more;
- Group K - [6.78; +∞] % of patients aged 65 and more;

2b. Population density in the municipality, where USF operates:

- Group L - [0; 209[people per km^2;
- Group M - [209; 651[people per km^2;
- Group N - [651; 1696[people per km^2;
- Group O - [1696; +∞] people per km^2;

2c. Percentage of diabetics in the patient list:

- Group P - [0; 15.41[% of diabetics;
- Group Q - [15.41; 17.82[% of diabetics;
- Group R - [17.82; 20.97[% of diabetics;
- Group S - [20.97; +∞] % of diabetics.

The selection of each of these variables for clustering has its own reasoning and might be relevant for DEA assessment of USFs' performance results. In particular, more experienced USFs and USFs Model B may have better performance results due to greater

[3] We re-calculated the ranges of quartiles for each populational characteristic, given that we use data for 2013 and we only consider USFs operating all 12 months in the contracting year

organisational maturity, while greater number of doctors may result in better organisational accessibility of the provider. Patients aged over 65 usually require more health care services, and this non-discretionary variable is frequently accounted for risk adjustment of the DEA scores in primary health care assessments (e.g. Pelone et al., 2012; Ramírez-Valdivia et al., 2011; Milliken et al., 2011; Luoma, 1996). Similarly, the demand for primary health care services might be greater in the areas with high population density, which was also considered in previous studies devoted to the DEA-based performance assessments of primary care providers (e g., Cordero-Ferrera et al., 2011; Ramírez-Valdivia et al., 2011). Lastly, the percentage of diabetics in patient lists serves as a proxy for the level of morbidity of the population covered. As concluded by Starfield, Weiner, Mumford and Steinwachs (1991), patients with several disorders tend to require more consultations per year, which results in less time available for other patients and activities. Therefore, we expect that USFs with lower percentage of diabetics should perform better.

To verify the relevance of the proposed clusters for DEA analysis, we calculated the index suggested by Camanho and Dyson (2006) for each of the six characteristics. For the calculation of the performance scores, we applied model (4) with restrictions (5) and (6) and the value φ^* calculated for the whole dataset for each of the components of the index (7), that was also modified for the case of three and four groups. The resulting scores on the index and its components are recorded in Table 3.4.

Table 3.4 Results on the Malmquist indices

	Number of doctors			Number of years	Model type	% of patients aged 65 and more					
Group comparison	A to B	B to C	A to C	D to E	F to G	H to I	I to G	H to G	H to K	I to K	G to K
Within-group performance spread (IE)	0.9539	1.0066	0.9602	0.9655	0.8506	1.0376	1.0160	1.0542	1.1512	1.1095	1.0920
Productivity gap between frontiers (IF)	0.9536	1.0211	0.9738	0.9215	0.9970	0.9932	0.9851	0.9785	0.9826	0.9893	0.9811
Malmquist Index (MI)	0.9096	1.0279	0.9350	0.8896	0.8480	1.0305	1.0009	1.0315	1.1311	1.0976	1.0714

Table 3.4 Results on the Malmquist indices *(continued)*

	Population density						% of diabetics					
Group comparison	L to M	M to N	L to N	L to O	M to O	N to O	P to Q	Q to R	P to R	P to S	Q to S	R to S
Within-group performance spread (IE)	0.9285	1.0188	0.9459	1.0054	1.0829	1.0629	1.0414	1.0183	1.0604	1.1374	1.0922	1.0726
Productivity gap between frontiers (IF)	1.0199	1.0002	1.0201	1.0077	0.9880	0.9789	0.9661	0.9969	0.9631	0.9748	1.0090	0.9882
Malmquist Index (MI)	0.9470	1.0191	0.9650	1.0131	1.0699	1.0405	1.0060	1.0151	1.0213	1.1087	1.1021	1.0599

In general, the observed results confirm our assumptions and a common understanding regarding the influence of the specified environmental factors on the best practice frontier. In particular, we can say that the number of years in operation has the most significant effect on the frontier, as the biggest productivity gap was observed between two groups of USFs Model A operating less than and more than 2 years (IF^{DE}=0.9215). Furthermore, a negative effect on the level of USFs' productivity results from a higher percentage of patients with diabetes (IF^{PR}=0.9631, IF^{RS}=0.9882) and from a higher percentage of patients aged 65 and more (IF^{HG}=0.9785, IF^{GK}=0.9811). The Malmquist results have also confirmed the expected effect of population density: it has a positive impact on productivity up to a certain density level (around 1696 patients per $km^{2,}$ considering that IF^{LN}=1.0201), but after reaching this limit, the effect of higher population density is negative (IF^{NO}=0.9789). This can be explained by the fact that very high population density is usually associated with deprivation and lower socio-economic background of the population covered, which is known to be an adverse factor for performance in primary health care (Amado and Dyson, 2009). Based on the Malmquist results, we can also say that USFs Model A and Model B have almost identical frontiers (IF^{FG}=0.9970), which means that USF model type does not have a significant effect on productivity level. As for the number of doctors, better productivity levels were observed among USFs with more than five doctors (IF^{AB}=0.9536, IF^{AC}=0.9738).

To complement the analysis of the indices, we have also calculated the Pearson correlation coefficients to see the relationship between target achievement on the 10 institutional PIs (y1,….,y10), the DEA scores derived from model (4) with restrictions (5) and (6) and the value φ^* calculated for the whole dataset, and the specified non-discretionary factors (with exception of model type, since it is a categorical variable). We have added two more variables– number of nurses and number of administrative staff members - to include all staff categories into analysis. The resulting coefficients are recorded in Table 3.5.

As we can see from Table 3.5, only four variables confirmed their influence on P4P target attainment by USFs. In particular, more experienced USFs, USFs that operate in municipalities with low population density, USFs that have a low percentage of patients aged 65 and more and/or a low percentage of diabetics are likely to achieve better performance results in terms of the explored DEA scores and PIs on clinical performance. Additionally, a low population density, a low percentage of diabetics, and more years in operation as USF have a positive effect on accessibility.

Table 3.5 Pearson correlation coefficients

Domain		Number of doctors	Number of nurses	Number of administrative staff members	Number of years in operation[4]	% of patients aged 65 and more	Population density	% of diabetics
	y1	0,009	-0,022	-0,047	0,013	-0,043	-0,008	-0,102
	y2	-0,091	-0,075	-0,020	**0,180****	-0,083	**-0,368****	-0,010
Accessibility	y3	0,080	0,066	0,057	**0,290****	0,041	-0,020	-0,079
	y4	-0,053	0,015	-0,024	0,031	-0,017	-0,029	**-0,215****
	y5	0,028	0,028	0,027	**0,278****	**-0,375****	-0,091	**-0,348****
	y6	0,071	0,059	0,047	**0,349****	**-0,446****	**-0,111***	**-0,285****
Clinical	y7	0,094	0,091	0,094	**0,382****	**-0,262****	**-0,144****	**-0,279****
Performance	y8	**-0,105***	-0,051	-0,004	**0,118***	**-0,166****	**-0,300****	**-0,227****
	y9	-0,021	0,059	0,030	**0,283****	**-0,302****	**-0,263****	**-0,300****
	y10	0,033	0,076	0,052	**0,201****	**-0,281****	**-0,192****	**-0,238****
DEA score		0,028	0,040	0,020	**0,310****	**-0.239****	**-0,151****	**-0.249****

* significant at $p<0.05$
**significant at $p<0.01$

As we can see from Table 3.5, only four variables confirmed their influence on P4P target attainment by USFs. In particular, more experienced USFs, USFs that operate in municipalities with low population density, USFs that have a low percentage of patients aged 65 and more and/or a low percentage of diabetics are likely to achieve better performance results in terms of the explored DEA scores and PIs on clinical performance. Additionally, a low population density, a low percentage of diabetics, and more years in operation as USF have a positive effect on accessibility.

As a result of the analysis of the Malmquist indices based on the methodology proposed by Camanho and Dyson (2006) and the Pearson correlation coefficients, we chose four variables – number of years in operation, percentage of patients aged 65 and more, population density, and percentage of diabetics - as the clustering criteria for USFs. We have not chosen model type for identifying clusters for several reasons. Firstly, model type is highly correlated with the number of years in operation, as more mature USFs are likely to be Model B. Secondly, by calculating the index by Camanho and Dyson (2006), we confirm that there is almost no gap between the frontiers of USFs Model A and USFs Model B, suggesting that USFs of both model types can be evaluated within a single group.

[4] Calculated for both USFs Model A and USFs Model B

We have also decided to leave the number of doctors out of the clustering analysis of USFs. There is a fixed number of patients that can be registered per family doctor, established by the public administration legislation, and no USF is obliged to accept patients beyond this limit. Since all PIs are formulated as ratios, where denominator represents/accounts for the number of registered/eligible patients, we cannot say that a greater number of family doctors or other staff members will lead to better performance on the explored PIs. As a proof of this assumption, the Pearson correlation coefficients generally do not confirm the influence of USFs' staff composition (number of doctors/nurses/administrative staff members) on the 10 institutional PIs and the DEA scores.

3.4.3 Cluster analysis results

As we have described earlier, we identified four clustering criteria for classifying USFs, including the number of years in operation, the percentage of patients aged 65 and more, population density, and the percentage of diabetics. After choosing the clustering criteria for USFs, we have tested five different methods of cluster analysis (average linkage within groups, centroid method, k-means clustering, two-step clustering, and Ward's method) to find the best solution for our case. In making this decision, we have considered the following parameters:

- the value of DB indices and CH indices for deciding on the most appropriate number of clusters;
- the size of the clusters by looking for the most balanced distribution of USFs among the proposed clusters, as the number of DMUs in the group materially affects the DEA results; in this respect, it is desirable to form clusters of a relatively similar size, or, at least, to avoid forming clusters with a very small number of units included in them;
- the relative contribution of each of the clustering criteria to separation of the clusters by exploring F-ratios of a one-way ANOVA test.

Furthermore, since we chose the squared Euclidean distance as a measure of distance between cases in cluster analysis, graphical representation also helps to find the best way in which clusters can be formed.

Table 3.6 presents the results for the DB indices and the CH indices calculated for 2-5 clusters for each of the tested cluster methods.

Table 3.6 Internal indices for cluster validation

	Average linkage within groups		Centroid method		K-means clustering		Two-step clustering		Ward's method	
	DB Index	CH Index	DB Index	CH Index	DB Index	CH Index	DB Index	CH Index	DB Index	CH Index
2 clusters	2.46	31.64	1.67	5.65	**4.13**	20.69	**2.37**	6.79	**2.00**	85.18
3 clusters	**2.27**	110.12	**1.25**	**664.73**	6.46	**674.38**	3.02	**653.72**	18.51	**604.89**
4 clusters	7.67	**342.81**	3.55	445.38	6.82	463.12	18.44	439.35	29.72	403.21
5 clusters	7.94	257.69	3.36	333.14	9.58	336.79	20.87	331.30	8.21	324.69

The optimal number of clusters for each method corresponds to the maximum CH index and the minimum DB index, as discussed earlier, and these values are highlighted in bold in Table 3.6. As suggested by the analysis of the indices, two or three clusters are the best options for classifying USFs.

Table 3.7 and Table 3.8 present the number of USFs included into each cluster for the case of three clusters and the case of two clusters respectively.

Table 3.7 Distribution of USFs for the case of three clusters

	Average linkage within groups	Centroid method	K-means clustering	Two-step clustering	Ward's method
Large cluster	226 USFs	301 USFs	228 USFs	294 USFs	234 USFs
Medium cluster	111 USFs	44 USFs	86 USFs	41 USFs	83 USFs
Small cluster	19 USFs	11 USFs	42 USFs	21 USFs	39 USFs

Table 3.8 Distribution of USFs for the case of two clusters

	Average linkage within groups	Centroid method	K-means clustering	Two-step clustering	Ward's method
Large cluster	245 USFs	345 USFs	240 USFs	336 USFs	234 USFs
Small cluster	111 USFs	11 USFs	116 USFs	20 USFs	122 USFs

Additionally, Figure B1 and Figure B2 in Appendix B present scatter plots for the distributions of USFs between three and two clusters in relation to three of the four clustering criteria – population density, number of years in operation, and percentage of diabetics.

Finally, Table 3.9 and Table 3.10 record F-ratios derived from a one-way ANOVA test of the clustering criteria. As we mentioned before, the magnitude of the F-ratios can give us an idea about the relative importance of each of the criteria to assining cluster membership of USFs. In particular, a high value of F-ratio with a low significance value implies that the explored variable significantly varies across clusters and it contributes to the separation of the clusters.

Table 3.9 F-ratios for the clustering criteria for the case of two clusters

	Average linkage within groups	Centroid method	K-means clustering	Two-step clustering	Ward's method
Number of years in operation	282.44**	7.37**	2.41	2.48	305.22**
Percentage of diabetics	35.06**	121.79**	478.99**	196.34**	6.55*
Percentage of patients aged 65 and more	11.28**	50.14**	270.99**	76.91**	2.83
Population density	31.04**	5.65*	20.69**	6.79**	85.18**

* significant at p<0.05
** significant at p<0.01

Table 3.10 F-ratios for the clustering criteria for the case of three clusters

	Average linkage within groups	Centroid method	K-means clustering	Two-step clustering	Ward's method
Number of years in operation	306.00**	6.75**	3.15*	4.98**	180.37**
Percentage of diabetics	17.49**	60.73**	298.02**	103.87**	3.62*
Percentage of patients aged 65 and more	10.10**	43.51**	128.01**	66.20**	13.15**
Population density	110.12**	664.74**	674.39**	653.72**	604.89**

* significant at p<0.05
** significant at p<0.01

We can clearly see from the values of F-ratios (Table 3.9 and Table 3.10) that three of the five methods – centroid method, k-means clustering, and two-step clustering – suggest assigning the least relative importance to the number of years in operation as a clustering criterion (or even do not consider it at all, as suggested by the observed significance levels) in case of three clusters and in case of two clusters. This means that USFs with different levels of experience may belong to the same clusters formed by these three metthods, which is

confirmed by the respective scatter plots in Figure B1 and Figure B2. However, accoring to our previous analysis, the number of years in operation as USF is the most impornant predictor of USFs' performance results among the four clustering criteria, as it has the highest positive correlation with the DEA scores and also positively affects target attainment on eight of the 10 institutional PIs (Table 3.5). For this reason, we believe that the relative contribution of this criterion should be high enough to classify USFs in accordance to their experience. Although in the case of three clusters this requirement is fulfilled in the solution provided by the average linkage within groups (Table 3.9 and Figure B1), this method suggests the creation of a very small group consisting of 19 USFs (Table 3.7), which contradicts our initial requirement regarding the size of the clusters.

In the case of two clusters, both the average linkage within groups and Ward's method provide similar solutions: they consider the number of years in operation as a principal clustering criterion (Table 3.9) and split USFs into two groups, respecting approximately the same proportion (about 2:1). At the same time, in our view, Ward's method provides a better clustering of USFs in comparison to the average linkage within groups for several reasons. Firstly, Ward's method classifies USFs into two clusters with a smaller difference in size than between the clusters suggested by the average linkage within groups (Table 3.8). Secondly, Ward's method has the lowest DB index for the case of two clusters, while none of the cluster validation indices confirmed the case of two clusters to be the best solution for the average linkage within groups (Table 3.6). Finally, Ward's method explicitly distinguishes population density as a clustering criterion with a significant relative importance in the cluster separation, while the average linkage within groups more equally weighs the remaining three criteria (Table 3.9). In this way, the solution by Ward's method can be more easily interpreted in terms of environmental effects, as one cluster gathers 122 USFs with less experience and/or with a higher population density (adverse environmental conditions), while the other consists of 234 more experienced USFs that operate in municipalities with a relatively lower population density (positive environmental conditions) (Figure B2). As a result, we have selected the solution provided by Ward's method for the case of two clusters among all the tested solutions.

3.4.4 The DEA assessment results and alternative P4P targets for USFs

By identifying two clusters of USFs that perform in different environmental conditions, we ensure the homogeneity of USFs that belong to the same cluster. Following the procedure

by Löber and Staat (2010), we run distinct DEA models for these two clusters: for USFs of Cluster 1 (122 USFs) we calculate the performance scores by including observations from this cluster exclusively, and for obtaining results for USFs of Cluster 2 (234 USFs), we consider data from both clusters. In this way, we will compare each USF only with those USFs that operate in the same or worse environmental conditions.

In order to proceed to the performance evaluation of the USFs, we have to specify the value of φ^* that represents the limit for the minimum difference in virtual weights in the DEA-based assessment of USFs.

For this purpose, we have implemented the two-step procedure proposed by Ramón et al. (2010). Firstly, we have run the benefit-of-the-doubt model (2) with the weight restrictions (5) and (6) to identify extreme efficient units for Cluster 1 and for Cluster 2. In this model, we used a constant input that is equal to one and 10 outputs corresponding to the 10 institutional PIs described in Table 3.3.

As a result, we identified three extreme efficient USFs for Cluster 1 and six extreme efficient USFs for Cluster 2. The information about the optimal virtual weights of these USFs is recorded in Table 3.11 and Table 3.12.

Table 3.11 Optimal virtual weights for the extreme efficient USFs for Cluster 1

	$\mu_1 y_1$	$\mu_2 y_2$	$\mu_3 y_3$	$\mu_4 y_4$	$\mu_5 y_5$	$\mu_6 y_6$	$\mu_7 y_7$	$\mu_8 y_8$	$\mu_9 y_9$	$\mu_{10} y_{10}$
USF Ramalde	12.31%	12.99%	12.31%	12.39%	6.17%	12.99%	6.17%	6.19%	12.31%	6.17%
USF Porta do Sol	6.97%	19.36%	11.68%	11.99%	6.97%	13.57%	6.97%	6.97%	6.97%	8.53%
USF Barcel Saúde	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%

Table 3.12 Optimal virtual weights for the extreme efficient USFs for Cluster 2

	$\mu_1 y_1$	$\mu_2 y_2$	$\mu_3 y_3$	$\mu_4 y_4$	$\mu_5 y_5$	$\mu_6 y_6$	$\mu_7 y_7$	$\mu_8 y_8$	$\mu_9 y_9$	$\mu_{10} y_{10}$
USF Vale do Vez	2.22%	26.37%	12.59%	8.82%	0.35%	26.37%	8.82%	8.58%	5.26%	2.22%
USF Portus Alacer	5.88%	26.82%	5.88%	11.42%	5.88%	20.60%	5.88%	5.88%	5.88%	5.88%
USF Nós e Vós Saúde	2.43%	37.83%	7.30%	2.43%	2.43%	37.83%	2.43%	2.43%	2.43%	2.43%
USF Nova Via	2.73%	31.92%	8.18%	7.17%	2.73%	31.92%	2.73%	7.17%	2.73%	2.73%
USF Porta do Sol	0.00%	35.25%	7.17%	7.59%	0.00%	31.78%	5.57%	0.00%	6.54%	0.00%
USF Mais Saúde	6.51%	28.02%	8.96%	6.51%	6.34%	18.31%	6.34%	6.34%	6.34%	6.51%

The extreme efficient units specified by the benefit-of-the-doubt model (2) with the weight restrictions (5) and (6) will remain as such even after adding the φ^* related restriction to the model specification, as argued by Ramón et al. (2010). Therefore, we can conclude that eight of the 356 USFs are considered best performers under the proposed DEA-based methodology for performance assessment of USFs.

If we look at performance profiles of best performing USFs, we can see that most of them have top-quartile or best results on the contracted PIs. For example, USF Portus Alacer has the best result on the global utilisation rate among all USFs (83.35%), which is the most valued PI, according to the ACSS methodology, and it also has the best result on the rate of home visits by the nurse (per 100 patients) (28.76%), which is the third most valued PI. Likewise, USF Nós e Vós Saúde has the highest results on the percentage of women aged 25-49 who have had a colpocytology in the past 3 years (85.81%) among all USFs, which is the second most important PI. USF Vale do Vez, in turn, has top quartile results on the following five PIs among all USFs: the global utilisation rate of medical consultations (81.90%), the rate of home visits by the doctor (per 100 patients) (5.76%), the rate of home visits by the nurse (per 100 patients) (19.17%), the percentage of diabetics with at least 2 HbA1C tests registered during the past 12 months and covering 2 semesters (93.59%), and the percentage of first consultations in the first trimester of pregnancy (95.12%). Additionally, USF Vale do Vez has 100% of first consultations in the first trimester of pregnancy and all 2-year-old children fully immunized in accordance with the National Vaccine Programme.

In Cluster 1, USF Ramalde has top quartile results on all PIs, except for the proportion of consultations provided by a designated family doctor. Similarly, USF Barcel Saúde has the best results in Cluster 1 with regards to the percentage of women aged 25-49 who have had a colpocytology in the past 3 years (80.16%), the percentage of patients with hypertension with at least one record of blood pressure in each semester (97.73%), and the percentage of diabetics with at least 2 HbA1C tests registered during the past 12 months and covering 2 semesters (97.06%). However, both USF Ramalde and USF Barcel Saúde are outperformed by best performing USFs from Cluster 2. By contrast, USF Porta do Sol acts as a best performer in both clusters, which means that even if compared to USFs with better environmental conditions, USF Porta do Sol still achieves top-ranked performance results.

Following the procedure by Ramón et al. (2010), we then maximised the ratios between the minimum and the maximum virtual weights of the extreme efficient USFs by solving

model (3) with weight restrictions (5) and (6). By maximising φ, we ensure that these units cannot unbalance the distribution of the virtual weight any further in order to be treated as fully efficient in the evaluation of each cluster, and, therefore, we avoid the problem of alternative optimal solutions. The results on the optimal values of φ along with the corresponding optimal virtuals are recorded in Table 3.13 and Table 3.14.

Table 3.13 The maximised φ of the extreme efficient USFs for Cluster 1

	φ	$\mu_1 y_1$	$\mu_2 y_2$	$\mu_3 y_3$	$\mu_4 y_4$	$\mu_5 y_5$	$\mu_6 y_6$	$\mu_7 y_7$	$\mu_8 y_8$	$\mu_9 y_9$	$\mu_{10} y_{10}$
USF Ramalde	54.52%	11.58%	13.42%	11.58%	13.42%	7.32%	13.42%	7.32%	7.32%	7.32%	7.32%
USF Porta do Sol	53.44%	11.38%	13.62%	13.62%	11.38%	7.28%	13.62%	7.28%	7.28%	7.28%	7.28%
USF Barcel Saúde	6.57%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%

Table 3.14 The maximised φ of the extreme efficient USFs for Cluster 2

	φ	$\mu_1 y_1$	$\mu_2 y_2$	$\mu_3 y_3$	$\mu_4 y_4$	$\mu_5 y_5$	$\mu_6 y_6$	$\mu_7 y_7$	$\mu_8 y_8$	$\mu_9 y_9$	$\mu_{10} y_{10}$
USF Vale do Vez	40.06%	6.39%	15.94%	15.94%	11.73%	6.39%	15.94%	6.39%	6.39%	8.51%	6.39%
USF Portus Alacer	40.84%	8.56%	16.44%	8.56%	16.44%	6.71%	16.44%	6.71%	6.71%	6.71%	6.71%
USF Nós e Vós Saúde	11.51%	4.28%	37.17%	4.28%	4.28%	4.28%	28.62%	4.28%	4.28%	4.28%	4.28%
USF Nova Via	10.31%	3.84%	30.86%	8.67%	6.63%	3.18%	30.86%	3.18%	6.41%	3.18%	3.18%
USF Porta do Sol	13.65%	5.66%	29.72%	7.16%	7.46%	4.06%	29.72%	4.06%	4.06%	4.06%	4.06%
USF Mais Saúde	52.43%	11.19%	13.81%	13.81%	11.19%	7.24%	13.81%	7.24%	7.24%	7.24%	7.24%

As we can see from Table 3.13 and Table 3.14, the minimum of the optimal φ is 6.57% for Cluster 1 and it equals 10.31% for Cluster 2. We used these values in model (4) (with restrictions (5) and (6)) as the bound φ* to establish the minimum level of dissimilarity between the optimal virtual weights of the underperforming USFs. The results on performance evaluation of all USFs (DEA scores and optimal virtual weights) are presented in Table C1 in Appendix C.

