博士 (経営学) 学位論文

MEASURING EFFICIENCY OF PUBLIC SERVICE ORGANIZATIONS UNDER REFORM AND AUSTERITY: THE CASES OF ENGLISH HOSPITALS, NEW ZEALAND DISTRICT HEALTH BOARDS, AND JAPANESE NATIONAL UNIVERSITIES

「改革と緊縮財政下での公共サービス組織の効率性評価：英国病院、ニュージーランド地区保健委員会、および日本の国立大学の事例」

2020年 9月

Thai Quoc Khanh

東京都立大学大学院　　社会科学研究科

ACKNOWLEDGMENT

First and foremost, I would like to express my deepest gratitude to my supervisor, Professor Masayoshi Noguchi, for his invaluable instructions and dedicated support as well as encouragements throughout the three years of my Ph.D. study. He did not only enthusiastically mentor me with his profound knowledge but also provided me favorable research conditions, which are the decisive factors for the completion of my doctoral program.

I would like to extend my sincere appreciation to my co-supervisor, Dr. Joseph Drew, University of Technology Sydney. I have learned a lot from his rich academic experiences and supportive advice. His dedicated support and constructive criticism have offered me more insights into carrying out research.

I am particularly grateful for the assistance given by Professor Paul Rouse, University of Auckland. His guidance and experiences on performance measurement in the healthcare sector have had a significant contribution to my study.

My thanks also go to the Tokyo Metropolitan Government for providing me a generous scholarship. I am also indebted to Da Nang People's Committee and Department of Finance of Da Nang city for providing me the necessary conditions and assistance to facilitate my study overseas.

I also thank my lab mates, Mr. Tran Thien Vu and Ms. Dana Kathleen McQuestin, for their constructive comments and useful discussions, which helped me fortify my understanding throughout my study.

Finally, I would like to express my special gratitude to my beloved wife, my children, and my entire family for their unconditional love. I am so lucky to have such immeasurable sacrifice and persistent supports that without them, my study would have been impossible.

Tokyo, September 2020

Thai Quoc Khanh

SUMMARY

Measuring the performance of public service organizations has increasingly become more intensive since the emergence of the New Public Management (NPM) in the 1980's. Since public service organizations have been often perceived as inefficient and unresponsive in meeting policymakers' demand and citizens' needs, measuring the public service efficiency is expected to continuously improve the performance of public service organizations and increase public support. Therefore, the need to measure the efficiency of these organizations, *inter alia*, has been a key concern of policymakers, managers, taxpayers, and other stakeholders. In addition, the pressure on public services' efficiency has increased following the financial crises in 2008, social-demographic change, climate change, and technological change. Accordingly, in this study, we investigate the changes and determinants of technical efficiency of the public service organizations in the three countries experiencing the structural reforms coupled with the tightening budget and socio-economic change, including English hospitals, New Zealand District Health Boards (DHBs), and Japanese national universities.

As structural reform is a component of the public management reform, we reviewed key movements in the public management reform since the 1980s, including New Public Management, Neo-Weberian state, and New Public Governance. Also, reforms in the context of the study (the UK healthcare, New Zealand healthcare, and Japanese national universities) are more elaborated. It is likely that the core elements of NPM have been maintained and employed; the emergence of new reform waves such as Neo-Weberian state and New Public Governance do not totally break with NPM but rather complement additional features or modify certain aspects. In addition, we also summarized the key aspects of financial management reform, including performance budgeting, the modernization of the accounting system, and the responses to the financial crisis. Although there have been different approaches to the reform, improving the performance

is considered as the ultimate goal, and the performance measurement has been a growing concern. Therefore, we closely summarized the literature related to performance measurement and performance management in the public sector. In this regard, we also described the approaches applied to measure technical efficiency, specifically to the two-stage Data Envelopment Analysis. Along with quantitative analysis used in measuring efficiency, qualitative content analysis is also employed to complement the findings and policy implications.

Following the overview of public management reforms and performance management, we investigated the manifestations of performance paradox in the implementation of performance measurement in the New Zealand DHBs, Australian hospitals, and Vietnamese hospitals. The semi-structured interviews were first conducted (including two New Zealand DHBs, one Australian hospital, and two Vietnamese hospitals), and then the transcripts were used for content analysis. Based on the evidence of performance paradox, such as intended errors, unintended errors, and synecdoche, it is suggested that performance indicators should focus more on healthcare outcomes, especially patient experiences. Moreover, our key contribution to the literature was that the exogenous factors such as epidemics, natural disasters, media, and political scrutiny might also affect the perception and interpretation of performance against the targets. In addition, we found that a succinct measure such as technical efficiency score can benchmark the efficiency of each unit and provide useful information for performance management. This also provides an underlying reason for the following three main empirical analyses.