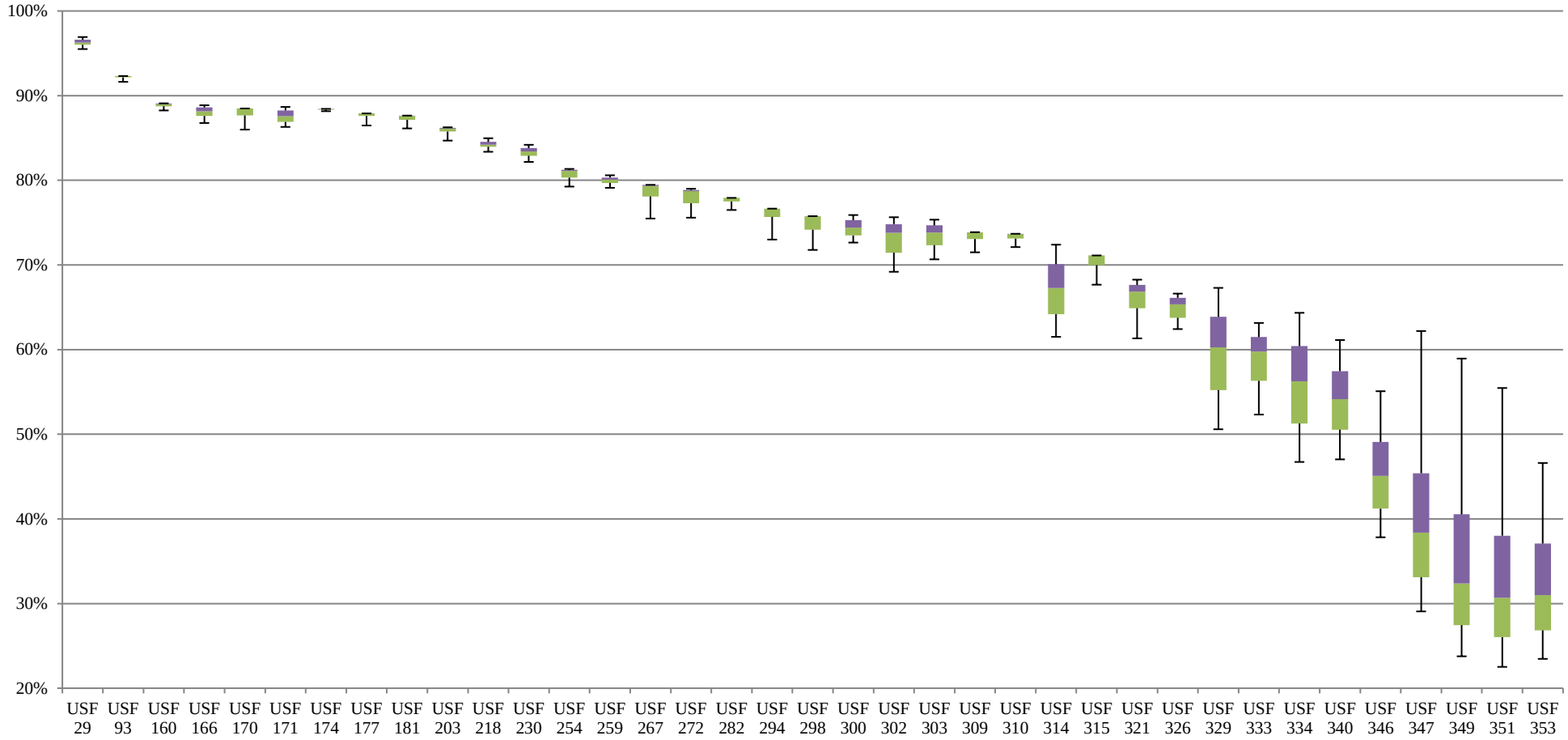
In order to provide further insight regarding the influence of the value φ* on the resulting scores, we undertook a sensitivity analysis of the DEA results to the specification of this value. In particular, we have repeated the evaluation of 122 USFs from Cluster 1 under eleven different scenarios, going from φ*=0 to φ*=0.5 by 0.05. Figure 3.1 presents the box-plots for the undertaken sensitivity analysis for a sample of 37 USFs of ARS Lisbon and Tagus River (all USFs belong to Cluster 1). The maximum resulting scores for each USF correspond to the scenario with φ*=0. In this case, the model allows for a total flexibility in the level of dissimilarity between the optimal virtuals, being the least restrictive scenario

among all. By contrast, if φ*=0.5, the model restricts the affordable level of dissimilarity between the optimal virtuals by 50%, which means that the minimum optimal weight cannot be lower than 50% of the maximum optimal weight. In this scenario, the resulting scores are either equal or worse when compared to the results from the other runs with lower values of φ*.

As we can see from the box-plots in Figure 3.1, the level of variability of the DEA scores among 23 USFs that scored more than 70% as the minimum score (e.g. USF 303, USF 294, USF 93) is relatively low (less than 5%). By contrast, the level of variability of the DEA scores among 4 USFs that scored less than 30% as the minimum score (USF 353, USF 351, USF 349, USF 347) is very high (more than 20%). There is also an intermediate group of 10 USF that have the minimum score of 30-70% (e.g. USF 340, USF 333, USF 321) with a varying degree of robustness of the DEA scores under the explored scenarios. In this way, we can verify that the value φ* (in combination with weight restrictions (5) and (6)) affects the resulting scores of severely underperforming USFs to a much greater extent than it affects the assessment results of top performing USFs. At the same time, we suggest that only those USFs that scored 80% and more under the proposed DEA methodology are rewarded with a bonus associated with relative performance assessment, and in this group of USFs the methodology is very robust.

A comparative analysis of the scores under the ACSS methodology and the DEA methodology can also provide interesting insights. For example, USF 170 and USF 93 from ARS Lisbon and Tagus River achieved a 100% score under the ACSS methodology, while they are not considered to be the best performers under the suggested DEA methodology: USF 170 scored 88.48%, while USF 93 scored 92.32%.

Figure 3.1 Sensitivity analysis of DEA scores to the specification of the value φ* for a sample of 37 USFs of ARS Lisbon and Tagus River



This means that both USFs had relatively undemanding targets on some of the explored PIs for the contracting year of 2013, when compared to the DEA targets established by our approach. Table 3.15 and Table 3.16 present the actual results and the contracted targets for USF 170 and USF 93 respectively along with the DEA targets, suggested by our model.

Besides the actual DEA targets that are labeled as "100%", Table 3.15 and Table 3.16 also present a 90% value of the DEA targets and an 80% value of the DEA targets to show the respective thresholds, associated with the measurement of the absolute target attainment for the following year, according to the suggested methodology for P4P contracting with USFs.

Table 3.15 Performance results and targets of USF 170

	Actual results	Contracted targets	DEA targets		
			100%	90%	80%
• Proportion of consultations provided by a designated family doctor (y1)	89.06%	88.30%	86.51%	77.86%	69.21%
• Global utilisation rate of medical consultations (y2)	73.16%	73.00%	76.66%	68.99%	61.33%
• Rate of home visits by the doctor (per 100 patients) (y3)	3.54%	2.90%	3.58%	3.22%	2.86%
• Rate of home visits by the nurse (per 100 patients) (y4)	11.22%	10.80%	21.89%	19.70%	17.51%
• Percentage of patients with hypertension with at least one record of blood pressure in each semester	87.51%	83.00%	92.46%	83.21%	73.97%
• Percentage of women aged 25-64 who have had a colpocytology in the past 3 years (y6)	56.45%	55.00%	76.13%	68.52%	60.90%
• Percentage of diabetics with at least 2 HbA1C tests registered during the past 12 months and covering 2	85.23%	75.00%	90.56%	81.50%	72.45%
• Percentage of first consultations in the first 28 days of the newborn's life (y8)	96.97%	90.00%	93.81%	84.43%	75.05%
• Percentage of fully immunised 2-year-old children in accordance with the National Vaccine Programme	93.81%	96.00%	98.11%	88.30%	78.49%
• Percentage of first consultations in the first trimester of pregnancy (y10)	95.40%	86.00%	91.78%	82.53%	73.36%

As discussed earlier, the DEA targets were drawn from the performance profiles of the relevant benchmarks that belong to the production frontier. In particular, USF Ramalde from ARS Norte acts as a single benchmark to USF 170, while USF Porta do Sol from ARS Norte serves as a single benchmark to USF 93. From Table 3.15 and Table 3.16, we can clearly see the difference in the actual targets and the targets identified by the benchmarks for both USFs.

Table 3.16 Performance results and targets of USF 93

	Actual results	Contracted targets	DEA targets		
			100%	90%	80%
• Proportion of consultations provided by a designated family doctor (y1)	79.73%	80.00%	84.34%	75.91%	67.47%
• Global utilisation rate of medical consultations (y2)	73.91%	73.00%	77.46%	69.71%	61.97%
• Rate of home visits by the doctor (per 100 patients) (y3)	4.45%	3.00%	3.76%	3.38%	3.01%
• Rate of home visits by the nurse (per 100 patients) (y4)	15.23%	14.00%	18.53%	16.68%	14.82%
• Percentage of patients with hypertension with at least one record of blood pressure in each semester (y5)	90.81%	92.00%	95.74%	86.17%	76.59%
• Percentage of women aged 25-64 who have had a colpocytology in the past 3 years (y6)	59.06%	60.00%	78.10%	70.29%	62.48%
• Percentage of diabetics with at least 2 HbA1C tests registered during the past 12 months and covering 2	86.88%	92.00%	94.92%	85.43%	75.94%
• Percentage of first consultations in the first 28 days of the newborn's life (y8)	85.45%	85.00%	96.55%	86.90%	77.24%
• Percentage of fully immunised 2-year-old children in accordance with the National Vaccine Programme	89.37%	97.00%	98.74%	88.87%	78.99%
• Percentage of first consultations in the first trimester of pregnancy (y10)	93.96%	85.00%	96.90%	87.21%	77.52%

For example, the actual target for the most valued PI - global utilisation rate of medical consultations – was 73% for both USFs, while our approach suggests setting different targets on this PI: 76.66% for USF 170 and 77.46% for USF 93. It is interesting to note that both USFs achieved results on this PI that are very close to the contracted targets – 73.16% (USF 170) and 73.91% (USF 93). We can also track a similar situation with regards to the proportion of consultations provided by a designated family doctor (USF 170 – 89.06% achieved vs. 88.30% targeted, USF 93 – 79.73% achieved vs. 80.00% targeted), the rate of home visits by the nurse (per 100 patients) (USF 170 – 11.22% achieved vs. 10.80% targeted, USF 93 – 15.23% achieved vs. 14.00% targeted), and percentage of women aged 25-64 who have had a colpocytology in the past 3 years (USF 170 – 56.45% achieved vs. 55.00% targeted, USF 93 – 59.06% achieved vs. 60.00% targeted). In this situation, we might expect that the effect of goal gradient comes into effect. This means that rather than exhibiting the maximum possible performance level on these PIs, as suggested by more demanding targets of the identified benchmarks (for example, the targets of 76.13% and 78.10% are suggested for the percentage of women aged 25-64 who have had a colpocytology in the past 3 years for USF 170 and USF 93 respectively), these USFs aimed to achieve the results that are as close to the contracted targets as possible.

From Table 3.15 we can also see that, for example, the DEA target on the proportion of consultations provided by a designated family doctor (86.51%) is lower than the actual result (89.06%) and the contracted target (88.30%) of USF 170. As pointed out by Allen et al. (1997), a DEA model with weight restrictions may produce the targets for underperforming DMUs that involve deterioration to the current level of some of the outputs in order to enable performance improvement with regard to the other outputs. This means that in order to score better under the DEA methodology, USF 170 should reconsider its current mix of outputs and relocate a part of the resources used for the provision of consultations by a designated family doctor to the provision of other health care services that have greater importance in the context of this assessment. A similar situation can be observed in relation to the other DEA targets suggested to USF 170, including the DEA targets for the percentage of first consultations in the first 28 days of the newborn's life and percentage of first consultations in the first trimester of pregnancy.

There are also some USFs that scored higher under the DEA approach compared to the actual scores on the P4P programme. For example, while USF 309 scored 55% on the ACSS assessment, it has a higher score of 73.87% according to our approach. Unlike the previous examples, this USF has two benchmarks – USF Ramalde and USF Porta do Sol. This means that the targets for USF 309 can be obtained by a convex combination of the achieved performance results of these two best performers, considering the relevant intensity coefficients λ. The information about the actual results, the contracted targets, and the DEA targets along with the two absolute thresholds (labeled as "90%" and "80%") in the DEA target attainment is recorded in Table 3.17.

Similarly to the previous cases, all targets suggested by our approach to USF 309 are higher than the contracted targets. However, by avoiding the all-or-nothing property of absolute thresholds, we allow this USF to score better than it has actually scored according to the ACSS methodology.

In other words, while the ACSS methodology disregards any result that is lower than 80% of the targeted values for the calculation of a total achievement score (in case of USF 309, these are the rate of home visits by the doctor (per 100 patients) (79.56% achievement rate of the contracted target) and percentage of diabetics with at least 2

Table 3.17 Performance results and targets of USF 309

	Actual results	Contracted targets	USF Ramalde (λ=0.01)	USF Porta do Sol (λ=0.99)	DEA targets		
					100%	90%	80%
• Proportion of consultations provided by a designated family doctor (y1)	92.20%	85.00%	86.51%	84.34%	85.46%	76.91%	68.37%
• Global utilisation rate of medical consultations (y2)	68.47%	70.00%	76.66%	77.46%	77.05%	69.35%	61.64%
• Rate of home visits by the doctor (per 100 patients) (y3)	1.91%	2.40%	3.58%	3.76%	3.67%	3.30%	2.94%
• Rate of home visits by the nurse (per 100 patients) (y4)	11.40%	13.00%	21.89%	18.53%	20.26%	18.23%	16.21%
• Percentage of patients with hypertension with at least one record of blood pressure in each semester (y5)	69.02%	80.00%	92.46%	95.74%	94.05%	84.65%	75.24%
• Percentage of women aged 25-64 who have had a colpocytology in the past 3 years (y6)	48.90%	55.00%	76.13%	78.10%	77.08%	69.37%	61.66%
• Percentage of diabetics with at least 2 HbA1C tests registered during the past 12 months and covering 2 semesters (y7)	53.71%	80.00%	90.56%	94.92%	92.67%	83.40%	74.14%
• Percentage of first consultations in the first 28 days of the newborn's life (y8)	75.61%	80.00%	93.81%	96.55%	95.14%	85.63%	76.11%
• Percentage of fully immunised 2-year-old children in accordance with the National Vaccine Programme (y9)	91.89%	95.00%	98.11%	98.74%	98.42%	88.58%	78.74%
• Percentage of first consultations in the first trimester of pregnancy (y10)	78.17%	85.00%	91.78%	96.90%	94.26%	84.83%	75.41%

HbA1C tests registered during the past 12 months and covering 2 semesters (67.14% achievement rate of the contracted target)), the DEA assessment takes into account all results on PIs proportionally to the established best practice frontier for calculating the resulting score.

This is particularly illustrative on the example of the rate of home visits by the doctor (per 100 patients), as by failing to provide at least two more doctor home visits (or to improve the result on this PI by 0.02%), target attainment on this PI is fully disregarded for the estimation of the resulting score of USF 309, as suggested by the ACSS methodology.

Another example is USF 353 that scored relatively low on both methodologies (ACSS score – 20.00%, DEA score – 41.33%), showing significant potential for performance improvement. The information about the actual results, the contracted

targets, and the DEA targets along with the two absolute thresholds (labeled as “90%” and “80%”) in the DEA target attainment is recorded in Table 3.18.

Table 3.18 Performance results and targets of USF 353

	Actual results	Contracted targets	DEA targets		
			100%	90%	80%
Proportion of consultations provided by a designated family doctor (y1)	87.88%	85.00%	84.34%	75.91%	67.47%
Global utilisation rate of medical consultations (y2)	47.39%	65.00%	77.46%	69.71%	61.97%
Rate of home visits by the doctor (per 100 patients) (y3)	0.26%	1.20%	3.76%	3.38%	3.01%
Rate of home visits by the nurse (per 100 patients) (y4)	2.60%	14.50%	18.53%	16.68%	14.82%
Percentage of patients with hypertension with at least one record of blood pressure in each semester (y5)	62.43%	80.00%	95.74%	86.17%	76.59%
Percentage of women aged 25-64 who have had a colpocytology in the past 3 years (y6)	30.84%	45.00%	78.10%	70.29%	62.48%
Percentage of diabetics with at least 2 HbA1C tests registered during the past 12 months and covering 2	49.15%	70.00%	94.92%	85.43%	75.94%
Percentage of first consultations in the first 28 days of the newborn’s life (y8)	66.67%	80.00%	96.55%	86.90%	77.24%
Percentage of fully immunised 2-year-old children in accordance with the National Vaccine Programme (y9)	95.89%	98.00%	98.74%	88.87%	78.99%
Percentage of first consultations in the first trimester of pregnancy (y10)	72.73%	85.00%	96.90%	87.21%	77.52%

USF Porta do Sol from ARS Norte is a single benchmark to USF 353 and it managed to achieve top performance results, despite working under similar environmental conditions. In this way, we believe that the establishment of a training programme between these two USFs would contribute to transferring the necessary knowledge from USF Porta do Sol to USF 353 on how to improve the USF’s performance.

3.5 Conclusions

In this study we aim to contribute to enhancing P4P programme design in health care, both conceptually and in practice, using the example of the P4P contracts concluded with USFs, the Portuguese primary care providers, in relation to the method

used for performance assessment and target setting. To achieve this aim, we propose a DEA-based methodological approach under the benefit-the-doubt model that within a single optimization procedure evaluates the USF's performance against the 10 institutional PIs and also derives realistic targets in case further improvement is possible. The developed DEA model includes a unique combination of weight restrictions - a parameter that regulates the level of dissimilarity between optimal virtuals (Ramón et al., 2010) along with ordinal and proportional virtual weight restrictions (Sarrico and Dyson, 2004; Wong and Beasley, 1990) – that, in our view, provides an optimal balance between the exogenously derived preferences regarding the relative importance of each individual PI and domain and the endogenously defined affordable level of dissimilarity in virtual weight assignment for each provider. The established restrictions still allow for a great degree of flexibility in weight assignment, as opposed to any fixed weighting system, preserving in this way one of the main advantages of DEA. A particularly valuable feature of the proposed DEA model is that even considering the incorporated weight restrictions, it does not assume any externally defined numerical thresholds in weight assignment that are usually difficult to justify and in certain situations may lead to infeasibility problems, as pointed out by Sarrico and Dyson (2004). At the same time, the model guarantees the assignment of non-zero weights for the 10 institutional PIs, so that all individual PIs are considered in the calculation of composite scores of performance. Another valuable advantage of the proposed DEA methodology is that it appears to be robust to the specification of the value φ^* in the group of top performing USFs, so that there is a clearly defined group of the USFs that are eligible to receive a bonus payment based on the DEA assessment results. Furthermore, the model can be easily adapted for the use of a different set of PIs or a different number of performance domains, making possible to develop further applications of the proposed model within and outside P4P context.

Another advantage of the proposed DEA methodology is that it explicitly accounts for the influence of environmental factors on P4P target attainment by USFs. We identified four relevant environmental factors - number of years in operation as USF, population density, percentage of patients aged 65 and more, and percentage of diabetics – via analysis of the Malmquist indices by Camanho and Dyson (2006) and the Pearson correlation coefficients. Following this identification, we performed a cluster analysis to provide a single classification of USFs in relation to their

environmental conditions, by contrast to the existing ACSS methodology. As a result of this cluster analysis, we were able to form a small cluster of 122 USFs that operate in less favourable conditions and a large cluster of 234 USFS that operate in more favourble conditions. In order to guarantee a fair comparison of USFs, we performed a DEA assessment of each USF by comparing its results only with those peers that operate in the same or worse environments, as suggested by Löber and Staat (2010). As a result of the DEA assessment, we identified 8 best performing USFs and 348 underperforming USFs.

On the basis of the DEA assessment results, we propose an alternative methodology for P4P contracting with USFs. This methodology suggests that a distinct system of graduated bonus payments should be linked to absolute performance assessment of the USFs that were classified as underperforming in the previous year and have their absolute targets derived from the respective benchmarks, and a similar bonus system should also be linked to relative performance assessment of all USFs in order to reward those units that scored relatively high in the undertaken DEA assessment. This combination aims to reinforce the effect of P4P on providers' performance, as while absolute targets set clear goals to work towards and offer greater control to the providers with regards to the attainability of a reward, relative assessment encourages continual improvement and stimulates the achievement of best possible performance results. It is also suggested that a benchmarking programme is developed in order to contribute to transferring knowledge from the best practices to the units that did not score 100%. Our suggestion is that this benchmarking programme can be linked with several supplementary rewards for the best performing USFs.

Whilst this study improves our understanding regarding the implementation of P4P in primary care, we recognise that it has some limitations. Firstly, for the estimation of relative performance scores of USFs, we have only considered those PIs that are currently included into the P4P contracts with USFs, subject to data availability. The majority of the studied indicators represent the volume of health care services provided by a USF to its registered patients, and these, in many cases, can only provide a partial view on performance of primary care providers (Amado and Dyson, 2008). Therefore, to enable a more comprehensive and in-depth analysis of USFs' performance, it would be highly desirable that future studies also include outcome measures of USFs' performance as well measures related to the cost of health care delivery by USFs.

Secondly, in future studies it is also important to take into account a wider range of environmental factors that can potentially affect USFs' performance. These are, for example, the socio-economic characteristics of the population, considering that, according to previous research, these characteristics are confirmed to have a significant effect on outcomes achievement in primary care delivery (e.g. Amado and Dyson, 2009). Since our model does not include any outcome measures, it is expected that the socio-economic characteristics of the population would not significantly affect the possibility of achieving better/worse performance results against the explored PIs, although this assumption should be carefully verified in future research. Finally, in this study we have adopted the benefit-of-the-doubt approach, in which all inputs are substituted with a single input equal to one and outputs represent the 10 institutional PIs, in order to perform a P4P assessment and produce the targets for USFs, being in line with the ACSS methodology. However, the use of an assessment model that is based on volume measures may provide a fairer comparison of USFs, as it accounts for potential differences in resources. In our view, the development of such model should also be addressed in future studies.

Chapter 4

CONCLUSIONS

4.1 General discussion of the results

The purpose of this thesis was to explore the methodological issues related to performance assessment and incentive regulation in the context of primary health care and to propose specific ways for improvement of performance measurement in this research area.

In the first article included in this thesis, we developed a conceptual framework that incorporates three stages of performance assessment (measurement, analysis, and action) and captures four performance dimensions (accessibility, equity, efficiency, and effectiveness), following previous theoretical works in this research field. In the developed conceptual framework, we have also linked each dimension with the relevant sub-dimensions to provide a better understanding of the performance aspects measured by each dimension. Furthermore, after undertaking a systematic review of the empirical literature on primary care assessments, we were able to propose a functional classification of PIs for each of the performance dimensions. In this respect, we believe that the developed framework and classification fully describe the main conceptual elements that are needed for a comprehensive assessment of primary care providers and also help to reveal the relationships between these elements. It is also our conviction that the use of these theoretical developments in future studies on primary care assessments can facilitate the design of balanced measurement systems and can contribute to choosing appropriate performance dimensions, sub-dimensions, and PIs according to the specified objectives of such assessments.

In the development of the framework, we have also emphasised the need to involve the targeted stakeholders at the measurement, analysis, and action stages of performance assessment. However, the undertaken literature review revealed that the involvement of stakeholders at different stages of organisational performance

assessment has been rare, limiting in this way the impact of the reviewed research for the targeted performance improvement in health care policy and practice. Furthermore, we have also concluded that the involvement of managers and health care professionals into primary care assessments is more frequent than the involvement of other relevant stakeholders. In this situation, we might expect that the differences in stakeholders' perspectives regarding important conceptual and practical aspects in the performance assessment of primary care providers are not fully accounted for, which can adversely affect face validity of the studies' research findings.

The second article aimed to explore the potential of DEA, a non-parametric frontier technique proposed by Charnes et al. (1978), for enhancing P4P programme design in primary health care. In particular, we aimed to propose a DEA methodology for performance assessment of USFs, primary care providers in Portugal, and a new methodology for P4P contracting based on the DEA assessment results.

In the development of a DEA methodology for performance assessment of USFs, we modified the standard DEA model by adopting the benefit-of-the-doubt approach and introducing a unique combination of weight restrictions - a parameter that regulates the level of dissimilarity between optimal virtual weights, proposed by Ramón et al. (2010), combined with ordinal and proportional virtual weight restrictions (Sarrico and Dyson, 2004; Wong and Beasley, 1990). In our view, such combination of the weight restrictions provides an optimal balance between the exogenously derived preferences regarding the relative importance of each PI and the endogenously defined affordable level of dissimilarity in virtual weight assignment for each provider. The established restrictions still allow for a great degree of flexibility in weight assignment, as opposed to any fixed weighting system, preserving in this way one of the main advantages of DEA.

The proposed DEA-based approach also explicitly accounts for the influence of environmental factors on P4P target attainment. In particular, having analysed the Malmquist indices by Camanho and Dyson (2006) and the Pearson correlation coefficients, we identified four relevant environmental factors – the number of years in operation, population density, the percentage of patients aged 65 and more, and the percentage of diabetics – and performed a cluster analysis to classify the providers in relation to their environmental conditions. As a result of this cluster analysis, we obtain

a small group of 122 USFs that operate in less favorable conditions and a large group of 234 USFs that operate in more favorable conditions.

We applied the proposed DEA methodology for the assessment of 356 USFs in relation to their results against the 10 institutional PIs in the contracting year of 2013. As suggested by Löber and Staat (2010), each USF was compared only with those peers that operate in the same or worse environments in accordance to the cluster membership of the USFs. As a result of the DEA assessment, we have identified 8 best performing USFs and 348 underperforming USFs.

We suggested a new methodology for P4P contracting with USFs based on the DEA assessment results. This methodology includes a combination of absolute and relative performance assessments that reinforces the effect of P4P on USFs with different performance levels. In particular, we proposed rewarding absolute target attainment of the USFs that were classified as underperforming in the previous year and have their absolute targets derived from the respective benchmarks and rewarding relative target attainment of the USFs that scored relatively high on the undertaken DEA assessment. Each of these two assessments is proposed to be linked to a distinct system of graduated rewards. We also suggested that a benchmarking programme should be created in order contribute to disseminating of best practices between the USFs with similar characteristics and similar performance priorities. The development of this benchmarking programme would be rewarded by several supplementary bonus payments for the best performing USFs.

In general, we believe that there are a number of valuable features of the proposed DEA methodology for performance evaluation and subsequent target setting in the context of P4P. In particular, our approach assumes a holistic performance assessment that recognizes the existence of multiple trade-offs between individual PIs and respects the relative strengths and weaknesses as well as different priorities in P4P target attainment of USFs. The proposed approach also considers the effect of the relevant environmental factors in the assessment of USFs, so that it effectively disentangles the variation in USFs' performance results that are attributable to environmental conditions from factors that are under direct control of health care managers and staff. In this way, we guarantee that the absolute targets established for underperforming USFs for subsequent periods are realistic and feasible, since these targets are based on adjusted

estimates of performance. Furthermore, if the proposed DEA methodology was adopted for P4P evaluations of USFs, it would encourage the providers to show their best possible results, given that DEA is based on a relative assessment of performance.

From a joint analysis of the research findings of the two articles, we can also draw several conclusions. In particular, we believe that performance assessment of USFs in the context of P4P can be significantly improved, if the proposed conceptual framework for primary care assessments is adopted. The present P4P contracts suggest assessing and rewarding USFs' performance against a small range of PIs and a limited number of sub-dimensions, leaving out important sub-dimensions included into the framework. For example, these contracts suggest the assessment of PIs on realised accessibility, however, it would be desirable to clarify beforehand whether all USFs have the same level of potential accessibility. This is an important piece of knowledge, since the differences in terms of economic accessibility, organisational accessibility, geographic accessibility, and service availability across USFs may ultimately result in a better or worse level of realised access to USFs among their registered patients and, therefore, may represent one of the areas for further improvement.

There is also no assessment of the effect of the Portuguese P4P programme on equity of access, equity of treatment, or equity of health care outcomes. However, it is recognised that the implementation of P4P programmes may exacerbate health care disparities in relation to specific patient groups by selecting those patients for whom it is easier to achieve good performance results (Conrad, 2009).

Furthermore, the P4P contracts include two measures on cost-efficiency - the average cost of medications prescribed per user and the average cost of complementary diagnostic and therapeutic means prescribed per user – with only a limited adjustment for the patient case-mix, such as accounting for differences in the percentage of patients aged 65 and more and accounting for differences in the percentage of diabetics. In our view, it would be desirable to compare the costs of health care delivery in each of the medical conditions separately, such as, for example, the average health care costs for patients with asthma, for patients with diabetes, for patients with hypertensions and so forth, considering the comorbidity status of the patients. Furthermore, comparing health care costs with the outputs and outcomes of health care delivery under a DEA

assessment can provide further insight regarding the cost and allocative efficiencies of USFs.

Regarding the assessment of effectiveness, there are some differences between the P4P contracts concluded in 2013 and 2016. In particular, the P4P contracts in 2016 suggested measuring the proportion of patients with hypertension aged less than 65 years and with arterial pressure less than 150/90 and the proportion of diabetics with the last HgbA1c $\leq$ 8.0% to assess USFs' outcomes achievement. These outcome measures were absent in the P4P contracts in 2013. Moreover, the assessment of clinical effectiveness in the latest P4P contracts included a larger range of activities associated with the management of maternal health, family planning, and child health in the first year of life. The measurement of patient satisfaction was also preserved in the P4P contracts in 2016. At the same time, we recommend to assess the effectiveness of interpersonal care within the context of P4P as one of the important predictors of outcomes achievement.

4.2 Suggestions for future studies

In this section we aim to outline the main suggestions for future research, based on the analysis of the two research articles included in this thesis.

1. Given that the undertaken systematic literature review showed that the four performance dimensions have mostly been explored separately, it is important that future studies investigate the relationships between different dimensions by exploring several or all dimensions simultaneously. Similarly, it is recommended to further explore the relationships between different sub-dimensions within a single dimension, such as, for example, the relationship between potential and realised accessibilities.

2. Considering an insufficient involvement of targeted stakeholders at different stages of performance assessment, we suggest that future studies should make more effort to involve these stakeholders into primary care assessments. Moreover, it is recommended that a greater emphasis is given to the involvement of patients and citizens among other stakeholders.

3. We also recommend exploring the possibility of conducting quality appraisal of the empirical studies devoted to health care performance assessments.

4. Another suggestion is to enlarge a set of sub-dimensions and the respective PIs included in the performance assessment of USFs by adopting the conceptual framework for primary care assessments developed in this thesis.

5. It is also recommended that future studies take into account a wider range of environmental factors that can potentially affect USFs' performance, such as the socio-economic characteristics of the population among other factors.

6. Finally, we recommend undertaking a DEA assessment based on volume measures, as it allows a fairer assessment of USFs by accounting for potential differences in resources between the providers.

References

1. Aakvik, A., & Holmås, T. H. (2006). Access to primary health care and health outcomes: the relationships between GP characteristics and mortality rates. Journal of Health Economics, 25(6), 1139-1153.
2. Abrams, R., Savela, T., Trinity, M. T., Falik, M., Tutunjian, B. & Ulmer, C. (1995) Performance of community health centers under managed care. *Journal of Ambulatory Care Management*, 18(3), 77-88.
3. ACSS (2012) *Cuidados de saúde primários: Metodologia de contratualização*, Lisbon, Ministry of Health.
4. ACSS (2015) *Relatório. Candidaturas e constituição de USF and UCC*, Lisbon, Ministry of Health.
5. ACSS (2016) *Termos de referência para contratualização nos cuidados de saúde primários em 2016*, Lisbon, Ministry of Health.
6. Agrell, P. J., Bogetoft, P. & Tind, J. (2002) Incentive plans for productive efficiency, innovation and learning. *International Journal of Production Economics*, 78(1), 1–11.
7. Akazili, J., Adjuik, M., Jehu-Appiah, C. & Zere, E. (2008) Using data envelopment analysis to measure the extent of technical efficiency of public health centres in Ghana. *BMC International Health and Human Rights*, 8, 11.
8. Allen, R., Athanassopoulos, A., Dyson, R. G. & Thanassoulis, E. (1997) Weight restrictions and value judgements in data envelopment analysis: evolution, development and future directions. *Annals of Operations Research*, 73, 13-34.
1. Allin, S., Hernández-Quevedo, C. & Masseria, C. (2009) Measuring equity of access to health care, in Smith, P.C., Mossialos, E., Papanicolas, I. & Leatherman S (eds.), *Performance measurement for health system improvement*, Cambridge, Cambridge University Press, 187-221.
2. Al-Taiar, A., Clark, A., Longenecker, J. C. & Whitty, C. J. (2010) Physical accessibility and utilization of health services in Yemen. *International Journal of Health Geographics*, 9, 38.
3. Amado, C. A. F. & Dyson, R. G. (2008) On comparing the performance of primary care providers. *European Journal of Operational Research*, 185(3), 915-932.
4. Amado, C. A. F. & Dyson, R. G. (2009) Exploring the use of DEA for formative evaluation in primary diabetes care: An application to compare English practices. *Journal of the Operational Research Society*, 60(11), 1469-1482.
5. Amado, C. A. F. & Santos, S. P. (2009) Challenges for performance assessment and improvement in primary health care: The case of the Portuguese health centres. *Health Policy*, 91(1), 43-56.
6. Amado, C. A. F., São José, J. M. S. & Santos, S. P. (2016) Measuring active ageing: A Data Envelopment Analysis approach. *European Journal of Operational Research*, 255(1), 207–223.

7. Andersen, R. M., McCutcheon, A., Aday, L. A., Chiu, G. Y. & Bell, R. (1983) Exploring dimensions of access to medical care. *Health Services Research*, 18(1), 49-74.
8. Andersen, P. & Petersen, N. C. (1993) A procedure for ranking efficient units in Data Envelopment Analysis. *Management Science*, 39(10), 1261–1264.
9. Andersen, R. & Aday, L. A. (1978) Access to medical care in the U.S.: realized and potential. *Medical Care*, 16(7), 533-546.
10. Andes, S., Metzger, L. M., Kralewski, J. & Gans, D. (2002) Measuring efficiency of physician practices using data envelopment analysis. *Managed Care*, 11(11), 48-54.
11. Arakawa, T., Arcêncio, R. A., Scatolin, B. E., Scatena, L. M., Ruffino-Netto, A. & Scatena Villa, T. C. (2011) Accessibility to tuberculosis treatment: assessment of health service performance. *Revista Latino-Americana de Enfermagem*, 19(4), 994-1002.
12. Aysola, J., Bitton, A., Zaslavsky, A. M. & Ayanian, J. Z. (2013) Quality and equity of primary care with patient-centered medical homes: results from a national survey. *Medical Care*, 51(1), 68-77.
13. Baker, R. (1992) General practice in Gloucestershire, Avon and Somerset: explaining variations in standards. *British Journal of General Practice*, 42, 415-418.
14. Banker, R. D. & Morey, R. C. (1986) The use of categorical variables in Data Envelopment Analysis. *Management Science*, 32(12), 1613–1627.
15. Barbaro, M. C., Lettiere, A, & Nakano, A. M. S. (2014) Prenatal care for adolescents and attributes of primary health care. *Revista Latino-Americana de Enfermagem*, 22(1), 108-114.
16. Bastian, N. D., Kang, H., Griffin, P. M. & Fulton, L. V. (2016) Measuring the effect of pay-for-performance financial incentives on hospital efficiency in the military health system. *IIE Transactions on Healthcare Systems Engineering*, 6(1), 33–41.
17. Basu, J. & Clancy, C. (2001) Racial disparity, primary care, and specialty referral. *Health Services Research*, 36(6), 64-77.
18. Bates, J., Baines, D. & Whynes, D., (1996) Measuring the efficiency of prescribing by general practitioners. *Journal of the Operational Research Society*, 47(12), 14-43.
19. Beaulieu, M. D., Haggerty, J. L., Beaulieu, C., Bouharaoui, F., Lévesque, J. F., Pineault, R., Burge, F. & Santor, D. A. (2011) Interpersonal communication from the patient perspective: comparison of primary healthcare evaluation instruments. *Health Policy*, 7, 108-123.
20. Bell, S., Wilson, K., Bissonnette, L. & Shah, T. (2013) Access to primary health care: does neighborhood of residence matter? *Annals of the Association of American Geographers*, 103, 85-105.
21. Bissonnette, L., Wilson, K., Bell, S. & Shah, T. I. (2012) Neighbourhoods and potential access to health care: the role of spatial and aspatial factors. *Health & Place*, 18, 841-853.

22. Bogetoft, P. (1994) Incentive efficient production frontiers: an agency perspective on DEA. *Management Science*, 40(8), 959-968.
23. Breyer, P. R. (1977) Neighborhood health centers: an assessment. *American Journal of Public Health*, 67(2), 179-182.
24. Bryce, C., Engberg, J. & Douglas, W. (2000) Comparing the agreement among alternative models in evaluating HMO efficiency. *Health Services Research*, 35(2), 509-528.
25. Buck, D. (2000) The efficiency of the community dental services in England: a data envelopment analysis. *Community Dentistry and Oral Epidemiology*, 28, 274-280.
26. Buja, A., Damiani, G., Gini, R.,Visca, M., Federico, B., Donato, D., Francesconi, P., Marini, A., Donatini, A., Brugaletta, S., Baldo, V. & Donata Bellentani, M. (2014) Systematic age-related differences in chronic disease management in a population-based cohort study: a new paradigm of primary care is required. *PLoS ONE*, 9(3), 1-9.
27. Calinski, T. & Harabasz, J. (1974) A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, 3(1), 1–27.
28. Camanho, A. S. & Dyson, R. G. (2006) Data envelopment analysis and Malmquist indices for measuring group performance. *Journal of Productivity Analysis*, 26(1), 35–49.
29. Campbell, R. J., Ramirez, A. M., Perez, K. & Roetzheim, R.G.(2003) Cervical cancer rates and the supply of primary care physicians in Florida. *Family Medicine*, 35, 60-64.
30. Campbell, S. M., Roland, M. O., & Buetow, S. A. (2000). Defining quality of care. Social Science & Medicine, 51(11), 1611-1625.
31. Campbell, S., Reeves, D., Kontopantelis, E., Middleton, E., Sibbald, B. & Roland, M. (2007) Quality of primary care in England with the introduction of pay for performance. *The New England Journal of Medicine*, 357(2), 181-190.
32. Carrier, E. R., Schneider, E., Pham, H. H. & Bach, P. B. (2011) Association between quality of care and the sociodemographic composition of physicians' patient panels: a repeat cross-sectional analysis. *Journal of General Internal Medicine*, 26(9), 987-994.
33. Cashin, C., Chi, Y-L., Smith, P.C., Borowitz, M. & Thomson, S. (eds) (2014) *Paying for performance in health care: Implications for health system performance and accountability,* Maidenhead, Open University Press.
34. Charnes, A., Cooper, W., Huang, Z. & Sun, D. (1990) Polyhedral Cone-Ratio DEA Models with an illustrative application to large commercial banks. *Journal of Econometrics*, 46(1-2), 73–91.
35. Charnes, A., Cooper, W. W. & Rhodes, E. (1978) Measuring efficiency of decision-making units. *European Journal of Operational Research*, 3, 429-444.
36. Cherchye, L., Moesen, W. Rogge, N., Van Puyenbroeck, T. Saisana, M., Saltelli, A. Liska, R. & Tarantola, S.(2008) Creating composite indicators with DEA and robustness analysis: the case of the Technology Achievement Index. *Journal of the Operational Research Society,* 59(2), 239-251.

37. Cherchye, L., Moesen, W., Rogge, N. & Puyenbroeck, T. V. (2007) An Introduction to 'Benefit of the Doubt' Composite Indicators. *Social Indicators Research*, 82(1), 111–145.
38. Chilingerian, J. A. & Sherman, H. D. (1996) Benchmarking physician practice patterns with DEA: a multi-stage approach for cost containment. *Annals of Operations Research,* 67(1), 83-116.
39. Chilingerian, J. A. & Sherman, H. D. (1997) DEA and primary care physicians report cards: deriving preferred practice cones from managed care service concepts and operating strategies. *Annals of Operations Research*, 73, 35-66.
40. Collier, D. A., Collier, C. E. & Kelly, T. M. (2006) Benchmarking physician performance, Part 2. *The Journal of Medical Practice Management*, 21(5), 273-279.
41. Conrad, D. A. (2009) Incentives for health-care performance improvement, in Smith, P. C., Mossialos, E., Papanicolas, I. & Leatherman S. (eds.), *Performance measurement for health system improvement*, Cambridge, Cambridge University Press, 582-612.
42. Cooper, W. W., Ruiz, J. L. & Sirvent, I. (2007) Choosing weights from alternative optimal solutions of dual multiplier models in DEA. *European Journal of Operational Research,* 180(1), 443–458.
43. Cooper, W. W., Seiford, L. W. & Zhu, J. (eds) (2011) *Handbook on Data Envelopment Analysis* (2nd ed.), New York, Springer.
44. Cordero-Ferrera, J. M., Crespo-Cebada, E. & Murillo-Zamorano, L. R. (2011) Measuring technical efficiency in primary health care: the effect of exogenous variables on results. *Journal of Medical Systems*, 35(4), 545-554.
45. Cordero-Ferrera, J. M., Crespo-Cebada, E. & Murillo-Zamorano, L. R. (2014) The effect of quality and socio-demographic variables on efficiency measures in primary health care. *The European Journal of Health Economics*, 15(3), 289-302.
46. Coughlin, S. S., Leadbetter, S., Richards, T. & Sabatino, S. A. (2008) Contextual analysis of breast and cervical cancer screening and factors associated with health care access among United States women, 2002. *Social Science & Medicine*, 66(2), 260-275.
47. Cromwell, J., Trisolini, M. G., Pope, G. C., Mitchell, J. B. & Greenwald, L. M. (eds) (2011) *Pay for performance in health care: methods and approaches*, Research Triangle Park, North Carolina, RTI Press.
48. Dahrouge, S., Hogg, W., Tuna, M., Russell, G., Devlin, R. A., Tugwell, P. & Kristjansson, E. (2010) An evaluation of gender equity in different models of primary care practices in Ontario. *BMC Public Health*, 10, 151-162.
49. Dahrouge, S., Hogg, W., Ward, N., Tuna, M., Devlin, R. A., Kristjansson, E., Tugwell, P. & Pottie, K. (2013) Delivery of primary health care to persons who are socio-economically disadvantaged: does the organizational delivery model matter? *BMC Health Services Research*, 13, 1-22.
50. Davies, D. L. & Bouldin, D. W. (1979) A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2), 224–227.

51. Defelice, L. C. & Bradford, W. D. (1997) Relative Inefficiencies in Production between Solo and Group Practice Physicians. *Health Economics*, 6(5), 455-465.
52. Deidda, M., Lupiáñez-Villanueva, F., Codagnone, C. & Maghiros, I. (2014) Using data envelopment analysis to analyse the efficiency of primary care units. *Journal of Medical Systems*, 38(10), 122.
53. Despotis, D. (2005) Measuring human development via data envelopment analysis: the case of Asia and the Pacific. *Omega*, 33(5), 385–390.
54. Donabedian, A. (1980) *The definition of quality and approaches to its assessment*. MI, Health Administration Press, Ann Arbor.
55. Donabedian, A. (1988) The quality of care. How can it be assessed? *Journal of the American Medical Associatio*n, 260(12), 1743-1748.
56. Doran, T., Kontopantelis, E., Reeves, D., Sutton, M. & Ryan, A. M. (2014) Setting performance targets in pay for performance programmes: what can we learn from QOF? *BMJ*, 348.
57. Draper, D. A., Solti, I. & Ozcan, Y. A. (2000) Characteristics of health maintenance organizations and their influence on efficiency. *Health Services Management Research*, 13(1), 40-56.
58. Drouin, J. P., Hediger, V., Henke, N., Kanzler, L., Leydon, E. F. & De Santis, P. (2008). The health care century. *Health International*, 7, 6-17.
59. Dusheiko, M., Gravelle, H., Martin, S., Rice, N. & Smith, P. C. (2011) Does better disease management in primary care reduce hospital costs? Evidence from English primary care. *Journal of Health Economics*, 30(5), 919-932.
60. Edward, A., Kumar, B., Kakar, F., Salehi, A. S., Burnham, G. & Peters, D. H. (2011). Configuring balanced scorecards for measuring health system performance: evidence from 5 years' evaluation in Afghanistan. *PLoS Medicine*, 8(7), 1-9.
61. Eijkenaar, F. (2011) Key issues in the design of pay for performance programs. *The European Journal of Health Economics*, 14(1), 117–131.
62. Eijkenaar, F., Emmert, M., Scheppach, M. & Schöffski, O. (2013) Effects of pay for performance in health care: A systematic review of systematic reviews. *Health Policy*, 110(2-3), 115–130.
63. Escoval, A. (2010) O processo de contratualização na saúde em Portugal (1996-2005). *Revista Portuguesa de Saúde Pública*, 9, 7-24.
64. Färe, R. & Lovell, C. K. (1978) Measuring the technical efficiency of production. *Journal of Economic Theory*, 19(1), 150–162.
65. Farrell, M. J. (1957) The measurement of productive efficiency. *Journal of the Royal Statistical Society*, Series A (General), 120 (3), 253-281.
66. Ferreira, C., Marques, R. C. & Nicola, P. (2013) On evaluating health centers groups in Lisbon and Tagus Valley: efficiency, equity and quality. *BMC Health Services Research*, 13, 529.
67. Filippetti, A. & Peyrache, A. (2011) The Patterns of Technological Capabilities of Countries: A Dual Approach using Composite Indicators and Data Envelopment Analysis. *World Development*, 39(7), 1108–1121.

68. Furtado, M. C. C., Braz, J. C., Pina, J. C., de Mello, D. F. & de Lima, R.A. (2013) Assessing the care of children under one year old in Primary Health Care. *Revista Latino-Americana de Enfermagem*, 21(2), 554-561.
69. García, F., Marcuello, C., Serrano, D. & Urbina, O. (1999) Evaluation of efficiency in primary health care centres: an application of data envelopment analysis. *Financial Accountability & Management*, 15(1), 67-83.
70. Geboers, H., Mokkink, H., van Montfort, P., van den Hoogen, H., van den Bosch, W. & Grol, R. (2002) Comprehensive assessment of the quality of care in small-scale primary care practices. *Journal of Clinical Governance*, 10, 51-61.
71. Gesler, W., Sherman, J. E., Spencer, J. Preisser, J. S., Arcury, T. A. & Perin, J. (2006) Exploring inequalities in health care coverage by degree of rurality in Western North Carolina. *Southeastern Geographer*, 46(1), 97-120.
72. Gilson, L., Magomi, M. & Mkangaa, E. (1995) The structural quality of Tanzanian primary health facilities. *Bulletin of the World Health Organization*, 73(1), 105-114.
73. Giuffrida, A. (1999) Productivity and efficiency changes in primary care: a Malmquist index approach. *Health Care Management Science*, 2(1), 11-26.
74. Giuffrida, A. & Gravelle, H. (2001) Measuring performance in primary care: econometric analysis and DEA. *Applied Economics*, 33(2), 163-175.
75. Goddard, M. & Jacobs, R. (2009) Using composite indicators to measure performance in health care, in Smith, P.C., Mossialos, E., Papanicolas, I. & Leatherman S. (eds.), *Performance measurement for health system improvement*, Cambridge, Cambridge University Press, 339-368.
76. Goddard, M. & Smith, P. (2001) Equity of access to health care services: Theory and evidence from the UK. *Social Science & Medicine*, 53, 1149-1162.
77. Gök, M. Ş. & Altındağ, E. (2014) Analysis of the cost and efficiency relationship: experience in the Turkish pay for performance system. *The European Journal of Health Economics*, 16(5), 459-469.
78. Gorey, K. M., Luginaah, I. N., Holowaty, E. J., Fung, K. Y. & Hamm, C. (2009) Associations of physician supplies with breast cancer stage at diagnosis and survival in Ontario, 1988 to 2006. *Cancer*, 115(15), 3563–3570
79. Gray, J., Millett, C., Saxena, S., Netuveli, G., Khunti, K. & Majeed, A. (2007) Ethnicity and quality of diabetes care in a health system with universal coverage: population-based cross-sectional survey in primary care. *Journal of General Internal Medicine*, 22(9), 1317-1320.
80. Green, R. H., Doyle, J. R. & Cook, W. D. (1996) Efficiency bounds in Data Envelopment Analysis. *European Journal of Operational Research*, 89(3), 482–490.
81. Guagliardo, M. F., Ronzio, C. R., Cheunge, I., Chackoe, E. & Joseph, J. G. (2004) Physician accessibility: an urban case study of pediatric providers. *Health & Place*, 10(3), 273-283.

82. Gulliford, M., Figueroa-Munoz, J., Morgan, M., Hughes, D., Gibson, B., Beech, R. & Hudson, M. (2002) What does 'access to health care' mean? *Journal of Health Services Research and Policy*, 7(3), 186-188.
83. Gusmano, M. K., Weisz, D., Rodwin, V. G., Lang, J., Qian, M., Bocquier, A., Moysan, V. & Verger, P. (2013) Disparities in access to health care in three French regions. *Health Policy*, 114(1), 31-40.
84. Habib, F., Hanafi, M. I. & El-Sagheer, A. (2011) Antenatal care in primary health care centres in Medina, Saudi Arabia, 2009: a cross-sectional study. *Eastern Mediterranean Health Journal*, 17(3), 196-202.
85. Haggerty, J., Pineault, R., Beaulieu, M. D., Brunelle, Y. , Gauthier, J., Goulet, F. & Rodrigue, J. (2007) Room for improvement: patients' experiences of primary care in Quebec before major reforms. *Canadian Family Physician*, 53(6), 1056.
86. Haggerty, J., Pineault, R., Beaulieu, M. D., Brunelle, Y. , Gauthier, J., Goulet, F. & Rodrigue, J. (2008) Practice features associated with patient-reported accessibility, continuity, and coordination of primary health care. *Annals of Family Medicine*, 6(2), 116-123.
87. Hall, A. G., Lemak, C.H., Steingraber, H. & Schaffer, S. (2008) Expanding the definition of access: it isn't just about health insurance. *Journal of Health Care for the Poor & Underserved*, 19(2), 625-638.
88. Halsteinli, V., Kittelsen, S. A. & Magnussen, J. (2010) Productivity growth in outpatient child and adolescent mental health services: the impact of case-mix adjustment. *Social Science & Medicine*, 70(3), 439-446.
89. Heard, A., Nath, D. K. & Loevinsohn, B. (2013) Contracting urban primary healthcare services in Bangladesh - effect on use, efficiency, equity and quality of care. *Tropical Medicine & International Health*, 18(7), 861-870.
90. Hedeen, A. N., Heagerty, P. J., Fortney J. C., Borowsky, S. J., Walder, D. J. & Chapko, M. K. (2002) VA community-based outpatient clinics: quality of care performance measures. *Medical Care*, 40(7), 570-577.
91. Hollingsworth, B. (2003) Non-parametric and parametric applications measuring efficiency in health care. *Health Care Management Science*, 6(4), 203-218.
92. Hollingsworth, B. (2008) The measurement of efficiency and productivity of health care delivery. *Health Economics*, 17, 1107-1128.
93. Hollingsworth, B., Dawson P. & Maniadakis, N. (1999) Efficiency measurement of health care: a review of nonparametric methods and applications. *Health Care Management Science*, 2(3), 161-172.
94. Huang, Y. G. & McLaughlin, C. P. (1989) Relative efficiency in rural primary health care: an application of data envelopment analysis. *Health Services Research*, 24(2), 143-158.
95. Huerta Munoz, U. & Källestål, C. (2012) Geographical accessibility and spatial coverage modeling of the primary health care network in the Western Province of Rwanda. *International Journal of Health Geographics*, 11, 40.
96. Idzerda, L., Adams, O., Patrick, J., Schrecker, T. & Tugwell, P. (2011) Access to primary healthcare services for the Roma population in Serbia: a secondary data analysis. *BMC International Health and Human Rights*, 11, 10.