The first empirical research is based on measuring the technical efficiency of English acute foundation trusts from 2009 to 2016 when the English healthcare sector experienced unprecedented fiscal hardship and structural reform. Employing the Data Envelopment Analysis (DEA) approach, we estimated the technical efficiency of 70 English acute foundation trusts and found an improvement in efficiency over the entire period examined,

which mainly resulted from the reduction in the number of hospital beds and staff. In addition, a two-stage analysis was adopted to provide insights on the determinants of the efficiency scores, which is likely to be a major concern for policymakers and hospital managers. Accordingly, the results obtained from the two-stage analysis suggested factors such as patient characteristics and asset utilization have a significant influence on efficiency. Therefore, in parallel with the optimization of hospital beds and the hospital length of stay, hospitals should better manage fixed assets (building and information technology) such as carefully evaluating asset conditions and rearranging hospital sites and hospital services to achieve higher efficiency. In the absence of recent DEA studies on English hospitals (most of them were conducted prior to 2006 and used crude outputs), especially those employed second-stage analysis to identify the determinants of efficiency, our study, therefore, complements the current literature on UK hospitals. Also, by adjusting hospital outputs for complexity (rather using crude outputs), we expect that the reliability of efficiency scores estimated can be improved. In addition, along with the factors that have been well examined in the extant literature (e.g. old patients, bed occupancy), our study made pioneering efforts to identify a number of internal factors (e.g. asset utilization, diversity of services provided) which can be more easily targeted by hospital managers to improve efficiency.

With a similar area – the healthcare sector – the second research is based on measuring the technical efficiency of New Zealand DHBs. However, New Zealand DHBs are different from English acute foundation trusts since they are a combination of local authorities and hospitals. While DHBs consume around three-quarters of the public health budget, measuring the efficiency of the healthcare sector in New Zealand has been a challenging issue and elusive. Two-stage DEA analysis was also employed to measure efficiency and finds the determinants of efficiency. Based on efficiency scores estimated, New Zealand DHBs seem to have improved their efficiency from 2013 to 2016. It seems that the success

in controlling the personnel cost and savings in non-clinical services expenses contributed to improvement in efficiency of DHBs. We also found the association between assets, demographics, financial plan, and performance against targets and efficiency. Therefore, DHBs may further improve efficiency through deliberate budget preparation, better utilization and sufficient investment in buildings and information technology, and optimization of patient flow at emergency departments. As previous studies mainly focus on different aspects of New Zealand healthcare, our study contributes to the current literature by providing a more comprehensive picture of the efficiency of the New Zealand healthcare sector. In addition, whilst attempts to measure the efficiency of the New Zealand healthcare sector has been tough going; in the context of budget constraints and the pressure of meeting growing demand, suggestions to improve the efficiency can be of great interest to policy-makers as well as decision-makers.

In the final research, we expand the efficiency measurement to public higher education. In this case, Japanese national universities were selected as they did not only experience structural reform and fiscal distress but were also characterized by a reduction in student numbers, intensive competition, and language barriers. In this context, it is assumed that national universities were motivated to improve their efficiency. However, it seems that little is known about the technical efficiency of Japanese national universities and unlike many other developed countries, there is a scarcity of DEA literature applied to measure the performance of public universities in Japan. Therefore, along with using the two-stage DEA, we augmented findings and policy implications through content analysis using semi-structured interviews. In contradiction to our expectations, we could not identify any systematic decline or improvement in the technical efficiency of Japanese national universities from 2010 to 2016. Probably, national universities were unable to contain resources usage in response to the shrinking in student numbers. The investigation of the factors that may have possibly influenced efficiency suggested a higher proportion of

government grants can reduce efficiency, and universities with less than 10,000 students, with attached hospitals or science faculty, are less likely to be efficient. Therefore, national universities should less rely on government grants, use resources more efficiently, increase the scale through mergers or recruitment of overseas students, and improve the performance of the attached hospitals. In addition, the government should scrutinize funding regulations to mitigate disparity and inefficiency, direct and provide guidelines to streamline science faculties. Thus, our study might be among the pioneer attempts to provide insight understanding of whether Japanese national universities use their resources efficiently, which complement the DEA literature on measuring the efficiency of public higher education.

In summary, the mainstream of the thesis is the efficiency measurement of public service organizations under reform and austerity. Our research is expected to provide insights into the efficiency of the public service organizations examined and has certain contributions to the extant literature on performance measurement of the public sector. However, it should be noted that public services are complex, and measuring efficiency is a challenging task. Therefore, what we have found might be just a part of a broad picture, and the policy implications that we have made definitely require a thorough assessment prior to implementation. Also, there still exist avenues for future research such as incorporating the quality aspect, measuring efficiency of component divisions, or testing the robustness of regression models.