97. Iezzoni, L. I. (2009) Risk adjustment for performance measurement, in Smith, P. C., Mossialos, E., Papanicolas, I. & Leatherman S. (eds.), *Performance measurement for health system improvement*, Cambridge, Cambridge University Press, 251-368.
98. Jaakkimainen, R. L., Barnsley, J., Klein-Geltink, J., Kopp, A. & Glazier, R. H. (2011) Did changing primary care delivery models change performance? A population based study using health administrative data. *BMC Family Practice*, 12(1), 44-58.
99. Jacobs, R., Martin, S., Goddard, M., Gravelle, H. & Smith, P. (2006a) Exploring the determinants of NHS performance ratings: lessons for performance assessment systems. *Journal of Health Services Research & Policy*, 11(4), 211-217.
100. Jacobs, R., Smith, P. & Goddard, M. (2004) Measuring performance: An examination of composite performance indicators: A report for the Department of Health. *CHE Technical Paper Series 29.*
101. Jacobs, R., Smith, P. C. & Street, A. (2006b). *Measuring efficiency in health care: analytic techniques and health policy*, Cambridge, Cambridge University Press.
102. Jahangoshai Rezaee, M., Moini, A. & Haji-Ali Asgari, F. (2012) Unified performance evaluation of health centers with integrated model of Data Envelopment Analysis and bargaining game. *Journal of Medical Systems*, 36(6), 3805-3815.
103. Kalinichenko, O., Amado, C. A. F. & Santos, S.P. (2013) Performance assessment in primary health care: A systematic literature review. *CEFAGE-UE Working Paper 2013/03*.
104. Kelaher, M., Dunt, D., Day, S. & Feldman, P. (2006) Effects of financial disadvantage on use and non-use of after hours care in Australia. *Health Policy*, 79(1), 16-23.
105. Keller, S. C., Yehia, B. R., Momplaisir, F. O., Eberhart, M. G., Share, A. & Brady, K. A. (2014) Assessing the overall quality of health care in persons living with HIV in an urban environment. *AIDS Patient Care & STDs*, 28 (4), 198-205.
106. Kern, L. M., Edwards, A. & Kaushal, R. (2014) The patient-centered medical home, electronic health records, and quality of care. *Annals of Internal Medicine*, 160(11), 741-749.
107. Kirigia, J. M., Emrouznejad, A., Sambo, L. G., Munguti, N. & Liambila, W. (2004) Using Data Envelopment Analysis to measure the technical efficiency of public health centers in Kenya. *Journal of Medical Systems*, 28(2), 155-166.
108. Kirigia, J. M., Emrouznejad, A., Vaz, R.G., Bastiene, H. & Padayachy, J. (2008) A comparative assessment of performance and productivity of health centres in Seychelles. *International Journal of Productivity & Performance Management*, 57(1), 72-92.

109. Kirigia, J. M., Sambo, L. G. & Scheel, H. (2001) Technical efficiency of public clinics in Kwazulu-Natal province of South Africa. *East African Medical Journal*, 78(2), S1-S13.
110. Kirigia, J. M., Sambo, L. G., Renner, A., Alemu, W., Seasa, S. & Bah, Y. (2011) Technical efficiency of primary health units in Kailahun and Kenema districts of Sierra Leone. *International Archives of Medicine*, 4(1), 15-28.
111. Kirschner, K., Braspenning, J., Jacobs, J. A. & Grol, R. (2012) Design choices made by target users for a pay-for-performance program in primary care: an action research approach. *BMC Family Practice*, 13, 25.
112. Klemenc-Ketis, Z., Kravos, A., Poplas-Susič, T., Svab, I. & Kersnik, J. (2014) New tool for patient evaluation of nurse practitioner in primary care settings. *Journal of Clinical Nursing*, 23(9-10), 1323-1331.
113. Kontodimopoulos, N., Moschovakis, G., Aletras, V. H. & Niakas, D. (2007) The effect of environmental factors on technical and scale efficiency of primary health care providers in Greece. *Cost Effectiveness & Resource Allocation*, 5, 14.
114. Kontodimopoulos, N., Nanos, P. & Niakas, D. (2006) Balancing efficiency of health services and equity of access in remote areas in Greece. *Health Policy*, 76(1), 49-57.
115. Kringos, D. S., Boerma, W. G. W., Bourgueil, Y., Cartier, T., Hasvold, T., Hutchinson, A., Lember, M., Oleszczyk, M., Pavlic, D. R., Svab, I., Tedeschi, P., Wilson, A., Windak, A., Dedeu, T. & Wilm, S. (2010a) The european primary care monitor: structure, process and outcome indicators. *BMC Family Practice*, 11, 81-88.
116. Kringos, D. S., Boerma, W. G. W., Hutchinson, A., van der Zee, J. & Groenewegen, P. P. (2010b) The breadth of primary care: a systematic literature review of its core dimensions. *BMC Health Services Research*, 10, 65-77.
117. Kumbhakar, S. C. & Lovell, C. A. K. (2003) *Stochastic Frontier Analysis*, Cambridge, Cambridge University Press.
118. Lamarche, P. A., Pineault, R., Haggerty, J., Hamel, M., Levesque, J. F. & Gauthier, J. (2010) The experience of primary health care users: a rural-urban paradox. *Canadian Journal of Rural Medicine*, 15(2), 61-67.
119. Lebrun, L. A., Shi, L., Zhu, J., Sharma, R., Sripipatana, A., Hayashi, A. S., Daly, C. A. & Ngo-Metzger, Q. (2013) Racial/ethnic differences in clinical quality performance among health centers. *Journal of Ambulatory Care Management*, 36(1), 24-34.
120. Lee, J. H., Choi, Y. J., Sung, N. J., Kim, S. Y., Chung, S. H., Kim, J., Jeon, T. H. & Park, H. K. (2009) Development of the Korean primary care assessment tool--measuring user experience: tests of data quality and measurement performance. *International Journal for Quality in Health Care*, 21(2), 103-111.
121. Lemos, E. F., Alves, A. M. S., Oliveira, G. C., Rodrigues, M. P., Martins N. D. G. & Croda, J. (2014) Health-service performance of TB treatment for

indigenous and non-indigenous populations in Brazil: a cross-sectional study. *BMC Health Services Research*, 14, 384-400.

122. Lester, H. & Roland, M. (2009) Performance measurement in primary care, in Smith, P. C., Mossialos, E., Papanicolas, I. & Leatherman S. (eds.), *Performance measurement for health system improvement*, Cambridge, Cambridge University Press, 371-405.
123. Lewis, P. C. & Holcomb, B. (2012) A model for patient-centered army primary care. *Military Medicine*, 177(12), 1502-1507.
124. Linna, M., Nordblad, A. & Koivu, M. (2003) Technical and cost efficiency of oral health care provision in Finnish health centres. *Social Science & Medicine*, 56(2), 343-353.
125. Liu, J. S., Lu, L. Y. Y., Lu, W.-M. & Lin, B. J. Y. (2013) A survey of DEA applications. *Omega*, 41(5), 893-902.
126. Liu, Y., Li, Z., Xiong, H., Gao, X. & Wu, J. (2010) Understanding of Internal Clustering Validation Measures. *2010 IEEE International Conference on Data Mining*.
127. Llanwarne, N. R., Abel, G. A., Elliott, M. N., Paddison, C. A., Lyratzopoulos, G., Campbell, J. L. & Roland, M. (2013) Relationship between clinical quality and patient experience: analysis of data from the English Quality and Outcomes Framework and the national GP Patient Survey. *Annals of Family Medicine*, 11(5), 467-472.
128. Löber, G. & Staat, M. (2010) Integrating categorical variables in Data Envelopment Analysis models: A simple solution technique. *European Journal of Operational Research*, 202(3), 810–818.
129. Luo, W. & Qi, Y. (2009) An enhanced two-step floating catchment area (E2SFCA) method for measuring spatial accessibility to primary care physicians. *Health & Place*, 15(4), 1100-1107.
130. Luo, W. & Wang, F. (2003) Measures of spatial accessibility to health care in a GIS environment: synthesis and a case study in the Chicago region. *Environment & Planning B: Planning & Design*, 30(6), 865-884.
131. Luo, W., Wang, F. & Douglass, C. (2004) Temporal changes of access to primary health care in Illinois (1990–2000) and policy implications. *Journal of Medical Systems*, 28(3), 287-299.
132. Luoma, K. (1996) Financial incentives and productive efficiency in Finnish health centres. *Health Economics*, 5(5), 435-445.
133. Macinko, J., Almeida, C. & de Sá, P. K. (2007) A rapid assessment methodology for the evaluation of primary care organization and performance in Brazil. *Health Policy & Planning*, 22(3), 167-167.
134. Manzoli, L., Flacco, M. E., De Vito, C., Arcà, S., Carle., F., Capasso, L., Marzuillo, C., Muraglia, A., Samani, F. & Villari, P. (2014) AHRQ prevention quality indicators to assess the quality of primary care of local providers: a pilot study from Italy. *European Journal of Public Health*, 24(5), 745-750.

135. Marathe, S., Wan, T. T., Zhang, J. & Sherin, K. (2007) Factors influencing community health centers' efficiency: a latent growth curve modeling approach. *Journal of Medical Systems*, 31(5), 365-374.
136. Marschall, P. & Flessa, S. (2011) Assessing the efficiency of rural health centres in Burkina Faso: an application of Data Envelopment Analysis. *Health Economics Review*, 1, 5.
137. Masiye, F., Kirigia, J. M., Emrouznejad, A., Sambo, L. G., Mounkaila, A., Chimfwembe, D. & Okello, D. (2006) Efficient management of health centres human resources in Zambia. *Journal of Medical Systems*, 30(6), 473-481.
138. Maxwell, R. J. (1992) Dimensions of quality revisited: from thought to action. *Quality in Health Care*, 1(3), 171-177.
139. McGrail, M. R. (2012) Spatial accessibility of primary health care utilising the two step floating catchment area method: an assessment of recent improvements. *International Journal of Health Geographics*, 11, 50-61.
140. McGrail, M. R. & Humphreys, J. S. (2009) Measuring spatial accessibility to primary care in rural areas: Improving the effectiveness of the two-step floating catchment area method. *Applied Geography*, 29(4), 533-541.
141. Mehrotra, A., Sorbero, M. E. & Damberg, C. L. (2010) Using the lessons of behavioral economics to design more effective pay-for-performance programmes. *American Journal of Managed Care*, 16(7), 497-503.
142. Millett, C., Gray, J., Bottle, A. & Majeed, A. (2008) Ethnic disparities in blood pressure management in patients with hypertension after the introduction of pay for performance. *Annals of Family Medicine*, 6(6), 490-496.
143. Millett, C., Gray, J., Wall, M. & Majeed, A. (2009a) Ethnic disparities in coronary heart disease management and pay for performance in UK. *Journal of General Internal Medicine*, 24(1), 8-13.
144. Millett, C., Netuveli, G., Saxena, S. & Majeed, A. (2009b) Impact of pay for performance on ethnic disparities in intermediate outcomes for diabetes: a longitudinal study. *Diabetes Care*, 32(3), 404-409.
145. Milliken, O., Devlin, R. A., Barham, V., Hogg, W., Dahrouge, S. & Russell, G. (2011) Comparative efficiency assessment of primary care service delivery models using Data Envelopment Analysis. *Canadian Public Policy*, 37(1), 85-109.
146. Mills, A., Palmer, N., Gilson, L., McIntyre, D., Schneider, H., Sinanovic, E. & Wadee, H. (2004) The performance of different models of primary care provision in Southern Africa. *Social Science & Medicine*, 59(5), 931-943.
147. Montgomery, J. E., Irish, J. T., Wilson, I. B., Chang, H., Li, A. C., Rogers, W. H. & Safran, D. G. (2004) Primary care experiences of Medicare beneficiaries, 1998 to 2000. *Journal of General Internal M*edicine, 19(10), 991-998.
148. Morais, P. & Camanho, A. S. (2011) Evaluation of performance of European cities with the aim to promote quality of life improvements. *Omega*, 39(4), 398–409.

149. Morris, E. & Landes, D. (2006). The equity of access to orthodontic dental care for children in the North East of England. *Public Health*, 120(4), 359-363.
150. Morris, S., Sutton, M. & Gravelle, H. (2005) Inequity and inequality in the use of health care in England: an empirical investigation. *Social Science & Medicine*, 60(6), 1251-1266.
151. Mosquera, P. A., Hernández, J., Vega, R., Martínez, J. & San Sebastián, M. (2013) Performance evaluation of the essential dimensions of the primary health care services in six localities of Bogota-Colombia: a cross-sectional study. *BMC Health Services Researc*h 13, 315.
152. Murillo-Zamorano, L. R. & Petraglia, C. (2011) Technical efficiency in primary health care: does quality matter? *European Journal of Health Economics*, 12(2), 115-125.
153. Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffmann, A. & Giovannini, E. (2008) *Handbook on Constructing Composite Indicators: Methodology and User Guide.* Paris, OECD Publishing.
154. Newton, J. T., Williams, A. C. & Bower, E. J. (2008) Access to dental services in Scotland: an analysis of dentist-population ratios. *British Dental Journal*, 204(3), E6.
155. Nuti, S., Daraio, C., Speroni, C. & Vainieri, M. (2011) Relationships between technical efficiency and the quality and costs of health care in Italy. *International Journal for Quality in Health Care*, 23(3), 324-330.
156. Observatório Português dos Sistemas de Saúde (2009) *Relatório de Primavera 2009: 10/30 anos: razões para continuar*. Lisbon, ENSP/UNL.
157. OECD (2015) *OECD Reviews of Health Care Quality: Portugal 2015: Raising Standards*, Paris, OECD Publishing.
158. OECD (2016) *Better ways to pay for health care*, Paris, OECD Publishing.
159. Osei, D., D'Almeida, S., George, M. O., Kirigia, J. M., Mensah, A. O. & Kainyu, L. H. (2005) Technical efficiency of public district hospitals and health centres in Ghana: a pilot study. *Cost Effectiveness & Resource Allocation*, 3, 9.
160. Ozcan, Y. A. (1998) Physician benchmarking: measuring variation in practice behavior in treatment of otitis media. *Health Care Management Science*, 1(1), 5-17.
161. Ozcan, Y. A., Jiang, H. J. & Pai, C. W. (2000) Do primary care physicians or specialists provide more efficient care? *Health Services Management Research*, 13(2), 90-96.
162. Pai, C. W., Ozcan, Y. A. & Jiang, H. J. (2000) Regional variation in physician practice pattern: an examination of technical and cost efficiency for treating sinusitis. *Journal of Medical Systems*, 24(2), 103-117.
163. Parchman, M. L. & Culler, S. (1994) Primary care physicians and avoidable hospitalizations. *Journal of Family Practice*, 39(2), 123-129.
164. Pelletier, L. R., Johnson, S. A., Westrick, E. R., Fontaine, E. R., Krinsky, A. D., Klugman, R. A. & Ash, A.S. (2014) Composite model for profiling

physicians across domains of care. *Journal of Healthcare Management*, 59(3), 224-237.

165. Pelone, F., Kringos, D. S., Romaniello, A., Archibugi, M., Salsiri, C. & Ricciardi, W. (2015) Primary care efficiency measurement using Data Envelopment Analysis: A Systematic Review. *Journal of Medical Systems*, 39 (156).

166. Pelone, F., Kringos, D. S., Valerio, L., Romaniello, A., Lazzari, A., Ricciardi, W. & Giulio de Belvis, A. (2012) The measurement of relative efficiency of general practice and the implications for policy makers. *Health Policy*, 107(2/3), 258-268.

167. Perronnin, M., Pichetti, S. & Sermet, C. (2014) Using Data Envelopment Analysis (DEA) to rank GPs according to their P4P quality and efficiency scores. *Value in Health*, 17(7), A416.

168. Perry, B. & Gesler, W. (2000) Physical access to primary health care in Andean Bolivia. *Social Science & Medicine*, 50(9), 1177-1188.

169. Petticrew, M. & Roberts, H. (2006) *Systematic reviews in the social sciences: a practical guide*, Oxford, Blackwell.

170. Pietilä, T., Sintonen, H., Pietilä, I., Widström, E., Varrela, J. & Alanen, P. (1998) Cost and productivity analysis of orthodontic care in Finland. *Community Dentistry & Oral Epidemiology*, 26(4), 283-288.

171. Pina, V. & Torres, L. (1992) Evaluating the efficiency of nonprofit organizations: an application of data envelopment analysis to the public health service. *Financial Accountability & Management*, 8(3), 213-224.

172. Portela, M. C. A. S. & Thanassoulis, E. (2006) Zero weights and non-zero slacks: Different solutions to the same problem. *Annals of Operations Research*, 145(1), 129–147.

173. Prior, D. & Surroca, J. (2007) A reasonable benchmarking frontier using DEA: An incentive scheme to improve efficiency in public hospitals. *DEE – Working Papers, Business Economics Series*, 07-11.

174. Puig-Junoy, J. & Ortún, V. (2004) Cost efficiency in primary care contracting: a stochastic frontier cost function approach. *Health Economics*, 13(12), 1149-1165.

175. Rabetti, A. C. & Freitas, S. F. T. (2011) Evaluation of actions concerning systemic arterial hypertension in primary healthcare. *Revista de Saúde Pública*, 45(2), 258-268.

176. Radford, A., Pink, G. & Ricketts, T. (2007) A comparative performance scorecard for federally funded community health centers in North Carolina. *Journal of Healthcare Management*, 52(1), 20-31.

177. Rahman, M. A. & Capitman, J. A. (2012) Can more use of supporting primary care health practitioners increase efficiency of health clinics? Evidence from California's San Joaquin Valley. *Journal of Health Care Finance*, 38(3), 78-92.

178. Ramírez-Valdivia, M. T., Maturana, S. & Salvo-Garrido, S. (2011) A multiple stage approach for performance improvement of primary healthcare practice. *Journal of Medical Systems*, 35(5), 1015-1028.
179. Ramón, N., Ruiz, J. L. & Sirvent, I. (2010) A multiplier bound approach to assess relative efficiency in DEA without slacks. *European Journal of Operational Research*, 203(1), 261–269.
180. Renner, A., Kirigia, J. M., Zere, E. A., Barry, S. P., Kirigia, D. G., Kamara, C. & Muthuri, L. H. (2005) Technical efficiency of peripheral health units in Pujehun district of Sierra Leone: a DEA application. *BMC Health Services Research*, 5, 77.
181. Ricketts, T. C., Randolph, R., Howard, H. A., Pathman, D. & Carey, T. (2001) Hospitalization rates as indicators of access to primary care. *Health Place*, 7(1), 27-38.
182. Ridde, V., Haddad, S. & Heinmüller, R. (2013) Improving equity by removing healthcare fees for children in Burkina Faso. *Journal of Epidemiology & Community Health*, 67(9), 751-757.
183. Roeger, L. S., Reed, R. L. & Smith, B. P. (2010). Equity of access in the spatial distribution of GPs within an Australian metropolitan city. *Australian Journal of Primary Health*, 16(4), 284-290.
184. Roetzheim, R. G., Pal, N., Durme, D. J. V., Wathington, D., Ferrante, J. M., Gonzalez, E. C. & Krischer, J. P. (2000) Increasing supplies of dermatologists and family physicians are associated with earlier stage of melanoma detection. *Journal of the American Academy of Dermatology*, 43(2), 211–218.
185. Roll, Y., Cook, W. D. & Golany, B. (1991) Controlling Factor Weights in Data Envelopment Analysis. *IIE Transactions*, 23(1), 2–9.
186. Rollins, J., Lee, K., Xu, Y. & Ozcan, Y. A. (2001) Longitudinal study of health maintenance organization efficiency. *Health Services Management Research*, 14(4), 249-262.
187. Rosenman, R. & Friesner, D. (2004) Scope and scale inefficiencies in physician practices. Health Economics, 13(11), 1091-1116.
188. Rosenman, R., Siddharthan, K. & Ahern, M. (1997) Output efficiency of health maintenance organizations in Florida. *Health Economics*, 6(3), 295-302.
189. Ryvicker, M., Gallo, W. T. & Fahs, M. C. (2012) Environmental factors associated with primary care access among urban older adults. *Social Science & Medicine*, 75(5), 914-921.
190. Safran, D. G., Taira, D. A., Rogers, W. H., Kosinski, M., Ware, J. E. & Tarlov, A. R. (1998) Linking primary care performance to outcomes of care. *Journal of Family Practice*, 47(3), 213-220.
191. Safran, D. G., Wilson, I. B., Rogers, W. H., Montgomery, J. E. & Chang, H. (2002) Primary care quality in the Medicare program: comparing the performance of Medicare health maintenance organizations and traditional fee-for-service Medicare. *Archives of Internal Medicine*, 162(7), 757-765.

192. Salinas-Jiménez, J. & Smith, P. (1996) Data envelopment analysis applied to quality in primary health care. *Annals of Operations Research*, 67(1), 141-161.
193. Saronga, H. P., Duysburgh, E., Massawe, S., Dalaba, M. A., Savadogo, G., Tonchev, P., Dong, H., Sauerborn, R. & Loukanova, S. (2014) Efficiency of antenatal care and childbirth services in selected primary health care facilities in rural Tanzania: a cross-sectional study. *BMC Health Services Research*, 14, 96.
194. Sarrico, C. & Dyson, R. (2004) Restricting virtual weights in data envelopment analysis. *European Journal of Operational Research*, 159(1), 17–34.
195. Schäfer, W., Groenewegen, P.P., Hansen, J. & Black, N. (2011). Priorities for health services research in primary care. *Quality in Primary Care*, 19(2), 77-83.
196. Schinnar, A. P., Kamis-Gould, E., Delucia, N. & Rothbard, A. B. (1990) Organizational determinants of efficiency and effectiveness in mental health partial care programs. *Health Services Research*, 25(2), 387-420.
197. Schmacker, E. R. & McKay, N. L. (2008) Factors affecting productive efficiency in primary care clinics. *Health Services Management Research*, 21(1), 60-70.
198. Scott, A., Sivey, P., Ouakrim, D. A., Willenberg, L., Naccarella, L., Furler, J. & Young, D. (2011) The effect of financial incentives on the quality of health care provided by primary care physicians. *Cochrane Database of Systematic Reviews.*
199. Sebastian, M. S. & Lemma, H. (2010) Efficiency of the health extension programme in Tigray, Ethiopia: a data envelopment analysis. *BMC International Health and Human Rights*, 10, 16-23.
200. Seid, M. & Stevens, G. D. (2005) Access to care and children's primary care experiences: results from a prospective cohort study. *Health Services Research*, 40(6), 1758-1780.
201. Sequist, T. D., Cullen, T., Bernard, K., Shaykevich, S., Orav, E. J. & Ayanian, J. Z. (2011) Trends in quality of care and barriers to improvement in the Indian Health Service. *Journal of General Internal Medicine*, 26(5), 480-486.
202. Sequist, T. D., Schneider, E.C., Anastario, M., Odigie, E. G., Marshall, R., Rogers, W. H. & Safran, D. G. (2008) Quality monitoring of physicians: linking patients' experiences of care to clinical quality and outcomes. *Journal of General Internal Medicine*, 23(11), 1784-1790
203. Shah, B. R., Gunraj, N. & Hux J. E. (2003) Markers of access to and quality of primary care for aboriginal people in Ontario, Canada. *American Journal of Public Health*, 93, 798-802.
204. Shahidzadeh-Mahani, A., Omidvari, S., Baradaran, H. R. & Azin, S. A. (2008) Factors affecting quality of care in family planning clinics: A study from Iran. *International Journal for Quality in Health Care*, 20(4), 284-290.

205. Shi, L., Lebrun, L. A., Zhu, J., Hayashi, A. S., Sharma, R., Daly, C. A., Sripipatana, A. & Ngo-Metzger, Q. (2012) Clinical quality performance in U.S. health centers. *Health Services Research*, 47(6), 2225-2249.
206. Shi, L., Lebrun-Harris, L. A., Parasuraman, S. R., Zhu, J. & Ngo-Metzger, Q. (2013) The quality of primary care experienced by health center patients. *Journal of the American Board of Family Medicine*, 26(6), 768-777.
207. Shi, L., Macinko, J., Starfield, B., Politzer, R. & Xu, J. (2005) Primary care, race, and mortality in US states. *Social Science & Medicine*, 61(1), 65–75.
208. Shi, L., Starfield, B., Kennedy, B. P. & Kawachi, I. (1999) Income inequality, primary care, and health indicators. *Journal of Family Practice*, 48(4), 275-284.
209. Shi, L., Starfield, B., Xu, J., Politzer, R. & Regan, J. (2003) Primary care quality: community health center and health maintenance organization. *Southern Medical Journal*, 96(8), 787-795.
210. Shwartz, M., Burgess, J. F. & Zhu, J. (2016) A DEA based composite measure of quality and its associated data uncertainty interval for health care provider profiling and pay-for-performance. *European Journal of Operational Research*, 253(2), 489–502.
211. Siddharthan, K., Ahern, M. & Rosenman, R. (2000) Data envelopment analysis to determine efficiencies of health maintenance organizations. *Health Care Management Science*, 3(1), 23-29.
212. Smith, P. C. (1996) A framework for analysing the measurement of outcome, in Smith, P. (ed.), *Measuring outcome in the public sector*, London, Taylor & Francis, 1-19.
213. Smith, P.C. (2002) *Measuring Up: Improving health system performance in OECD countries*, Paris, OECD Publishing.
214. Smith, P.C., Mossialos, E., Papanicolas, I. & Leatherman S. (eds) (2009) *Performance measurement for health system improvement*, Cambridge, Cambridge University Press.
215. Starfield, B. & Shi, L. (2002) Policy relevant determinants of health: an international perspective. *Health Policy*, 60(3), 201–218.
216. Starfield, B. (1998). *Primary care: balancing health needs, services and technology*, New York, Oxford University Press.
217. Starfield, B. (2009) Primary care and equity in health: the importance to effectiveness and equity of responsiveness to peoples' needs. *Humanity & Society*, 33(1-2), 56–73.
218. Starfield, B., Shi, L. & Macinko, J. (2005) Contribution of primary care to heath systems and health. *The Milbank Quarterly*, 83(3), 457-502.
219. Starfield, B., Weiner, J., Mumford, L. & Steinwachs, D. (1991) Ambulatory care groups: a categorization of diagnoses for research and management. *Health Services Research*, 26(1), 53-74.
220. Stevens, G. D. & Shi, L. (2002) Racial and ethnic disparities in the quality of primary care for children. *Journal of Family Practice*, 51(6), 573.

221. Stevens, G. D., Shi, L., Vane, C. & Peters, A. L. (2014) Do experiences consistent with a medical-home model improve diabetes care measures reported by adult Medicaid patients? *Diabetes Care*, 37(9), 2565-2571.
222. Sutton, M. (2002) Vertical and horizontal aspects of socio-economic inequity in general practitioner contacts in Scotland. *Health Economics*, 11, 537-549.
223. Szczepura, A., Davies, C., Fletcher, J. & Boussofiane, A. (1993) Efficiency and effectiveness in general practice. *Journal of Management in Medicin*e, 7(5), 36-47.
224. Teach, S. J., Guagliardo, M. F., Crain, E. F., McCarter, R. J., Quint, D. M., Shao, C. & Joseph, J. G. (2006) Spatial accessibility of primary care pediatric services in an urban environment: association with asthma management and outcome. *Pediatrics*, 117(4), S78-S85.
225. Testi, A., Fareed, N., Ozcan, Y. A. & Tanfani, E. (2013) Assessment of physician performance for diabetes: a bias-corrected data envelopment analysis model. *Quality in Primary Care*, 21(6), 345-357.
226. Thanassoulis, E., Boussofiane, A. & Dyson, R. G. (1995) Exploring output quality targets in the provision of perinatal care in England using data envelopment analysis. *European Journal of Operational Research*, 80(3), 588-607.
227. Thompson, R. G., Langemeier, L. N., Lee, C.-T., Lee, E. & Thrall, R. M. (1990) The role of multiplier bounds in efficiency analysis with application to Kansas farming. *Journal of Econometrics*, 46(1-2), 93–108.
228. Thompson, R. G., Singleton, F. D., Thrall, R. M. & Smith, B. A. (1986) Comparative Site Evaluations for Locating a High-Energy Physics Lab in Texas. *Interfaces*, 16(6), 35–49.
229. Tilley, C. J. & Chalkley, M. J. (2005) Measuring access to health services: General Dental Services in Scotland. *British Dental Journal*, 199(9), 599-601.
230. Toda, M., Opwora, A., Waweru, E., Noor, A., Edwards, T., Fegan, G., Molyneux, C. & Goodman, C. (2012) Analyzing the equity of public primary care provision in Kenya: variation in facility characteristics by local poverty level. *International Journal for Equity in Health*, 11(1), 75.
231. Tourigny, A., Aubin, M., Haggerty, J., Bonin, L., Morin, D., Reinharz, D., Leduc, Y., St-Pierre, M., Houle, N., Giguère, A., Benounissa, Z. & Carmichael, P. H. (2010) Patients' perceptions of the quality of care after primary care reform: Family medicine groups in Quebec. *Canadian Family Physician*, 56(7), 273-282.
232. Tsai, J., Shi, L., Yu, W. L., Hung, L. M. & Lebrun, L. A. (2010). Physician specialty and the quality of medical care experiences in the context of the Taiwan national health insurance system. *Journal of the American Board of Family Medicine*, 23(3), 402-412
233. Tukey, J.W. (1977) *Exploratory data analysis*, Reading, Massachusetts, Addison-Wesley.

234. Tyler, L. H., Ozcan, Y. A. & Wogen, S. E. (1995) Mental health case management and technical efficiency. *Journal of Medical Systems*, 19(5), 413-423.
235. Ulmer, C., Lewis-Idema, D., Von Worley, A., Rodgers, J., Berger, L. R., Darling, E. J. & Lefkowitz, B. (2000) Assessing primary care content: four conditions common in community health center practice. *Journal of Ambulatory Care Management*, 23(1), 23-38.
236. Utriainen, P. & Widström, E. (1990) Economic aspects of dental care in Finnish health centers. *Community Dentistry & Oral Epidemiology*, 18(5), 235-238.
237. Van Herck, P., De Smedt, D., Annemans, L., Remmen, R., Rosenthal, M. B. & Sermeus, W. (2010) Systematic review: Effects, design choices, and context of pay-for-performance in health care. *BMC Health Services Research*, *10* (247).
238. Varatharajan, D., Thankappan, R. & Jayapalan, S. (2004) Assessing the performance of primary health centres under decentralized government in Kerala, India. *Health Policy & Planning*, 19(1), 41-51.
239. Varela, P., Martins, G. A. & Fávero, L. P. (2010) Production efficiency and financing of public health: an analysis of small municipalities in the state of São Paulo - Brazil. *Health Care Management Science*, 13(2), 112-123.
240. Venning, P., Durie, A., Roland, M., Roberts, C. & Leese, B. (2000) Randomised controlled trial comparing cost effectiveness of general practitioners and nurse practitioners in primary care. *British Medical Journal*, 320(7241), 1048-1053.
241. Villiers, A., Koko-Mhlahlo, K. & Senekal, M. (2005) Nutritional well-being of young children in Duncan Village, East London, South Africa: accessibility of primary health care clinics. *Public Health Nutrition*, 8(5), 520-532.
242. Wagner, J. M. & Shimshak, D. G. (2000) Physician profiling using data envelopment analysis: a case study. *International Journal of Healthcare Technology and Management*, 2(1), 358-374.
243. Wagner, J. M., Shimshak, D. G. & Novak, M. A. (2003) Advances in physician profiling: the use of DEA. *Socio-Economic Planning Sciences*, 37(2), 141-163.
244. Wang, F. & Luo, W. (2005) Assessing spatial and nonspatial factors for healthcare access: towards an integrated approach to defining health professional shortage areas. *Health & Place*, 11(2), 131-146.
245. Wang, J. J., Winther, C. H., Cha, J., McCullough, C. M., Parsons, A. S., Singer, J. & Shih, S. C. (2014a) Patient-centered medical home and quality measurement in small practices. *American Journal of Managed Care*, 20(6), 481-489.
246. Wang, L. (2011) Analysing spatial accessibility to health care: a case study of access by different immigrant groups to primary care physicians in Toronto. *Annals of GIS*, 17(4), 237-251.

247. Wang, W., Shi, L., Yin, A., Lai, Y., Maitland, E. & Nicholas, S. (2014b). Development and validation of the Tibetan Primary Care Assessment Tool. *BioMed Research International*, Article ID 308739.
248. Weiss, E., Haslanger, K. & Cantor, J. C. (2001) Accessibility of primary care services in safety net clinics in New York City. *American Journal of Public Health*, 91(8), 1240-1245.
249. Wensing, M., J. Hermsen, Grol, R. & Szecsenyi, J. (2008) Patient evaluations of accessibility and co-ordination in general practice in Europe. *Health Expectations*, 11(4), 384-390.
250. Whitehead, M. (1990) *The concepts and principles of equity and health*, Copenhagen, WHO Regional Office for Europe.
251. Wong, S. Y. S., Kung, K., Griffiths, S. M., Carthy, T., Wong, M. C., Lo, S. V., Chung, V. C., Goggins, W. B. & Starfield, B. (2010) Comparison of primary care experiences among adults in general outpatient clinics and private general practice clinics in Hong Kong. *BMC Public Health*, 10, 397.
252. Wong, Y.-H. B. & Beasley, J. E. (1990) Restricting Weight Flexibility in Data Envelopment Analysis. *Journal of the Operational Research Society*, 41(9), 829–835.
253. World Health Organization (2008) *The World Health Report 2008 - Primary health care (now more than ever*, Geneva, WHO.
254. World Health Organization (2011) *Global health and aging*, World Health Organization and National Institute of Health.
255. Zallman, L., Ma, J., Xiao, L. & Lasser, K. E. (2010) Quality of US primary care delivered by resident and staff physicians. *Journal of General Internal Medicine*, 25(11), 1193-1197.
256. Zanella, A., Camanho, A. S. & Dias, T. G. (2013) Benchmarking countries' environmental performance. *Journal of the Operational Research Society*, 64, 426–438.
257. Zavras, A. I., Tsakos, G., Economou, C. & Kyriopoulos, J. (2002) Using DEA to evaluate efficiency and formulate policy within a Greek national primary health care network. Data Envelopment Analysis. *Journal of Medical Systems*, 26(4), 285-292.
258. Zuvekas, S. H. & Cohen, J. W. (2010) Paying physicians by capitation: is the past now prologue? *Health Affairs*, 29(9), 1661-1666.

Appendix A: Summary of the reviewed studies

Table A1 Summary of the reviewed studies

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	B	C	D
Aakvik and Holmås (2006)	To examine the relationship between the availability of physicians and health outcomes in Norwegian municipalities	General practitioners	Norway	General practice	Yes	No	No	No
Abrams et al. (1995)	To investigate the administrative and policy implications of managed care for community health centres and to assess the cost and utilisation performance of community health centres in managed care networks	7 community health centres	USA	General practice	No	No	Yes*	No
Akazili et al. (2008)	To determine the degree of efficiency of health centres in Ghana and recommend performance targets for the inefficient facilities	89 public health centres	Ghana	General practice	No	No	Yes	No
Al-Taiar et al. (2010)	To investigate the relationship between different measures of physical access, including straight-line distances, road distances and travel time and the impact of these measures on the vaccination of children in Yemen	Primary care centres	Yemen	Preventive care	Yes	No	No	No

A – accessibility, B – equity, C – efficiency, D – effectiveness

*Efficiency has been analysed through partial ratio-based measures (the first category of efficiency studies)

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Amado and Dyson (2009)	To examine the role of DEA in helping decision makers to understand and improve the performance of primary care practices	14 general practitioner practices	UK	Chronic disease management	No	Yes	Yes	Yes
Amado and Santos (2009)	To contribute to the discussion regarding appropriate ways to compare the performance of primary care providers	351 health care centres	Portugal	General practice	No	Yes	Yes	Yes
Andes et al. (2002)	To develop a model which helps to determine best practices of efficient physician offices while allowing for choices between inputs	115 primary care physician practices	USA	General practice	No	No	Yes	No
Arakawa et al. (2011)	To evaluate patient accessibility to tuberculosis treatment in different referral services in Ribeirão Preto, an interior city in Sao Paulo state	4 referral services of primary health units, district health units and referral centres	Brazil	Acute and non-acute infectious disease management	Yes	No	No	No
Aysola et al. (2013)	To determine if patient-centred medical homes (PCMHs) are associated with improved quality and equity in paediatric primary care	PCMHs	USA	Maternal and child care	Yes	Yes	No	Yes
Baker (1992)	To explain variations in the level of development among general practices in three counties of south west England	287 general practices	UK	General practice	Yes	No	Yes*	No

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Barbaro et al. (2014)	To evaluate prenatal care for adolescents in health units, in accordance with the attributes of Primary Health Care guidelines	1 Basic Health District Unit, 1 School-Based Health Centre, 6 Basic Health Units, 4 Family Health Units	Brazil	Maternal and child care	Yes	No	No	Yes
Bates et al. (1996)	To illustrate the nature of the problems arising when attempting to assess the efficiency with which the prescribing of drugs is achieved by general practitioners in UK National Health Service	106 general practices	UK	General practice	No	No	Yes	No
Bell et al. (2013)	To examine potential (geographical) access to primary health care in the city of Mississauga, Ontario, Canada	677 general practices	Canada	General practice	Yes	No	No	No
Bissonnette et al. (2012)	To examine neighbourhood access to primary health care in the city of Mississauga, Ontario, Canada	677 general practices	Canada	General practice	Yes	No	No	No
Breyer (1977)	To determine the degree to which original goals of a national health centre policy were met and to suggest areas for future improvement	9 health centres	USA	General practice	Yes	No	Yes*	Yes
Bryce et al. (2000)	To describe the efficiency of HMOs and to test the robustness of these findings across alternative models of efficiency	585 HMOs	USA	General practice	No	No	Yes	No

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Buck (2000)	To assess the efficiency with which health authorities' Community Dental Services provide dental care in England	100 Community Dental Services	UK	Dental care	No	No	Yes	No
Buja et al. (2014)	To compare the performance of primary health care services in managing diabetes, congestive heart failure and coronary heart disease, by age group	Local Health Units	Italy	Chronic disease management	No	Yes	No	No
Campbell et al. (2007)	To assess the quality of primary care in England with the introduction of pay-for-performance	42 primary care practices	UK	Chronic disease management	No	No	No	Yes
Carrier et al. (2011)	To test for associations between quality of care and the composition of a physician's patient panel	Non-federal US physicians	USA	General practice	No	Yes	No	No
Chilingerian and Sherman (1996)	To explore the use of DEA to locate best practice physicians and the potential savings if inefficient physicians adopt these best practice patterns	326 primary care physicians working in HMOs	USA	General practice	No	No	Yes	No
Chilingerian and Sherman (1997)	To present a detailed real application that demonstrates how to improve physician profiles with enhanced DEA models	326 primary care physicians working in HMOs	USA	General practice	No	No	Yes	No
Collier et al. (2006)	To benchmark relative performance assessment for primary care physicians who work in a university clinic	16 primary care physicians	USA	General practice	No	No	Yes	Yes

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Cordero-Ferrera et al. (2011)	To extend the existing literature about efficiency measurement in primary health care with the application of a recently developed method to deal with exogenous variables	97 primary care centres	Spain	General practice	No	No	Yes	No
Cordero-Ferrera et al. (2014)	To extend the literature on measuring efficiency in primary health care by considering the influence of quality indicators and environmental variables conjointly in a case study	94 primary care centres	Spain	General practice	No	No	Yes	No
Coughlin et al. (2008)	To explore the relationships between race/ethnicity and area factors affecting access to health care in the United States	US primary care physicians	USA	Preventive care	No	Yes	No	No
Dahrouge et al. (2010)	To evaluate whether gender differences in the primary care experience in each model of primary care exist and whether the extent of gender differences between models differs	137 primary care practices	Canada	General practice	No	Yes	No	No
Dahrouge et al. (2013)	To evaluate the delivery of primary health services for different socio-economic groups and assessed the performance of different organisational models in terms of equality of health care delivery in Ontario, Canada	137 primary care practices	Canada	General practice	No	Yes	No	No

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Defelice and Bradford (1997)	To define whether solo or group practice physicians are relatively more inefficient	924 primary care physicians	USA	General practice	No	No	Yes	No
Deidda et al. (2014)	To analyse the efficiency of primary care centres adopting Information and Communication Technology devices, using a new database on primary care centres in the Basque Region in Spain	130 primary care centres	Spain	General practice	No	No	Yes	No
Draper et al. (2000)	To examine the efficiency of HMOs	249 HMOs	USA	General practice	No	No	Yes	No
Edward et al. (2011)	To illustrate the performance trends in delivering the basic primary health services the first five-year period following elections in 2004 and to reflect on the potential and limitations of the scorecard as a performance management tool to measure and improve health service delivery in similar health care contexts	3 district hospitals,7 comprehensive health centres, 15 basic health centres	Afghanistan	General practice	Yes	No	No	Yes
Ferreira et al. (2013)	To contribute to a better understanding of the performance of Primary Health Care by measuring it in a Portuguese region (Lisbon and Tagus Valley) and identifying best practices	22 health care centres	Portugal	General practice	No	Yes	Yes	Yes
Friedberg et al. (2007)	To examine whether larger physician group size and affiliation with networks of multiple groups are associated with higher quality of care	132 physician groups	USA	General practice	No	No	No	Yes

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Furtado et al. (2013)	To analyse the presence and extent of Primary Health Care attributes and the strength of affiliation of children under one year old in a Family Health Unit	2 School Health centres, 7 Basic Health Units, 5 Family Health Units	Brazil	Maternal and child care	Yes	No	No	Yes
García et al. (1999)	To analyse the efficiency of primary care centres in the province of Zaragoza in Spain	54 primary care centres	Spain	General practice	No	No	Yes	No
Geboers et al. (2002)	To examine the possibility of performing a comprehensive assessment of the quality of care provided in primary care practices	39 primary care practices	Holland	General practice	Yes	No	No	Yes
Gesler et al. (2006)	To examine whether or not there is poorer geographic access to health care in more rural counties and to explore whether rural populations are adequately served in terms of national standards	284 primary care physicians	USA	General practice	No	Yes	No	No
Gilson et al. (1995)	To provide the quantitative analysis of structural quality, giving details of the nature of the quality weaknesses	58 primary health organisations	Tanzania	General practice	Yes	No	No	No
Giuffrida and Gravelle (2001)	To compare different methods of measuring the performance of English FHSAs in providing primary care	90 FHSAs	UK	General practice	Yes	No	Yes	Yes

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Giuffrida (1999)	To explore the productivity change in the provision of primary health care in England between the financial years 1990/91 and 1994/95 using routine administrative data from Family Health Services Authorities (FHSAs) and to decompose it into pure efficiency, technical and scale changes	90 FHSAs	UK	General practice	Yes	No	Yes	Yes
Gray et al. (2007)	To assess the quality of diabetes care and intermediate clinical outcomes within a multiethnic population after a sustained period of investment in quality improvement	32 primary care practices	UK	Chronic disease management	No	Yes	No	No
Guagliardo et al. (2004)	To propose a method for measuring and analysing spatial accessibility to physicians that is easily understood by health policy makers and is particularly useful for congested urban areas	Primary care physicians	USA	Maternal and child care	Yes	No	No	No
Gusmano et al. (2013)	To compare access to primary and specialty care in three metropolitan regions of France and to identify the factors that contribute to disparities in access to care within and among these regions	Primary and specialty care providers	France	General practice	No	Yes	No	No
Habib et al. (2011)	To analyse the pattern of antenatal care consultations at the health centres and the characteristics of the physicians and the women and to determine the predictors of good antenatal care performance	7 primary health care centres	Saudi Arabia	Maternal and child care	No	No	No	Yes

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Haggerty et al. (2007)	To investigate variations in accessibility, continuity of care, and coordination of services as experienced by patients in Quebec on the eve of major reforms, and to provide baseline information against which reforms could be measured	100 primary health care clinics	Canada	General practice	Yes	No	No	Yes
Haggerty et al. (2008)	To identify attributes of clinic organisation and physician practice that predict accessibility, continuity, and coordination of care as experienced by patients	100 primary health care clinics	Canada	General practice	Yes	No	No	Yes
Hall et al. (2008)	To expand the definition of access to physician services, using a survey of primary care providers in the Florida Medicaid program	3222 primary care providers	USA	General practice	Yes	No	No	No
Halsteinli et al. (2010)	To assess productivity growth in Norwegian outpatient child and adolescent mental health service units	From 48 to 60 outpatient child and adolescent mental health service units	Norway	Mental care	No	No	Yes	No
Heard et al. (2013)	To evaluate a large, ongoing effort to improve urban primary health care in Bangladesh through expansion of publicly funded urban health facilities and contracting with partner non-governmental organisations	Urban primary health care centres in non-governmental organisations	Bangladesh	General practice	Yes	No	Yes*	Yes

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	B	C	D
Hedeen et al. (2002)	To compare quality of care provided to veterans at Community-Based Outpatient Clinics and at traditional hospital-based VA Medical Centre clinics	20 Community-Based Outpatient Clinics, 20 VA Medical Centre clinics	USA	General practice	No	No	No	Yes
Huang and McLaughlin (1989)	To apply DEA to the evaluation of rural primary health care programs	77 primary care clinics	USA	General practice	No	No	Yes	No
Huerta Munoz and Källestål (2012)	To measure geographical accessibility, model spatial coverage of the existing primary health facility network, estimate the number of primary health facilities working under capacity and the population underserved in the Western Province of Rwanda	113 primary health care facilities	Rwanda	General practice	Yes	No	No	No
Idzerda et al. (2011)	To assess whether the Roma, as a vulnerable population, are able to effectively access primary care services, and if not, what barriers prevent them from doing so	Primary health care services	Serbia	Maternal and child care	No	Yes	No	No
Jaakkimainen et al. (2011)	To evaluate differences in performance between Family Health Network and Family Health Group and to compare performance before and after physicians joined these new primary care groups	3466 Family Health Network and 474 Family Health Group physicians	USA	General practice	No	No	No	Yes
Jacobs et al. (2006a)	To examine the degree to which variations in the performance of health care organisations are explained by a range of factors that are subject to differing degrees of managerial control	304 Primary Care Trusts	UK	General practice	Yes	No	No	Yes

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Jahangoshai Rezaee et al. (2012)	To introduce a new integrated approach to measure unified efficiency of the healthcare systems	54 health centres	USA	General practice	No	No	Yes	No
Jaiveer et al. (2006)	To assess the performance of 12 local practices using the Alphabet Strategy with respect to the diabetes quality indicators before and after the implementation of GMS2	13 (2004) and 12(2005) general practices	UK	Chronic disease management	No	No	No	Yes
Kelaher et al. (2006)	To examine the effects of socioeconomic disadvantage on access to after hours care and episodes of not seeking after hours care when needed among users and non-users of after hours care	General practitioners and emergency departments	Australia	General practice	No	Yes	No	No
Keller et al. (2014)	To assess the quality of primary care for people living with HIV (PLWH) and to explore factors associated with meeting Health Resources and Services Administration-identified HIV performance measures	Primary HIV medical care	USA	Acute and non-acute infectious disease management	No	No	No	Yes
Kern et al. (2014)	To compare quality of care provided by physicians in PCMHs with that provided by physicians using paper medical re-cords and, separately, with that provided by physicians using electronic health records without the PCMH (to determine whether effects were driven by electronic health records)	675 primary care physicians	USA	General practice	No	No	o	Yes

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Kirigia et al. (2001)	To investigate the technical inefficiencies among 155 primary health care clinics in Kwazulu-Natal Province of South Africa and to draw policy implications	155 primary health care clinics	South Africa	General practice	No	No	Yes	No
Kirigia et al. (2004)	To determine the degree of technical efficiency of individual primary health care facilities in Kenya; to recommend the performance targets for inefficient facilities; to estimate the magnitudes of excess inputs; and to recommend what should be done with those excess inputs	32 public health centres	Kenya	General practice	No	No	Yes	No
Kirigia et al. (2008)	To measure the technical and scale efficiency of health centres; to evaluate changes in productivity; and to highlight possible policy implications of the results for policy makers	17 public health centres	Seychelles	General practice	No	No	Yes	No
Kirigia et al. (2011)	To estimate the technical efficiency of community health centres (CHCs), community health posts (CHPs) and maternal and child health posts (MCHPs) in Kailahun and Kenema districts of Sierra Leone, estimate the output increases needed to make inefficient MCHPs, CHCs and CHPs efficient, and explore strategies for increasing technical efficiency of these institutions	22 community health centres, 21 community health posts,36 maternal and child health posts	Sierra Leone	General practice	No	No	Yes	No
Klemenc-Ketis et al. (2014)	To validate a tool for patient evaluation of nurse practitioners	7 family medicine practices	Slovenia	General practice	No	No	No	Yes

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Kontodimopoulos et al. (2006)	To investigate the efficiency of a set of small-scaled Greek hospitals known as hospital-health centres	17 hospital-health centres	Greece	General practice	No	No	Yes	No
Kontodimopoulos et al. (2007)	To compare technical and scale efficiency of primary care centres from the two largest Greek providers, the National Health System (NHS) and the Social Security Foundation (IKA) and to determine if, and how, efficiency is affected by various exogenous factors such as catchment population and location	194 primary care centres (103 NHS and 91 IKA)	Greece	General practice	No	No	Yes	No
Lamarche et al. (2010)	To assess the care experience of primary health care users, to determine whether assessments vary according to the geographical context in which services are obtained and to determine whether the observed variations are consistent across all components of the care experience	100 primary health care clinics	Canada	General practice	No	Yes	No	No
Lebrun et al. (2013)	To asses racial/ethnic disparities in clinical quality among US health centres	1131 health centres	USA	General practice	No	Yes	No	No
Lee et al. (2009)	To develop a tool for assessing the performance of primary care services in South Korea from the patient's perspective and to test the validity of the tool under the conceptual framework of the recently developed definition of primary care in Korea	16 primary care providers	South Korea	General practice	Yes	No	No	Yes

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Lemos et al. (2014)	To assess the performance of health services for indigenous and non-indigenous populations with regard to tuberculosis control	Primary care providers and specialists	Brazil	Acute and non-acute infectious disease management	No	Yes	No	No
Lewis and Holcomb (2012)	To evaluate a novel method of primary care delivery in a military family practice setting.	14 practice physicians, family nurse practitioners, and physician assistants	USA	General practice	Yes	No	Yes*	Yes
Linna et al. (2003)	To measure the productive efficiency of public dental health provision across Finland and to investigate various factors explaining the technical and cost efficiency of public dental care	228 public dental health centres	Finland	Dental care	No	No	Yes	No
Llanwarne et al. (2013)	To examine the relationship between clinical quality and patient experience using data from 2 established measures of quality in primary care in England	7759 general practices	UK	General practice	Yes	No	No	Yes
Luo and Qi (2009)	To present an enhancement of the two-step floating catchment area (2SFCA) method for measuring spatial accessibility and to measure the spatial access to primary care physicians in northern Illinois	Primary care physicians	USA	General practice	Yes	No	No	No

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	B	C	D
Luo and Wang (2003)	To synthesize two GIS-based accessibility measures into one framework and to apply the methods to examining spatial accessibility to primary health care in the Chicago ten-county region	Primary care physicians	USA	General practice	Yes	No	No	No
Luo et al. (2004)	To examine temporal changes of access to primary health care in Illinois between 1990 and 2000 in a GIS environment	Primary care physicians	USA	General practice	Yes	No	No	No
Luoma (1996)	To examine the productive efficiency of Finnish health centres by applying DEA and econometric methods	202 health centres	Finland	General practice	No	No	Yes	No
Macinko et al. (2007)	To presents methodology for the rapid assessment of the organisation and performance of primary care services at district level	40 primary care clinics	Brazil	General practice	Yes	No	No	Yes
Manzoli et al. (2014)	To compare the performance of local providers, such as Italian local health authorities (LHAs) and health districts, using prevention quality indicators	44 local health units and 11 health districts	Italy	General practice	Yes	No	No	No
Marathe et al. (2007)	To examine factors affecting the variation in technical and cost efficiency of community health centres	493 community health centres	USA	General practice	No	No	Yes	No
Marschall and Flessa (2011)	To determine the relative efficiency of primary care facilities in Nouna, a rural health district in Burkina Faso	20 health centres	Burkina Faso	General practice	No	No	Yes	No

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Masiye et al. (2006)	To estimate the degree of technical, allocative and cost efficiency in individual public and private health centres in Zambia; and to identify the relative inefficiencies in the use of various inputs among individual health centres	40 health centres	Zambia	General practice	No	No	Yes	No
McGrail and Humphreys (2009)	To critically appraise the two-step floating catchment area (2SFCA) method, a recent solution for measuring primary care service accessibility across rural areas of Victoria, Australia	Primary care physicians	Australia	General practice	Yes	No	No	No
McGrail (2012)	To assess measure recent improvements of the measure the two step floating catchment area (2SFCA) method and to measure spatial accessibility of primary health care providers in the state of Victoria, Australia	Primary care physicians	Australia	General practice	Yes	No	No	No
Millett et al. (2008)	To examine ethnic disparities in the management of hypertension among patients with and without cardiovascular comorbidities after the implementation of a major pay-for-performance incentive scheme in UK primary care	16 family practices	UK	Chronic disease management	No	Yes	No	No

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Millett et al. (2009a)	To examine disparities in coronary heart disease management and intermediate clinical outcomes within a multiethnic population before and after the introduction of a major pay for performance initiative in April 2004	36 family practices	UK	Chronic disease management	No	Yes	No	No
Millett et al. (2009b)	To examine the impact of a major pay for performance incentive on trends in the quality of diabetes care in white, black, and South Asian ethnic groups in an urban setting in the U.K.	16 family practices	UK	Chronic disease management	No	Yes	No	No
Milliken et al. (2011)	To compare the relative productive efficiencies of 4 models of primary care service delivery using the DEA method on 130 primary care practices in Ontario, Canada	130 primary care practices	Canada	General practice	No	No	Yes	No
Mills et al. (2004)	To examine the performance of different models of PC provision, in order to identify their strengths and weaknesses from the perspective of a government wishing to develop an overall strategy for improving PC provision	21 primary care sites	South Africa	General practice	Yes	No	No	Yes
Montgomery et al. (2004)	To examine changes in the quality of primary care experienced and reported by Medicare beneficiaries from 1998 to 2000	Primary care providers	USA	General practice	Yes	No	No	Yes

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Morris and Landes (2006)	To assess the equity of access to primary care orthodontic treatment in relation to deprivation in the County Durham and Tees Valley Strategic Health Authority area	General dental practitioners	UK	Dental care	No	Yes	No	No
Morris et al. (2005)	To investigate inequality and inequity in the use of general practitioner consultations, outpatient visits, day cases and inpatient stays in England	Primary care practices	UK	General practice	No	Yes	No	No
Mosquera et al. (2013)	To evaluate the performance of the essential dimensions of the PHC strategy in six localities geographically distributed throughout Bogotá city	Private and public primary health care clinics in Bogota, Colombia	Colombia	General practice	Yes	No	No	Yes
Murillo-Zamorano and Petraglia (2011)	To assess the incidence of quality on the measurement of technical efficiency in the primary health care sector	85 primary care centres	Spain	General practice	No	No	Yes	No
Newton et al. (2008)	To assess inequalities in the availability of NHS dental services between Health Boards in Scotland	2134 general dental practitioners	UK	Dental care	No	Yes	No	No
Nuti et al. (2011)	To measure technical efficiency of Tuscan Local Health Authorities and its relationship with quality and appropriateness of care	12 Local Health Authorities	Italy	General practice	No	No	Yes	No

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Osei et al. (2005)	To estimate the relative technical efficiency and scale efficiency of a sample of public hospitals and health centres in Ghana; and to demonstrate policy implications for health sector policy-makers	17 district hospitals, 17 health centres	Ghana	General practice	No	No	Yes	No
Ozcan (1998)	To analyse physician practice behaviour and develops measures of physician practice efficiency as a basis for improving productivity and reducing costs in otitis media treatment	160 physicians	USA	Acute and non-acute infectious disease management	No	No	Yes	No
Ozcan et al. (2000)	To compare resource utilisation between primary care physicians and specialists in the treatment of Medicaid sinusitis patients in Virginia	176 physicians (152 generalists and 24 otolaryngologists)	USA	Acute and non-acute infectious disease management	No	No	Yes	No
Pai et al. (2000)	To examine the physician efficiency and resulting cost patterns by region	178 physicians	USA	Acute and non-acute infectious disease management	No	No	Yes	No
Parchman and Culler (1994)	To examine the relationship between the availability of primary care physicians and access to health care as measured by the avoidable hospital condition rates	26 health service areas	USA	General practice	Yes	No	No	No

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	B	C	D
Pelletier et al. (2014)	To develop a hierarchical, freighted composite model to reliably compare primary care physicians across domains of care	199 primary care physicians	USA	General practice	No	No	No	Yes
Pelone et al. (2012)	To compare technical efficiency of general practice delivered by the twenty Regions of Italy's decentralized healthcare system and to determine if it was affected by contextual factors	General practices in 20 regions	Italy	General practice	Yes	No	Yes	No
Perry and Gesler (2000)	To assess physical access to primary health care in a remote and impoverished region of Andean Bolivia	Primary health care providers	Bolivia	General practice	Yes	No	No	No
Pietilä et al. (1998)	To investigate the costs of orthodontic care provided for children and adolescents up to the age of 18 by municipal health centres in Finland, and to study the productivity of these services	217 municipal health centres	Finland	Dental care	No	No	Yes	No
Pina & Torres (1992)	To examine the alternative approach to the analysis of the efficiency of the health care organisations	10 health centres	Spain	General practice	No	No	Yes	No
Puig-Junoy & Ortún (2004)	To improve knowledge of cost efficiency in primary care contracting in the Catalan health system	180 primary care teams	Spain	General practice	Yes	No	Yes	No
Rabetti & Freitas (2011)	To evaluate the efficiency of the Family Health Strategy in actions related to hypertension	66 municipalities	Brazil	Chronic disease management	No	No	Yes	No

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Radford et al. (2007)	To develop a scorecard that CHCs could use to evaluate organisational performance, to test the scorecard by providing CHCs with timely information about their own performance and comparative information about the performance of their peers, to assess the perceived usefulness of the scorecard, and to determine the interest in and feasibility of continuing the scorecard	14 community health centres	USA	General practice	Yes	No	Yes*	No
Rahman & Capitman (2012)	To explore the factors that may have contributed to productive efficiency gains of 67 primary health care centres	67 primary health care centres	USA	General practice	No	No	Yes	No
Ramírez-Valdivia et al. (2011)	To empirically determine which factors affect more the technical efficiency of the PHCs managed by each municipality using a multiple step analysis	259 municipalities	Chile	General practice	No	No	Yes	No
Renner et al. (2005)	To measure the technical efficiency and scale efficiency of a sample of public peripheral health units in Sierra Leone	37 peripheral health units	Sierra Leone	General practice	No	No	Yes	No
Ricketts et al. (2001)	To test the concept of ambulatory care sensitive condition rates as markers in a combined rural and urban context and explore their use as health program outcome measure	120 primary care service areas in North Carolina	USA	General practice	Yes	No	No	No

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Ridde et al. (2013)	To evaluate the effects on healthcare access inequities of an intervention exempting children under 5 years from user fees in Burkina Faso	18 primary health care centres	Burkina Faso	Maternal and child care	No	Yes	No	No
Roeger et al. (2010)	To examine the equity of access to GPs in an Australia capital city (Adelaide), utilising specialised geographic information systems	General practitioners	Australia	General practice	No	Yes	No	No
Rollins et al. (2001)	To analyse HMO efficiency longitudinally, using a DEA model, and to compare the various types of HMOs and their efficiency using DEA	36 HMOs	USA	General practice	No	No	Yes	No
Rosenman and Friesner (2004)	To compare the efficiency of single specialty and multispecialty group medical practices	156 primary care practices, 346 speciality care practices	USA	General practice	No	No	Yes	No
Rosenman et al. (1997)	To measure the relative technical efficiencies of 28 HMOs licensed to practice in the State of Florida in the autumn of 1994	28 HMOs	USA	General practice	No	No	Yes	No
Ryvicker et al. (2012)	To identify environmental and socio-demographic factors associated with primary care visits among older adults in New York City	49 Primary Care Service Areas	USA	General practice	No	Yes	No	No
Safran et al. (1998)	To evaluate the relationship between seven defining elements of primary care and three outcomes	Primary care physicians	USA	General practice	Yes	No	No	Yes

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Safran et al. (2002)	To compare the primary care received by seniors in Medicare HMOs with that of seniors in the traditional FFS Medicare program, and among HMOs, and to examine performance differences associated with HMO model-type and profit status	121 HMOs	USA	General practice	Yes	No	No	Yes
Salinas-Jiménez and Smith (1996)	To explore the role of quality indicators in primary care, and to examine the extent to which DEA gives useful insights into FHSA performance in terms of quality	85 FHSA	USA	General practice	Yes	No	Yes	Yes
Saronga et al. (2014)	To assess the actual dimension and distribution of the costs of providing antenatal care and childbirth services in selected rural primary health care facilities in Tanzania	11 health centres and dispensaries	Tanzania	Maternal and child care	No	No	Yes*	No
Schinnar et al. (1990)	To formulate measures of efficiency and effectiveness; examine the possible relationship between the two; and explore organisational, fiscal, and client variables that may account for the variance in these performance measures	54 mental health partial care programs	USA	Mental care	No	No	Yes	Yes
Schmacker (2008)	To examine factors affecting the productive efficiency of primary care clinics	from 85 to 91 primary care clinics (per year), 442 observations	USA	General practice	No	No	Yes	No

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	B	C	D
Sebastian and Lemma (2010)	To estimate the technical efficiency of a sample of health posts in rural Tigray, to identify those factors which might be explaining the efficiency results.	60 health posts	Ethiopia	General practice	No	No	Yes	No
Seid and Stevens (2005)	To examine whether and how different kinds of access to care (financial, potential, and realised) predict parent-report child primary care experiences in an urban community sample	Primary care providers	USA	Maternal and child care	Yes	No	No	No
Sequist et al. (2008)	To evaluate the association between clinical performance and patient experiences in a statewide sample of physician practice sites and a sample of physicians within a large physician group	373 practice sites and 119 individual primary care physicians	USA	General practice	Yes	No	No	Yes
Sequist et al. (2011)	To analyse trends in ambulatory quality of care and physician reports of barriers to quality improvement within the Indian Health Service	740 federally employed physicians	USA	General practice	No	Yes	No	No
Shah et al. (2003)	To evaluate primary care accessibility and quality for Ontario's aboriginal population	Primary care providers	Canada	General practice	No	Yes	No	No
Shahidzadeh-Mahani et al. (2008)	To determine factors affecting quality of family planning services	25 urban Primary Health Care clinics	Iran	Maternal and child care	No	No	No	Yes

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Shi et al. (2003)	To compare the primary health care quality of community health centres and health maintenance organisations in South California in order to elucidate the quality of CHC performance relative to mainstream settings such as the HMO	Community health centres and health maintenance organisations	USA	General practice	Yes	No	No	Yes
Shi et al. (2012)	To describe current clinical quality among the nation's community health centres and to examine health centre characteristics associated with performance excellence	1039 community health centres	USA	General practice	No	No	No	Yes
Shi et al. (2013)	To examine the quality of the primary care experienced by health centre patients and to investigate whether race/ethnicity and insurance coverage are significantly associated with patients' experiences	Federally funded health centres	USA	General practice	No	Yes	No	No
Siddharthan et al. (2000)	To measure the relative technical efficiencies of 164 HMOs licensed to practice in the United States	164 HMOs	USA	General practice	No	No	Yes	No
Staat (2003)	To analyse the differences regarding the efficiency of service provision by Austrian general practitioners	591 physicians	Austria	General practice	No	No	Yes	No
Stevens and Shi (2002)	To compared the quality of primary care experienced specifically by children of different racial and ethnic groups	Primary care physicians	USA	Maternal and child care	No	Yes	No	No

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Stevens et al. (2014)	To evaluate whether patient-reported indicators of care that are consistent with the medical-home model are associated with patient-reported receipt of recommended preventive diabetes services for adults considered vulnerable	Primary care physicians	USA	Chronic disease management	Yes	No	No	Yes
Szczepura et al. (1993)	To describe the exploratory use of a statistical technique called DEA	52 general practices	UK	General practice	No	No	Yes	No
Teach et al. (2006)	To test the hypotheses that higher spacial accessibility to primary care paediatric services among a population of urban and largely disadvantaged children would be associated with (1) more scheduled primary care visits for asthma, (2) better longitudinal asthma management, and (3) fewer unscheduled visits for asthma care	Paediatric primary care providers	USA	Maternal and child care	Yes	No	No	Yes
Testi et al. (2013)	To evaluate the performance of primary care practices in the treatment of their diabetic patients	96 family physicians	Italy	Chronic disease management	No	No	Yes	Yes
Thanassoulis et al. (1995)	To explore the use of DEA for the assessment of units providing perinatal care in England and estimate performance targets for them	83 District Health Authorities	UK	Maternal and child care	No	No	Yes	Yes
Tilley and Chalkley (2005)	To assess the extent of access to the publicly funded General Dental Service by adults in Scotland	General dental services	UK	Dental care	Yes	No	No	No

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Toda et al. (2012)	To assess whether poorer areas have poorer health services by investigating associations between public facility characteristics and the poverty level of the area in which the facility is located	248 public health centres and dispensaries	Kenya	General practice	No	Yes	No	No
Tourigny et al. (2010)	To evaluate how a primary care reform, which aimed to promote interprofessional and interorganisational collaborative practices, affected patients' experiences of the core dimensions of primary care	5 family medicine groups	Canada	General practice	Yes	No	No	Yes
Tsai et al. (2010)	To examine the relationship between physician specialty and the quality of primary medical care experiences	Primary care physicians and specialists	Taiwan	General practice	Yes	No	No	Yes
Tyler et al. (1995)	To assess whether Community Mental Health centres operate at different levels of technical efficiency in the production of case management services	39 community mental health centres	USA	Mental care	No	No	Yes	No
Ulmer et al. (2000)	To discuss results of medical records reviews assessing quality of care at CHCs for acute otitis media, diabetes, asthma, and hypertension	20 community health centres	USA	General practice	No	No	No	Yes
Utriainen and Widström (1990)	To assess the effectiveness, output and costs of dental care at different health centres and to analyse the relationships between these factors	34 health centres	Finland	Dental care	Yes	No	Yes*	Yes

Table A1 Summary of the reviewed studies *(continued)*

Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Varatharajan et al. (2004)	To provide an approach to assess primary health care performance under decentralized government	10 primary health centres	India	General practice	Yes	No	Yes*	Yes
Varela et al. (2010)	To measure the variations in performance of small municipalities in the State of São Paulo, Brazil, regarding the technical efficiency in the use of public funds in public primary health care actions concerning the funding profile, in a scenario of fiscal federalism	359 municipalities	Brazil	General practice	No	No	Yes	No
Venning et al. (2000)	To compare the cost effectiveness of general practitioners and nurse practitioners	20 general practices	UK	General practice	No	No	Yes*	Yes
Villiers et al. (2005)	To contribute to the nutritional well-being of young children living in Duncan Village by investigating factors that influence clinic attendance of mothers and to formulate recommendations for optimisation of accessibility of primary health care clinics in the area	Primary health care clinics	South Africa	Maternal and child care	Yes	No	No	Yes
Wagner and Shimshak (2000)	To evaluate the efficiency of primary care physicians from a managed care organisation	21 primary care physicians	USA	General practice	No	No	Yes	No
Wagner et al. (2003)	To examine physician practices within the organisation using DEA	21 primary care physicians	USA	General practice	No	No	Yes	No
Wang and Luo (2005)	To explore both spatial and non-spatial factors in examining accessibility to primary healthcare in Illinois	Primary care physicians	USA	General practice	No	Yes	No	No

Table A1 Summary of the reviewed studies *(continued)*

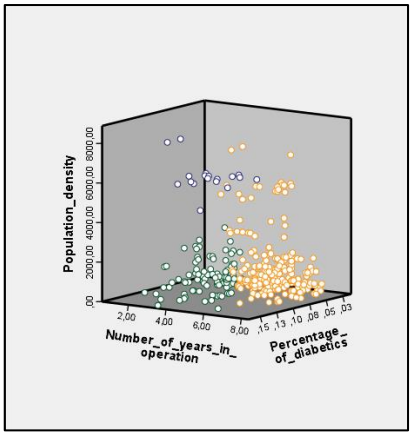
Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Wang (2011)	To analyse the spatial accessibility of a number of immigrant groups to linguistically diverse primary care (family) physicians in the Toronto Census Metropolitan Area	Primary care physicians	Canada	General practice	No	Yes	No	No
Wang et al. (2014a)	To assess performance on quality measures among small primary care practices that recently adopted an electronic health record (EHR), and how performance differs between practices that have achieved patient-centred medical home (PCMH) recognition and those that have not	150 primary care practices	USA	General practice	No	No	No	Yes
Wang et al. (2014b)	To develop a primary care assessment tool in Tibetan area and assess the primary care quality among different healthcare settings	Primary care providers (township health centres, county hospitals, prefecture hospitals)	China	General practice	Yes	No	No	Yes
Weiss et al. (2001)	To determine primary care accessibility for low-income patients	179 ambulatory care facilities	USA	General practice	Yes	No	No	No

Table A1 Summary of the reviewed studies *(continued)*

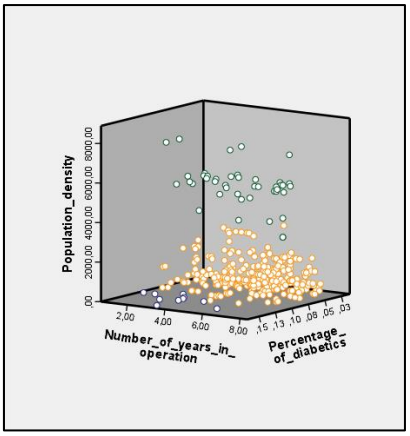
Paper's reference	Study aims	Units of analysis	Country of application	Medical condition studied	Performance dimensions explored in primary care delivery			
					A	**B**	**C**	**D**
Wensing et al. (2008)	To determine whether patient evaluations of the accessibility to general practice and co-ordination with other care providers were associated with characteristics of general practice organisation	284 general practices	Austria, Belgium, England, France, Germany, Israel, The Netherlands, Slovenia, Switzerland, and Wales	General practice	Yes	No	No	Yes
Wong et al. (2010)	To compare the primary care experiences of GOPC users and the users of care provided by private general practitioners in Hong Kong via a territory-wide telephone survey	General outpatient clinics, private general practice clinics	China	General practice	Yes	No	No	Yes
Zallman et al. (2010)	To assess the quality of outpatient care delivered by resident and staff physicians	Resident and staff physicians	USA	General practice	No	No	No	Yes
Zavras et al. (2002)	To critically evaluate the relative efficiency of primary health care centres of the principal Greek public insurance provider, the Social Security Institute	133 primary care centres	Greece	General practice	No	No	Yes	No
Total					**59**	**33**	**75**	**57**

Appendix B: Scatter plots for the distribution of USFs between clusters

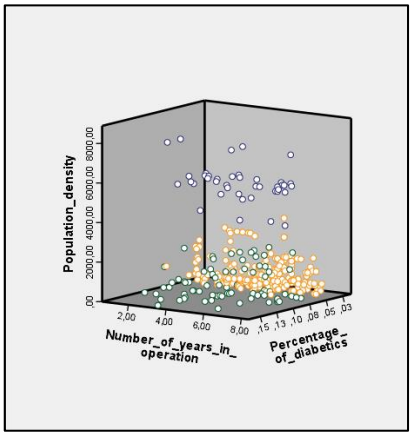
Figure B1 Scatter plots for the distribution of USFs between three clusters (orange dots-large cluster, green dots – medium cluster, blue dots – small cluster)



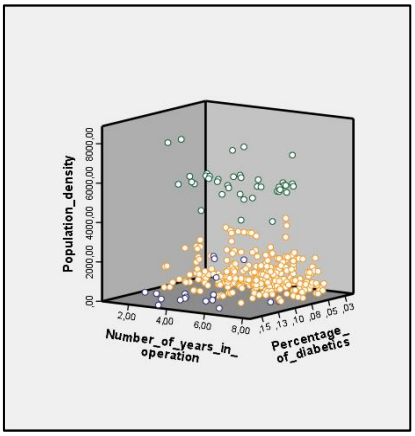
Average linkage within groups



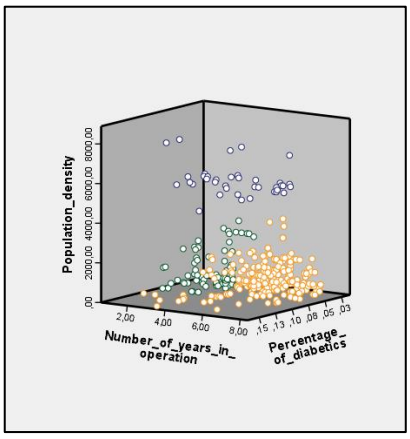
Centroid method



K-means clustering

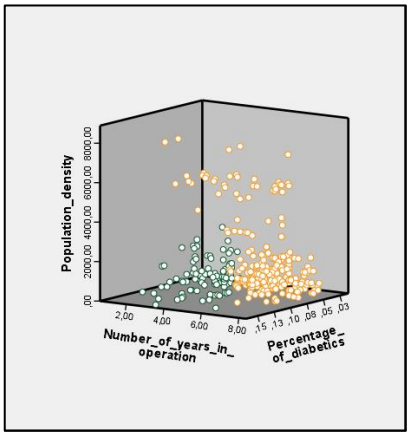


Two-step clustering

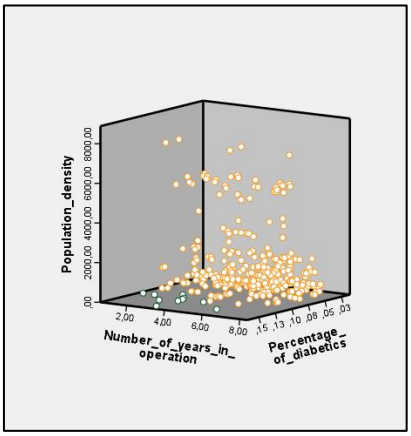


Ward's method

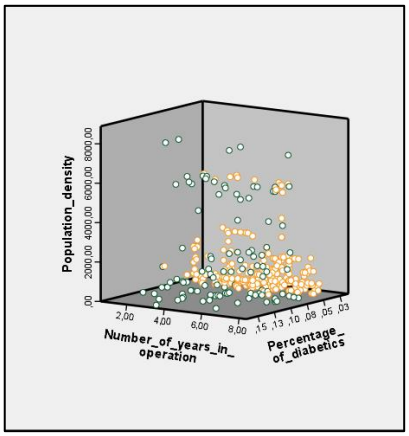
Figure B2 Scatter plots for the distribution of USFs between two clusters (orange dots-large cluster, green dots – small cluster)



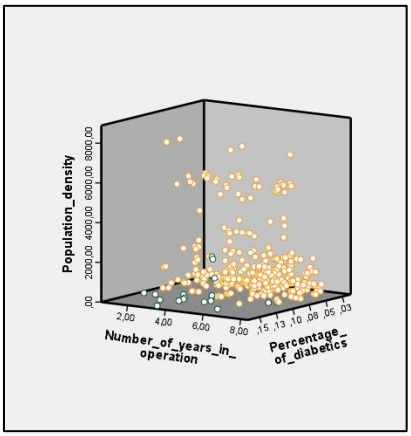
Average linkage within groups



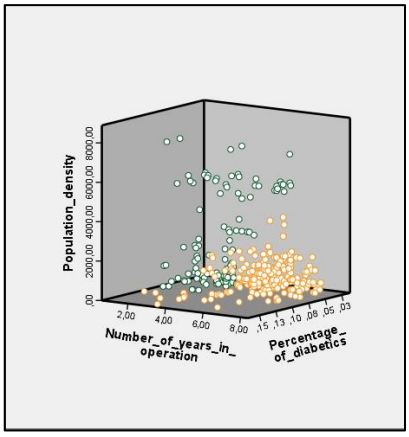
Centroid method



K-means clustering



Two-step clustering



Ward's method

Appendix C: The DEA assessment results of USFs

Table C1 The DEA scores and optimal virtual weights of USFs

USF	DEA score	$\mu_1 y_1$	$\mu_2 y_2$	$\mu_3 y_3$	$\mu_4 y_4$	$\mu_5 y_5$	$\mu_6 y_6$	$\mu_7 y_7$	$\mu_8 y_8$	$\mu_9 y_9$	$\mu_{10} y_{10}$
USF 1	100.00%	2.67%	25.89%	12.61%	8.84%	2.67%	25.89%	4.29%	8.84%	4.52%	3.80%
USF 2	100.00%	6.03%	26.58%	6.03%	11.37%	6.03%	19.86%	6.03%	6.03%	6.03%	6.03%
USF 3	100.00%	3.83%	37.15%	5.19%	3.83%	3.83%	30.85%	3.83%	3.83%	3.83%	3.83%
USF 4	100.00%	6.97%	20.73%	8.16%	14.14%	6.97%	15.14%	6.97%	6.97%	6.97%	6.97%
USF 5	100.00%	6.97%	19.36%	11.68%	11.99%	6.97%	13.57%	6.97%	6.97%	6.97%	8.53%
USF 6	100.00%	5.88%	30.13%	8.10%	5.88%	5.88%	20.58%	5.88%	5.88%	5.88%	5.88%
USF 7	100.00%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 8	100.00%	3.84%	30.86%	8.67%	6.63%	3.18%	30.86%	3.18%	6.41%	3.18%	3.18%
USF 9	99.84%	2.73%	41.60%	2.84%	2.84%	2.73%	36.13%	2.73%	2.84%	2.73%	2.84%
USF 10	99.79%	11.43%	13.57%	11.43%	13.57%	4.52%	13.57%	4.52%	4.52%	11.43%	11.43%
USF 11	99.22%	5.97%	27.54%	9.04%	7.45%	2.94%	27.54%	2.94%	2.94%	6.18%	7.45%
USF 12	98.73%	3.42%	30.49%	9.73%	6.36%	3.42%	30.30%	3.42%	6.02%	3.42%	3.42%
USF 13	98.73%	4.69%	35.08%	5.55%	4.69%	4.69%	26.57%	4.69%	4.69%	4.69%	4.69%
USF 14	98.29%	8.17%	25.50%	8.17%	8.17%	8.17%	9.17%	8.17%	8.17%	8.17%	8.17%
USF 15	98.01%	3.42%	33.17%	6.15%	7.27%	3.42%	30.17%	3.42%	3.42%	3.42%	6.15%
USF 16	97.95%	6.25%	18.75%	18.75%	6.25%	6.25%	18.75%	6.25%	6.25%	6.25%	6.25%
USF 17	97.88%	2.67%	40.71%	2.67%	3.95%	2.67%	36.63%	2.67%	2.67%	2.67%	2.67%
USF 18	97.80%	3.72%	31.42%	11.15%	3.72%	3.72%	31.42%	3.72%	3.72%	3.72%	3.72%
USF 19	97.72%	3.69%	35.78%	5.48%	5.05%	3.69%	26.13%	5.05%	5.05%	5.05%	5.05%
USF 20	97.70%	3.72%	36.07%	5.32%	4.89%	3.72%	26.73%	4.89%	4.89%	4.89%	4.89%
USF 21	97.59%	7.20%	26.99%	8.61%	7.20%	2.78%	26.99%	3.04%	2.78%	7.20%	7.20%
USF 22	97.54%	4.30%	37.09%	4.30%	4.30%	4.30%	28.48%	4.30%	4.30%	4.30%	4.30%
USF 23	97.42%	3.23%	31.33%	10.78%	4.66%	3.23%	28.11%	4.66%	4.66%	4.66%	4.66%
USF 24	97.37%	1.17%	17.84%	15.49%	15.49%	1.17%	15.49%	15.49%	1.17%	15.49%	1.17%
USF 25	97.12%	5.50%	30.84%	7.22%	6.44%	3.18%	30.84%	3.18%	3.18%	6.44%	3.18%
USF 26	97.08%	3.80%	36.83%	5.20%	4.17%	3.80%	29.89%	4.17%	4.17%	4.17%	3.80%
USF 27	96.97%	2.72%	41.37%	2.95%	2.95%	2.72%	35.47%	2.95%	2.95%	2.95%	2.95%
USF 28	96.84%	8.13%	24.38%	9.36%	8.13%	8.13%	9.36%	8.13%	8.13%	8.13%	8.13%
USF 29	96.75%	1.22%	18.56%	18.56%	11.67%	1.22%	18.56%	5.67%	11.67%	11.67%	1.22%
USF 30	96.71%	2.69%	40.90%	3.21%	3.21%	2.69%	35.00%	2.69%	3.21%	3.21%	3.21%
USF 31	96.57%	5.23%	34.16%	5.23%	5.39%	5.23%	23.87%	5.23%	5.23%	5.23%	5.23%
USF 32	96.54%	5.63%	31.60%	5.63%	7.15%	3.26%	24.55%	5.63%	5.63%	5.32%	5.63%
USF 33	96.51%	4.42%	35.39%	5.77%	4.42%	4.42%	27.90%	4.42%	4.42%	4.42%	4.42%
USF 34	96.49%	6.03%	28.00%	9.95%	6.03%	2.89%	23.00%	6.03%	6.03%	6.03%	6.03%
USF 35	96.44%	1.18%	17.98%	17.98%	12.86%	1.18%	17.98%	3.94%	12.86%	12.86%	1.18%
USF 36	96.41%	5.11%	32.77%	7.00%	5.11%	5.11%	24.44%	5.11%	5.11%	5.11%	5.11%
USF 37	96.33%	4.85%	33.52%	4.85%	6.77%	4.85%	25.74%	4.85%	4.85%	4.85%	4.85%
USF 38	96.29%	4.61%	36.17%	4.61%	4.61%	4.61%	26.95%	4.61%	4.61%	4.61%	4.61%
USF 39	96.17%	4.33%	35.49%	4.33%	5.85%	4.33%	28.36%	4.33%	4.33%	4.33%	4.33%
USF 40	96.15%	4.68%	35.97%	4.68%	4.68%	4.68%	26.61%	4.68%	4.68%	4.68%	4.68%
USF 41	95.95%	4.51%	36.46%	4.51%	4.51%	4.51%	27.43%	4.51%	4.51%	4.51%	4.51%
USF 42	95.83%	4.16%	29.22%	12.47%	4.16%	4.16%	29.22%	4.16%	4.16%	4.16%	4.16%
USF 43	95.72%	4.94%	35.19%	4.94%	4.94%	4.94%	25.32%	4.94%	4.94%	4.94%	4.94%

Table C1 The DEA scores and optimal virtual weights of USFs *(continued)*

USF	DEA score	$\mu_1 y_1$	$\mu_2 y_2$	$\mu_3 y_3$	$\mu_4 y_4$	$\mu_5 y_5$	$\mu_6 y_6$	$\mu_7 y_7$	$\mu_8 y_8$	$\mu_9 y_9$	$\mu_{10} y_{10}$
USF 44	95.68%	4.93%	33.76%	6.38%	4.93%	4.93%	25.36%	4.93%	4.93%	4.93%	4.93%
USF 45	95.62%	6.51%	29.71%	7.28%	6.51%	3.06%	27.79%	3.06%	6.51%	6.51%	3.06%
USF 46	95.46%	4.57%	36.30%	4.57%	4.57%	4.57%	27.17%	4.57%	4.57%	4.57%	4.57%
USF 47	95.45%	5.66%	31.26%	7.43%	5.66%	3.22%	31.26%	3.41%	5.66%	3.22%	3.22%
USF 48	95.43%	3.21%	40.36%	3.21%	3.21%	3.21%	33.93%	3.21%	3.21%	3.21%	3.21%
USF 49	95.37%	3.85%	30.76%	3.85%	11.54%	3.85%	30.76%	3.85%	3.85%	3.85%	3.85%
USF 50	95.35%	4.80%	35.57%	4.80%	4.83%	4.80%	25.98%	4.80%	4.80%	4.80%	4.80%
USF 51	95.29%	4.49%	35.36%	5.67%	4.49%	4.49%	27.57%	4.49%	4.49%	4.49%	4.49%
USF 52	95.19%	2.66%	40.53%	2.66%	4.15%	2.66%	36.69%	2.66%	2.66%	2.66%	2.66%
USF 53	95.16%	2.99%	41.03%	2.99%	2.99%	2.99%	35.05%	2.99%	2.99%	2.99%	2.99%
USF 54	95.10%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 55	94.97%	7.23%	27.49%	8.04%	7.23%	2.83%	27.49%	2.83%	2.83%	7.23%	6.77%
USF 56	94.82%	12.50%	12.50%	12.50%	12.50%	0.82%	12.50%	12.50%	0.82%	12.50%	10.86%
USF 57	94.68%	7.17%	27.27%	7.17%	8.38%	2.81%	27.27%	2.81%	2.81%	7.12%	7.17%
USF 58	94.58%	5.51%	31.39%	6.17%	6.93%	3.24%	25.01%	6.17%	3.24%	6.17%	6.17%
USF 59	94.56%	8.15%	25.56%	8.15%	8.15%	8.15%	9.27%	8.15%	8.15%	8.15%	8.15%
USF 60	94.44%	6.85%	28.73%	7.57%	6.85%	2.96%	23.53%	6.85%	2.96%	6.85%	6.85%
USF 61	94.32%	7.15%	28.56%	7.15%	7.15%	3.12%	28.56%	3.12%	7.15%	4.94%	3.12%
USF 62	94.28%	4.61%	36.17%	4.61%	4.61%	4.61%	26.95%	4.61%	4.61%	4.61%	4.61%
USF 63	94.23%	2.64%	40.16%	3.60%	3.60%	2.64%	35.85%	2.64%	2.64%	3.60%	2.64%
USF 64	94.19%	5.62%	32.16%	5.62%	6.59%	4.19%	26.18%	5.62%	4.19%	4.19%	5.62%
USF 65	94.16%	3.06%	40.83%	3.06%	3.06%	3.06%	34.72%	3.06%	3.06%	3.06%	3.06%
USF 66	94.13%	5.11%	33.37%	6.41%	5.11%	5.11%	24.43%	5.11%	5.11%	5.11%	5.11%
USF 67	94.00%	4.52%	34.14%	6.81%	4.52%	4.52%	27.38%	4.52%	4.52%	4.52%	4.52%
USF 68	93.92%	4.27%	36.21%	4.27%	5.26%	4.27%	28.67%	4.27%	4.27%	4.27%	4.27%
USF 69	93.92%	5.61%	32.85%	5.61%	5.93%	5.61%	21.95%	5.61%	5.61%	5.61%	5.61%
USF 70	93.91%	4.59%	32.99%	7.83%	4.59%	3.40%	32.99%	3.40%	3.40%	3.40%	3.40%
USF 71	93.86%	10.04%	14.96%	10.04%	14.96%	2.46%	14.96%	2.46%	10.04%	10.04%	10.04%
USF 72	93.81%	3.84%	25.00%	10.58%	10.58%	3.84%	10.58%	10.58%	10.58%	10.58%	3.84%
USF 73	93.81%	4.84%	34.31%	6.01%	4.84%	4.84%	25.80%	4.84%	4.84%	4.84%	4.84%
USF 74	93.75%	4.95%	35.15%	4.95%	4.95%	4.95%	25.25%	4.95%	4.95%	4.95%	4.95%
USF 75	93.64%	4.82%	35.55%	4.82%	4.82%	4.82%	25.91%	4.82%	4.82%	4.82%	4.82%
USF 76	93.64%	3.01%	40.97%	3.01%	3.01%	3.01%	34.96%	3.01%	3.01%	3.01%	3.01%
USF 77	93.33%	4.85%	35.14%	4.85%	5.16%	4.85%	25.76%	4.85%	4.85%	4.85%	4.85%
USF 78	93.33%	5.90%	31.79%	5.90%	6.40%	3.28%	28.36%	3.28%	5.90%	5.90%	3.28%
USF 79	93.24%	5.19%	32.37%	5.19%	7.26%	3.34%	25.91%	5.19%	5.19%	5.19%	5.19%
USF 80	93.24%	4.89%	34.51%	5.71%	4.89%	4.89%	25.54%	4.89%	4.89%	4.89%	4.89%
USF 81	92.89%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 82	92.87%	4.72%	35.85%	4.72%	4.72%	4.72%	26.42%	4.72%	4.72%	4.72%	4.72%
USF 83	92.81%	12.50%	12.50%	12.50%	12.50%	0.82%	12.50%	10.86%	12.50%	12.50%	0.82%
USF 84	92.81%	8.20%	25.41%	8.20%	8.20%	8.20%	9.02%	8.20%	8.20%	8.20%	8.20%
USF 85	92.75%	5.51%	32.78%	5.51%	6.20%	5.51%	22.46%	5.51%	5.51%	5.51%	5.51%
USF 86	92.75%	6.14%	30.36%	6.14%	7.37%	3.13%	25.66%	6.14%	3.13%	5.81%	6.14%
USF 87	92.73%	5.31%	32.23%	5.31%	7.15%	5.31%	23.45%	5.31%	5.31%	5.31%	5.31%
USF 88	92.69%	3.68%	35.65%	3.68%	7.00%	3.68%	31.62%	3.68%	3.68%	3.68%	3.68%
USF 89	92.67%	5.34%	33.68%	5.64%	5.34%	5.34%	23.29%	5.34%	5.34%	5.34%	5.34%
USF 90	92.58%	8.69%	15.18%	10.94%	15.18%	1.00%	15.18%	10.94%	10.94%	10.94%	1.00%

Table C1 The DEA scores and optimal virtual weights of USFs *(continued)*

USF	DEA score	$\mu_1 y_1$	$\mu_2 y_2$	$\mu_3 y_3$	$\mu_4 y_4$	$\mu_5 y_5$	$\mu_6 y_6$	$\mu_7 y_7$	$\mu_8 y_8$	$\mu_9 y_9$	$\mu_{10} y_{10}$
USF 91	92.57%	6.46%	29.19%	7.89%	6.46%	3.01%	24.61%	6.46%	3.01%	6.46%	6.46%
USF 92	92.35%	3.02%	40.93%	3.02%	3.02%	3.02%	34.88%	3.02%	3.02%	3.02%	3.02%
USF 93	92.32%	8.13%	24.39%	9.35%	8.13%	8.13%	9.35%	8.13%	8.13%	8.13%	8.13%
USF 94	92.23%	12.04%	13.87%	12.04%	12.04%	0.91%	12.04%	12.04%	12.04%	12.04%	0.91%
USF 95	92.21%	5.49%	33.14%	5.87%	5.49%	5.49%	22.54%	5.49%	5.49%	5.49%	5.49%
USF 96	92.21%	4.85%	33.56%	6.73%	4.85%	4.85%	25.73%	4.85%	4.85%	4.85%	4.85%
USF 97	92.12%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 98	92.03%	5.57%	30.19%	5.57%	8.68%	5.57%	22.16%	5.57%	5.57%	5.57%	5.57%
USF 99	92.02%	5.31%	18.94%	18.94%	6.81%	5.31%	18.94%	6.81%	6.81%	6.81%	5.31%
USF 100	92.01%	5.26%	33.91%	5.56%	5.26%	5.26%	23.68%	5.26%	5.26%	5.26%	5.26%
USF 101	91.95%	3.17%	40.48%	3.17%	3.17%	3.17%	34.13%	3.17%	3.17%	3.17%	3.17%
USF 102	91.93%	4.74%	34.69%	5.82%	4.74%	4.74%	26.28%	4.74%	4.74%	4.74%	4.74%
USF 103	91.90%	5.57%	26.43%	8.49%	9.52%	2.72%	24.85%	2.72%	2.72%	8.49%	8.49%
USF 104	91.88%	4.85%	34.29%	4.85%	6.00%	4.85%	25.73%	4.85%	4.85%	4.85%	4.85%
USF 105	91.79%	6.23%	30.54%	7.01%	6.23%	3.15%	28.10%	6.23%	3.15%	6.23%	3.15%
USF 106	91.76%	5.23%	33.22%	6.31%	5.23%	5.23%	23.83%	5.23%	5.23%	5.23%	5.23%
USF 107	91.73%	1.23%	18.75%	15.01%	15.01%	1.23%	15.01%	15.01%	2.51%	15.01%	1.23%
USF 108	91.72%	5.16%	34.51%	5.16%	5.16%	5.16%	24.18%	5.16%	5.16%	5.16%	5.16%
USF 109	91.68%	4.39%	36.83%	4.39%	4.39%	4.39%	28.05%	4.39%	4.39%	4.39%	4.39%
USF 110	91.65%	4.47%	35.24%	4.47%	5.81%	4.47%	27.63%	4.47%	4.47%	4.47%	4.47%
USF 111	91.63%	4.29%	34.95%	6.48%	4.29%	4.29%	28.57%	4.29%	4.29%	4.29%	4.29%
USF 112	91.61%	5.96%	26.43%	8.81%	8.81%	5.96%	8.81%	8.81%	8.81%	8.81%	8.81%
USF 113	91.58%	5.29%	32.67%	5.92%	6.13%	3.37%	28.06%	5.92%	3.37%	5.92%	3.37%
USF 114	91.55%	5.25%	34.24%	5.25%	5.25%	5.25%	23.74%	5.25%	5.25%	5.25%	5.25%
USF 115	91.51%	4.72%	35.83%	4.72%	4.72%	4.72%	26.39%	4.72%	4.72%	4.72%	4.72%
USF 116	91.50%	1.64%	25.03%	11.66%	11.66%	1.64%	11.66%	11.66%	11.66%	11.66%	1.71%
USF 117	91.47%	5.14%	34.50%	5.14%	5.22%	5.14%	24.29%	5.14%	5.14%	5.14%	5.14%
USF 118	91.46%	5.05%	32.50%	5.05%	7.39%	5.05%	24.74%	5.05%	5.05%	5.05%	5.05%
USF 119	91.41%	4.77%	35.25%	4.77%	5.22%	4.77%	26.17%	4.77%	4.77%	4.77%	4.77%
USF 120	91.40%	4.34%	35.27%	6.05%	4.34%	4.34%	28.32%	4.34%	4.34%	4.34%	4.34%
USF 121	91.35%	1.13%	17.13%	14.88%	16.87%	1.13%	16.87%	1.13%	1.13%	14.88%	14.88%
USF 122	91.28%	4.92%	34.74%	5.42%	4.92%	4.92%	25.40%	4.92%	4.92%	4.92%	4.92%
USF 123	91.23%	4.27%	37.20%	4.27%	4.27%	4.27%	28.66%	4.27%	4.27%	4.27%	4.27%
USF 124	91.22%	5.49%	32.20%	6.12%	6.18%	3.32%	27.79%	6.12%	3.32%	6.12%	3.32%
USF 125	91.17%	3.14%	30.48%	11.70%	4.68%	3.14%	28.15%	4.68%	4.68%	4.68%	4.68%
USF 126	91.12%	8.16%	24.49%	9.18%	8.16%	8.16%	9.18%	8.16%	8.16%	8.16%	8.16%
USF 127	91.10%	4.80%	34.44%	5.96%	4.80%	4.80%	26.00%	4.80%	4.80%	4.80%	4.80%
USF 128	90.94%	4.52%	36.43%	4.52%	4.52%	4.52%	27.39%	4.52%	4.52%	4.52%	4.52%
USF 129	90.91%	8.28%	25.03%	8.34%	8.34%	8.28%	8.34%	8.34%	8.34%	8.34%	8.34%
USF 130	90.91%	9.35%	21.96%	9.35%	9.35%	4.95%	12.06%	4.95%	9.35%	9.35%	9.35%
USF 131	90.85%	5.31%	33.64%	5.74%	5.31%	5.31%	23.44%	5.31%	5.31%	5.31%	5.31%
USF 132	90.81%	8.15%	25.54%	8.15%	8.15%	8.15%	9.23%	8.15%	8.15%	8.15%	8.15%
USF 133	90.73%	5.27%	32.68%	6.77%	5.27%	3.37%	29.08%	5.27%	3.63%	5.27%	3.37%
USF 134	90.73%	5.56%	32.62%	5.56%	6.26%	5.56%	22.20%	5.56%	5.56%	5.56%	5.56%
USF 135	90.66%	3.24%	31.46%	10.23%	5.07%	3.24%	30.13%	3.24%	5.07%	5.07%	3.24%
USF 136	90.65%	8.25%	24.75%	8.75%	8.25%	8.25%	8.75%	8.25%	8.25%	8.25%	8.25%
USF 137	90.64%	5.29%	34.13%	5.29%	5.29%	5.29%	23.56%	5.29%	5.29%	5.29%	5.29%

Table C1 The DEA scores and optimal virtual weights of USFs *(continued)*

USF	DEA score	$\mu_1 y_1$	$\mu_2 y_2$	$\mu_3 y_3$	$\mu_4 y_4$	$\mu_5 y_5$	$\mu_6 y_6$	$\mu_7 y_7$	$\mu_8 y_8$	$\mu_9 y_9$	$\mu_{10} y_{10}$
USF 138	90.48%	5.44%	31.62%	5.44%	7.50%	5.44%	22.81%	5.44%	5.44%	5.44%	5.44%
USF 139	90.39%	3.78%	32.84%	8.54%	4.84%	3.78%	27.93%	4.84%	3.78%	4.84%	4.84%
USF 140	90.38%	12.04%	12.96%	12.96%	12.04%	1.34%	12.96%	5.66%	5.97%	12.04%	12.04%
USF 141	90.29%	7.93%	23.78%	7.93%	10.37%	7.93%	10.37%	7.93%	7.93%	7.93%	7.93%
USF 142	90.25%	5.31%	30.91%	8.47%	5.31%	5.31%	23.44%	5.31%	5.31%	5.31%	5.31%
USF 143	90.23%	5.04%	31.27%	5.04%	8.65%	5.04%	24.80%	5.04%	5.04%	5.04%	5.04%
USF 144	90.19%	12.04%	13.87%	12.04%	12.04%	0.91%	12.04%	0.91%	12.04%	12.04%	12.04%
USF 145	90.14%	4.57%	35.10%	5.75%	4.57%	4.57%	27.13%	4.57%	4.57%	4.57%	4.57%
USF 146	90.10%	4.82%	35.53%	4.82%	4.82%	4.82%	25.88%	4.82%	4.82%	4.82%	4.82%
USF 147	89.89%	6.12%	29.71%	8.06%	6.12%	3.06%	26.17%	3.06%	6.12%	6.12%	5.48%
USF 148	89.71%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 149	89.67%	4.98%	35.06%	4.98%	4.98%	4.98%	25.09%	4.98%	4.98%	4.98%	4.98%
USF 150	89.58%	5.27%	34.18%	5.27%	5.27%	5.27%	23.63%	5.27%	5.27%	5.27%	5.27%
USF 151	89.53%	5.30%	33.76%	5.64%	5.30%	5.30%	23.50%	5.30%	5.30%	5.30%	5.30%
USF 152	89.49%	5.04%	33.13%	6.79%	5.04%	5.04%	24.80%	5.04%	5.04%	5.04%	5.04%
USF 153	89.46%	5.16%	33.71%	5.16%	5.98%	3.48%	25.90%	5.16%	5.16%	5.16%	5.16%
USF 154	89.46%	8.10%	25.69%	8.10%	8.10%	8.10%	9.48%	8.10%	8.10%	8.10%	8.10%
USF 155	89.35%	4.81%	35.58%	4.81%	4.81%	4.81%	25.96%	4.81%	4.81%	4.81%	4.81%
USF 156	89.33%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 157	89.31%	2.85%	41.44%	2.85%	2.85%	2.85%	35.74%	2.85%	2.85%	2.85%	2.85%
USF 158	89.22%	6.44%	28.86%	6.44%	8.25%	3.82%	25.66%	3.82%	3.82%	6.44%	6.44%
USF 159	89.22%	3.22%	40.33%	3.22%	3.22%	3.22%	33.88%	3.22%	3.22%	3.22%	3.22%
USF 160	89.08%	10.25%	14.75%	14.75%	10.25%	0.97%	14.75%	10.25%	10.25%	10.25%	3.53%
USF 161	89.07%	10.25%	19.26%	10.25%	10.25%	4.08%	11.11%	10.25%	10.25%	10.25%	4.08%
USF 162	89.07%	4.50%	36.51%	4.50%	4.50%	4.50%	27.52%	4.50%	4.50%	4.50%	4.50%
USF 163	89.06%	5.69%	32.13%	5.69%	6.50%	5.69%	21.56%	5.69%	5.69%	5.69%	5.69%
USF 164	89.02%	11.42%	13.58%	13.58%	11.42%	1.40%	13.58%	1.40%	10.78%	11.42%	11.42%
USF 165	88.81%	5.98%	31.03%	7.01%	5.98%	3.20%	22.89%	5.98%	5.98%	5.98%	5.98%
USF 166	88.72%	7.64%	15.83%	15.83%	10.70%	1.04%	15.83%	1.04%	10.70%	10.70%	10.70%
USF 167	88.60%	7.90%	23.69%	10.52%	7.90%	7.90%	10.52%	7.90%	7.90%	7.90%	7.90%
USF 168	88.57%	2.88%	41.37%	2.88%	2.88%	2.88%	35.61%	2.88%	2.88%	2.88%	2.88%
USF 169	88.52%	4.35%	34.91%	5.37%	5.37%	4.35%	24.17%	5.37%	5.37%	5.37%	5.37%
USF 170	88.48%	7.84%	23.53%	10.78%	7.84%	7.84%	10.78%	7.84%	7.84%	7.84%	7.84%
USF 171	88.47%	12.04%	13.87%	12.04%	12.04%	0.91%	12.04%	0.91%	12.04%	12.04%	12.04%
USF 172	88.46%	5.27%	34.20%	5.27%	5.27%	5.27%	23.66%	5.27%	5.27%	5.27%	5.27%
USF 173	88.45%	4.96%	35.12%	4.96%	4.96%	4.96%	25.19%	4.96%	4.96%	4.96%	4.96%
USF 174	88.44%	10.59%	14.41%	14.41%	10.59%	0.95%	14.41%	10.59%	2.88%	10.59%	10.59%
USF 175	88.36%	4.81%	35.58%	4.81%	4.81%	4.81%	25.96%	4.81%	4.81%	4.81%	4.81%
USF 176	87.90%	7.89%	23.66%	7.89%	10.57%	7.89%	10.57%	7.89%	7.89%	7.89%	7.89%
USF 177	87.90%	7.86%	23.58%	10.70%	7.86%	7.86%	10.70%	7.86%	7.86%	7.86%	7.86%
USF 178	87.86%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 179	87.84%	4.87%	35.38%	4.87%	4.87%	4.87%	25.64%	4.87%	4.87%	4.87%	4.87%
USF 180	87.81%	5.07%	34.80%	5.07%	5.07%	5.07%	24.66%	5.07%	5.07%	5.07%	5.07%
USF 181	87.65%	3.79%	24.09%	11.96%	10.15%	3.79%	11.96%	10.15%	3.79%	10.15%	10.15%
USF 182	87.59%	4.62%	36.13%	4.62%	4.62%	4.62%	26.89%	4.62%	4.62%	4.62%	4.62%
USF 183	87.57%	5.56%	30.70%	5.56%	8.19%	4.48%	24.36%	5.56%	4.48%	5.56%	5.56%
USF 184	87.54%	4.35%	36.95%	4.35%	4.35%	4.35%	28.25%	4.35%	4.35%	4.35%	4.35%

Table C1 The DEA scores and optimal virtual weights of USFs *(continued)*

USF	DEA score	$\mu_1 y_1$	$\mu_2 y_2$	$\mu_3 y_3$	$\mu_4 y_4$	$\mu_5 y_5$	$\mu_6 y_6$	$\mu_7 y_7$	$\mu_8 y_8$	$\mu_9 y_9$	$\mu_{10} y_{10}$
USF 185	87.51%	9.73%	20.82%	9.73%	9.73%	1.37%	9.73%	9.73%	9.73%	9.73%	9.73%
USF 186	87.42%	4.07%	35.63%	6.24%	4.07%	4.07%	29.67%	4.07%	4.07%	4.07%	4.07%
USF 187	87.38%	4.72%	33.03%	4.72%	7.54%	4.72%	26.42%	4.72%	4.72%	4.72%	4.72%
USF 188	87.35%	4.31%	34.80%	4.31%	6.57%	4.31%	28.44%	4.31%	4.31%	4.31%	4.31%
USF 189	87.33%	5.08%	34.77%	5.08%	5.08%	5.08%	24.62%	5.08%	5.08%	5.08%	5.08%
USF 190	87.27%	5.36%	33.10%	6.19%	5.36%	5.36%	23.22%	5.36%	5.36%	5.36%	5.36%
USF 191	87.18%	5.37%	33.01%	6.25%	5.37%	5.37%	23.13%	5.37%	5.37%	5.37%	5.37%
USF 192	87.15%	5.12%	29.95%	9.80%	5.12%	3.09%	26.42%	5.12%	5.12%	5.12%	5.12%
USF 193	87.08%	4.35%	36.96%	4.35%	4.35%	4.35%	28.26%	4.35%	4.35%	4.35%	4.35%
USF 194	87.03%	5.45%	31.45%	7.64%	5.45%	5.45%	22.73%	5.45%	5.45%	5.45%	5.45%
USF 195	86.98%	3.69%	35.83%	5.24%	5.24%	3.69%	25.36%	5.24%	5.24%	5.24%	5.24%
USF 196	86.81%	6.10%	31.26%	6.53%	6.10%	3.22%	25.25%	6.10%	3.22%	6.10%	6.10%
USF 197	86.80%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 198	86.76%	3.73%	32.56%	9.98%	3.73%	3.73%	31.33%	3.73%	3.73%	3.73%	3.73%
USF 199	86.63%	3.99%	38.04%	3.99%	3.99%	3.99%	30.06%	3.99%	3.99%	3.99%	3.99%
USF 200	86.48%	5.03%	34.90%	5.03%	5.03%	5.03%	24.84%	5.03%	5.03%	5.03%	5.03%
USF 201	86.44%	3.03%	40.90%	3.03%	3.03%	3.03%	34.83%	3.03%	3.03%	3.03%	3.03%
USF 202	86.31%	5.06%	34.81%	5.06%	5.06%	5.06%	24.68%	5.06%	5.06%	5.06%	5.06%
USF 203	86.22%	9.38%	15.62%	15.62%	9.38%	1.03%	15.62%	9.38%	5.21%	9.38%	9.38%
USF 204	86.02%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 205	85.92%	3.70%	35.92%	5.19%	5.19%	3.70%	25.54%	5.19%	5.19%	5.19%	5.19%
USF 206	85.87%	6.30%	31.11%	6.30%	6.30%	5.59%	19.94%	6.30%	5.59%	6.30%	6.30%
USF 207	85.77%	5.40%	33.79%	5.40%	5.40%	5.40%	22.98%	5.40%	5.40%	5.40%	5.40%
USF 208	85.58%	7.36%	22.08%	13.20%	7.36%	7.36%	13.20%	7.36%	7.36%	7.36%	7.36%
USF 209	85.50%	4.13%	35.44%	6.30%	4.13%	4.13%	29.33%	4.13%	4.13%	4.13%	4.13%
USF 210	85.41%	12.04%	13.87%	12.04%	12.04%	0.91%	12.04%	0.91%	12.04%	12.04%	12.04%
USF 211	85.36%	5.42%	33.53%	5.42%	5.63%	5.42%	22.90%	5.42%	5.42%	5.42%	5.42%
USF 212	85.28%	5.28%	34.16%	5.28%	5.28%	5.28%	23.59%	5.28%	5.28%	5.28%	5.28%
USF 213	85.18%	4.77%	35.68%	4.77%	4.77%	4.77%	26.13%	4.77%	4.77%	4.77%	4.77%
USF 214	85.11%	5.35%	32.99%	6.32%	5.35%	5.35%	23.27%	5.35%	5.35%	5.35%	5.35%
USF 215	84.99%	4.98%	35.07%	4.98%	4.98%	4.98%	25.11%	4.98%	4.98%	4.98%	4.98%
USF 216	84.99%	5.61%	33.16%	5.61%	5.61%	5.61%	21.93%	5.61%	5.61%	5.61%	5.61%
USF 217	84.90%	5.35%	32.42%	6.89%	5.35%	5.35%	23.25%	5.35%	5.35%	5.35%	5.35%
USF 218	84.71%	6.19%	31.42%	6.19%	6.19%	2.06%	31.42%	2.06%	2.06%	6.19%	6.19%
USF 219	84.68%	4.92%	33.35%	4.92%	6.81%	4.92%	25.41%	4.92%	4.92%	4.92%	4.92%
USF 220	84.64%	6.50%	18.50%	18.50%	6.50%	5.50%	18.50%	6.50%	6.50%	6.50%	6.50%
USF 221	84.59%	3.08%	40.77%	3.08%	3.08%	2.68%	35.42%	2.68%	3.08%	3.08%	3.08%
USF 222	84.58%	5.02%	33.35%	5.02%	6.60%	5.02%	24.88%	5.02%	5.02%	5.02%	5.02%
USF 223	84.47%	5.52%	33.44%	5.52%	5.52%	5.52%	22.40%	5.52%	5.52%	5.52%	5.52%
USF 224	84.35%	5.33%	32.00%	7.34%	5.33%	3.30%	27.41%	5.33%	3.30%	5.33%	5.33%
USF 225	84.21%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 226	84.15%	4.75%	35.75%	4.75%	4.75%	4.75%	26.25%	4.75%	4.75%	4.75%	4.75%
USF 227	84.15%	1.91%	29.01%	9.54%	9.54%	1.91%	9.92%	9.54%	9.54%	9.54%	9.54%
USF 228	84.14%	11.07%	13.95%	11.07%	13.91%	1.44%	13.91%	1.44%	11.07%	11.07%	11.07%
USF 229	84.11%	7.00%	27.55%	7.00%	8.45%	2.84%	19.17%	7.00%	7.00%	7.00%	7.00%
USF 230	83.99%	12.50%	12.50%	12.50%	12.50%	0.82%	12.50%	0.82%	10.86%	12.50%	12.50%

Table C1 The DEA scores and optimal virtual weights of USFs *(continued)*

USF	DEA score	$\mu_1 y_1$	$\mu_2 y_2$	$\mu_3 y_3$	$\mu_4 y_4$	$\mu_5 y_5$	$\mu_6 y_6$	$\mu_7 y_7$	$\mu_8 y_8$	$\mu_9 y_9$	$\mu_{10} y_{10}$
USF 231	83.93%	5.65%	33.06%	5.65%	5.65%	3.41%	24.01%	5.65%	5.65%	5.65%	5.65%
USF 232	83.90%	4.98%	35.07%	4.98%	4.98%	4.98%	25.12%	4.98%	4.98%	4.98%	4.98%
USF 233	83.84%	12.04%	13.87%	12.04%	12.04%	0.91%	12.04%	0.91%	12.04%	12.04%	12.04%
USF 234	83.76%	5.03%	34.90%	5.03%	5.03%	5.03%	24.84%	5.03%	5.03%	5.03%	5.03%
USF 235	83.55%	5.49%	32.36%	6.66%	5.49%	5.49%	22.54%	5.49%	5.49%	5.49%	5.49%
USF 236	83.23%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 237	83.20%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 238	82.88%	10.17%	19.48%	10.17%	10.17%	4.65%	10.17%	10.17%	10.17%	10.17%	4.65%
USF 239	82.81%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 240	82.75%	9.64%	21.08%	9.64%	9.64%	1.38%	11.48%	9.64%	9.64%	9.64%	8.21%
USF 241	82.72%	4.27%	34.83%	6.63%	4.27%	4.27%	28.63%	4.27%	4.27%	4.27%	4.27%
USF 242	82.71%	9.56%	21.31%	9.57%	9.56%	2.20%	9.57%	9.56%	9.56%	9.56%	9.56%
USF 243	82.51%	5.85%	32.26%	6.04%	5.85%	3.33%	25.79%	5.85%	5.85%	5.85%	3.33%
USF 244	82.45%	7.09%	27.96%	7.85%	7.09%	2.88%	27.17%	2.88%	2.88%	7.09%	7.09%
USF 245	82.19%	8.23%	25.30%	8.23%	8.23%	8.23%	8.84%	8.23%	8.23%	8.23%	8.23%
USF 246	82.06%	3.71%	35.96%	5.17%	5.17%	3.71%	27.09%	3.71%	5.17%	5.17%	5.17%
USF 247	82.03%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 248	81.88%	5.08%	33.25%	6.58%	5.08%	5.08%	24.58%	5.08%	5.08%	5.08%	5.08%
USF 249	81.81%	4.17%	29.13%	12.52%	4.17%	4.17%	29.13%	4.17%	4.17%	4.17%	4.17%
USF 250	81.74%	4.95%	35.16%	4.95%	4.95%	4.95%	25.26%	4.95%	4.95%	4.95%	4.95%
USF 251	81.67%	5.30%	34.11%	5.30%	5.30%	5.30%	23.52%	5.30%	5.30%	5.30%	5.30%
USF 252	81.54%	5.26%	34.22%	5.26%	5.26%	5.26%	23.69%	5.26%	5.26%	5.26%	5.26%
USF 253	81.37%	4.29%	37.14%	4.29%	4.29%	4.29%	28.57%	4.29%	4.29%	4.29%	4.29%
USF 254	81.30%	12.04%	13.87%	12.04%	12.04%	0.91%	12.04%	0.91%	12.04%	12.04%	12.04%
USF 255	81.11%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 256	80.85%	12.04%	13.87%	12.04%	12.04%	0.91%	12.04%	0.91%	12.04%	12.04%	12.04%
USF 257	80.69%	7.69%	26.92%	7.69%	7.69%	7.69%	11.53%	7.69%	7.69%	7.69%	7.69%
USF 258	80.61%	5.92%	32.25%	5.92%	5.92%	5.92%	20.42%	5.92%	5.92%	5.92%	5.92%
USF 259	80.45%	12.50%	12.50%	12.50%	12.50%	0.82%	12.50%	10.86%	0.82%	12.50%	12.50%
USF 260	80.28%	2.72%	41.33%	2.98%	2.98%	2.72%	35.64%	2.72%	2.98%	2.98%	2.98%
USF 261	80.11%	5.43%	30.04%	5.43%	9.10%	3.10%	25.18%	5.43%	5.43%	5.43%	5.43%
USF 262	80.11%	3.48%	33.72%	9.33%	3.48%	3.48%	32.62%	3.48%	3.48%	3.48%	3.48%
USF 263	80.01%	2.47%	37.64%	2.47%	7.42%	2.47%	37.64%	2.47%	2.47%	2.47%	2.47%
USF 264	79.73%	5.96%	32.12%	5.96%	5.96%	5.96%	20.19%	5.96%	5.96%	5.96%	5.96%
USF 265	79.71%	5.26%	34.23%	5.26%	5.26%	5.26%	23.72%	5.26%	5.26%	5.26%	5.26%
USF 266	79.53%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 267	79.46%	5.29%	34.14%	5.29%	5.29%	5.29%	23.57%	5.29%	5.29%	5.29%	5.29%
USF 268	79.40%	4.84%	35.49%	4.84%	4.84%	3.66%	28.18%	3.66%	4.84%	4.84%	4.84%
USF 269	79.20%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 270	79.10%	5.15%	33.13%	6.56%	5.15%	3.42%	25.97%	5.15%	5.15%	5.15%	5.15%
USF 271	79.01%	4.02%	35.57%	4.02%	6.40%	4.02%	29.92%	4.02%	4.02%	4.02%	4.02%
USF 272	78.93%	7.48%	27.56%	7.48%	7.48%	1.81%	18.26%	7.48%	7.48%	7.48%	7.48%
USF 273	78.83%	5.75%	32.76%	5.75%	5.75%	3.38%	23.64%	5.75%	5.75%	5.75%	5.75%
USF 274	78.58%	4.93%	35.22%	4.93%	4.93%	4.93%	25.37%	4.93%	4.93%	4.93%	4.93%
USF 275	78.47%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 276	78.36%	4.30%	37.09%	4.30%	4.30%	3.82%	28.97%	4.30%	4.30%	4.30%	4.30%
USF 277	78.23%	5.56%	33.33%	5.56%	5.56%	3.44%	24.34%	5.56%	5.56%	5.56%	5.56%

Table C1 The DEA scores and optimal virtual weights of USFs *(continued)*

USF	DEA score	$\mu_1 y_1$	$\mu_2 y_2$	$\mu_3 y_3$	$\mu_4 y_4$	$\mu_5 y_5$	$\mu_6 y_6$	$\mu_7 y_7$	$\mu_8 y_8$	$\mu_9 y_9$	$\mu_{10} y_{10}$
USF 278	78.15%	5.51%	33.46%	5.51%	5.51%	5.51%	22.44%	5.51%	5.51%	5.51%	5.51%
USF 279	78.13%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 280	78.12%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 281	78.05%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 282	77.92%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 283	77.80%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 284	77.74%	4.76%	35.73%	4.76%	4.76%	4.76%	26.22%	4.76%	4.76%	4.76%	4.76%
USF 285	77.68%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 286	77.59%	11.74%	14.79%	11.74%	11.74%	1.52%	11.74%	1.52%	11.74%	11.74%	11.74%
USF 287	77.57%	3.44%	33.41%	6.57%	6.57%	3.44%	20.26%	6.57%	6.57%	6.57%	6.57%
USF 288	77.47%	5.53%	33.42%	5.53%	5.53%	5.53%	22.36%	5.53%	5.53%	5.53%	5.53%
USF 289	77.42%	12.04%	13.87%	12.04%	12.04%	0.91%	12.04%	0.91%	12.04%	12.04%	12.04%
USF 290	77.20%	4.03%	35.37%	6.58%	4.03%	4.03%	29.87%	4.03%	4.03%	4.03%	4.03%
USF 291	77.03%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 292	76.72%	4.65%	36.04%	4.65%	4.65%	4.65%	26.73%	4.65%	4.65%	4.65%	4.65%
USF 293	76.65%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 294	76.65%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 295	76.63%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 296	76.28%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 297	76.14%	4.67%	36.00%	4.67%	4.67%	3.71%	27.62%	4.67%	4.67%	4.67%	4.67%
USF 298	75.76%	7.76%	26.73%	7.76%	7.76%	7.76%	11.21%	7.76%	7.76%	7.76%	7.76%
USF 299	75.59%	3.48%	33.74%	3.48%	9.30%	3.48%	32.61%	3.48%	3.48%	3.48%	3.48%
USF 300	75.54%	9.90%	20.29%	9.90%	9.90%	1.33%	17.63%	1.33%	9.90%	9.90%	9.90%
USF 301	75.48%	7.78%	26.66%	7.78%	7.78%	7.78%	11.09%	7.78%	7.78%	7.78%	7.78%
USF 302	75.37%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 303	75.01%	9.55%	21.36%	9.55%	9.55%	1.40%	10.42%	9.55%	9.55%	9.55%	9.55%
USF 304	74.94%	5.06%	32.44%	5.06%	7.45%	3.34%	28.14%	5.06%	3.34%	5.06%	5.06%
USF 305	74.57%	2.79%	41.63%	2.79%	2.79%	2.73%	36.16%	2.73%	2.79%	2.79%	2.79%
USF 306	74.36%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 307	74.14%	5.66%	20.70%	11.82%	11.82%	1.36%	11.82%	1.36%	11.82%	11.82%	11.82%
USF 308	73.92%	7.27%	21.80%	7.27%	13.66%	7.27%	13.66%	7.27%	7.27%	7.27%	7.27%
USF 309	73.87%	9.41%	21.77%	9.41%	9.41%	6.18%	9.41%	6.18%	9.41%	9.41%	9.41%
USF 310	73.68%	9.73%	20.82%	9.73%	9.73%	1.37%	9.73%	9.73%	9.73%	9.73%	9.73%
USF 311	73.25%	1.70%	25.90%	11.20%	11.20%	1.70%	12.99%	1.70%	11.20%	11.20%	11.20%
USF 312	72.73%	3.40%	32.99%	10.20%	3.40%	3.40%	32.99%	3.40%	3.40%	3.40%	3.40%
USF 313	72.59%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 314	71.77%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 315	71.12%	8.19%	25.43%	8.19%	8.19%	8.19%	9.05%	8.19%	8.19%	8.19%	8.19%
USF 316	70.93%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 317	70.21%	4.70%	35.89%	4.70%	4.70%	4.70%	26.48%	4.70%	4.70%	4.70%	4.70%
USF 318	70.15%	5.41%	33.76%	5.41%	5.41%	5.41%	22.94%	5.41%	5.41%	5.41%	5.41%
USF 319	69.69%	3.40%	32.99%	10.20%	3.40%	3.40%	32.99%	3.40%	3.40%	3.40%	3.40%
USF 320	68.80%	3.97%	38.08%	3.97%	3.97%	3.93%	30.23%	3.97%	3.93%	3.97%	3.97%
USF 321	68.16%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 322	67.92%	3.97%	38.09%	3.97%	3.97%	3.97%	30.15%	3.97%	3.97%	3.97%	3.97%
USF 323	67.20%	5.71%	32.86%	5.71%	5.71%	3.39%	23.76%	5.71%	5.71%	5.71%	5.71%
USF 324	67.11%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%

Table C1 The DEA scores and optimal virtual weights of USFs *(continued)*

USF	DEA score	$\mu_1 y_1$	$\mu_2 y_2$	$\mu_3 y_3$	$\mu_4 y_4$	$\mu_5 y_5$	$\mu_6 y_6$	$\mu_7 y_7$	$\mu_8 y_8$	$\mu_9 y_9$	$\mu_{10} y_{10}$
USF 325	66.67%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 326	66.38%	12.04%	13.87%	12.04%	12.04%	0.91%	12.04%	0.91%	12.04%	12.04%	12.04%
USF 327	66.10%	4.63%	36.10%	4.63%	4.63%	4.63%	26.83%	4.63%	4.63%	4.63%	4.63%
USF 328	65.90%	4.66%	36.03%	4.66%	4.66%	4.66%	26.72%	4.66%	4.66%	4.66%	4.66%
USF 329	65.88%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 330	65.70%	9.56%	21.32%	9.56%	9.56%	2.20%	9.56%	9.56%	9.56%	9.56%	9.56%
USF 331	63.56%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 332	63.14%	9.56%	21.32%	9.56%	9.56%	2.20%	9.56%	9.56%	9.56%	9.56%	9.56%
USF 333	62.60%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 334	62.56%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 335	62.24%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 336	61.99%	12.04%	13.87%	12.04%	12.04%	0.91%	12.04%	0.91%	12.04%	12.04%	12.04%
USF 337	61.80%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 338	61.59%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 339	59.61%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 340	59.59%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 341	59.00%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 342	58.58%	8.33%	25.00%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%	8.33%
USF 343	58.44%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 344	57.52%	6.44%	30.67%	6.44%	6.44%	6.44%	17.78%	6.44%	6.44%	6.44%	6.44%
USF 345	54.73%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 346	52.16%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 347	51.47%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 348	48.91%	3.40%	32.99%	3.40%	10.20%	3.40%	32.99%	3.40%	3.40%	3.40%	3.40%
USF 349	47.52%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 350	44.78%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 351	43.97%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 352	41.90%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 353	41.33%	2.74%	41.77%	2.74%	2.74%	2.74%	36.28%	2.74%	2.74%	2.74%	2.74%
USF 354	39.69%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%
USF 355	37.97%	5.07%	34.78%	5.07%	5.07%	3.59%	27.61%	3.59%	5.07%	5.07%	5.07%
USF 356	32.26%	3.94%	38.19%	3.94%	3.94%	3.94%	30.31%	3.94%	3.94%	3.94%	3.94%