TABLE OF CONTENTS

FIGURES

TABLES

ABBREVIATIONS

A&E	Accident and Emergency
BCC	Banker, Charnes, and Cooper
CCG	Clinical Commission Group
CCR	Charnes, Cooper, and Rhodes
CPI	Consumer Price Index
CRS	Constant Return to Scale
DEA	Data Envelopment Analysis
DGP	Data Generating Process
DHBs	District Health Boards
DMUs	Decision Making Units
ED	Emergency Department
EU	European Union
FTE	Full-Time Equivalent
GDP	Gross Domestic Products
HBL	Health Benefit Limited
HRG	Healthcare Resource Group
IT	Information Technology
JUG	Joint-Up Governance
KPI	Key Performance Indicator
LOS	Length of Stay
MEXT	Ministry of Education, Culture, Sports, Science, and Technology
ML	Maximum Likelihood
NHS	National Health Service
NPfIT	National Programme for Information Technology
NPG	New Public Governance
NPM	New Public Management

NWS	Neo-Weberian State
OECD	Organization for Economic Co-operation and Development
OLS	Ordinary Least Square
PBFF	Population-Based Funding Formula
PCTs	Primary Care Trusts
PFM	Public Financial Management
PHOs	Primary Healthcare Organizations
PLICS	Patient Level Costing System
SBA	Slacked-based Model
SFA	Stochastic Frontier Analysis
SLR	Service Line Report
UK	United Kingdom
US	United States
VRS	Variable Return to Scale
WoG	Whole-of-Governance
WTE	Whole Time Equivalent

CHAPTER 1: INTRODUCTION

The two first sections of this chapter outline the background and context of this research. The next section specifies the motivation and research purposes. The final section outlines this thesis.

1.1 Background

Public services are essential such that most people, if not all, from the cradle to grave, use them at least once, and citizens may judge governments through their failure or success in providing public services (see Pollitt, 2010; Schick, 2013). In addition, not only the recipients of public services but also other stakeholders are involved and have concerns, such as taxpayers, employees, managers, and politicians (Latham and Prowle, 2012). The key role of public services may also come from its central part of any government, in terms of both public spending and staff. Indeed, governments are often the key providers for public goods and services such as social care, healthcare, and education. For instance, in the Organization for Economic Cooperation and Development (OECD) countries, 70% of final-consumption expenditures on healthcare and 84% on education are covered by the government (Lau *et al.*, 2017). Therefore, since they are amongst the largest budget spenders, the performance of public services organizations in areas such as healthcare and education has a considerable impact on the performance of the entire public sector.

The participation and intervention of the government in providing public services are on the grounds of overcoming the defects of free markets (e.g. services without profitably might not be provided). However, there is no guarantee that the government will outperform the private sector and use the resources more efficiently in the absence of market forces (e.g. supply and demand) coupled with nonmarket complications (e.g. political institutions; motivation) (Andrews and Entwistle, 2014). Therefore, given that public service efficiency is a vital issue in public management, public service

organizations have often been perceived as inefficient and unresponsive in meeting policymakers' demands and citizens' needs. Unlike the private sector where efficiency can be evaluated through profitability, market share, customers' satisfaction, or bankruptcy, no such indicator is available in the public sector. Therefore, measuring the public service efficiency is an integral component of public management (Andrews and Entwistle, 2014), which can help public service organizations better manage and provide services and continuously improve the performance and increase public support (Pidd, 2012).

Although performance measurement of public service organizations is certainly not a new notion, it has increasingly become more intensive since the emergence of the New Public Management (NPM) in the 1980's (J. M. Lewis, 2015). Indeed, "NPM focused on efficiency as a central objective, claiming that there were major efficiency problems in public organizations, which has been disputed and seen as an ideological argument" (Christensen and Laegreid, 2011, p.14). The underlying nature of NPM is a market-based reform advocating the implementation of market principles such as competition and more choices in public service provision. Also, it is worth noting that unlike the normal free market, the term quasi-market is often used in the public sector, where users normally do not directly buy goods or services but, rather, the government pays for these based on the users' choices. More particularly, public services can be provided by diversified providers in both the public and private sectors, allowing users to have more options to choose alternative providers. When this choice is combined with competition, it will provide incentives for higher efficiency, quality, and responsiveness (Le Grand, 2007). For example, patients can choose a hospital for treatment and the type of medical treatment, and parents (or students) will be allowed to choose any school/university and the curriculum to study. Thus, the hospital might lose its clients or school/university can lose pupils/students to other competitors, which might threaten its viability. As a result, the

providers (both knaves and knights) are encouraged to continuously improve their performance to benefit both providers and users (see Le Grand, 2007).

As a component of public sector management reforms, restructuring public service organizations is a common solution to streamline, renovate the public sector bodies, and incentivize the participation of the private sector, gearing at motivating the competition and improving efficiency. Embedded with restructuring, the pressure on public services efficiency has been intensified due to the financial crises of 2008, social-demographic change, climate change, and technological change (Griffiths and Kippin, 2017; Pollitt, 2016).

1.2 Context of the research

Improving public service efficiency is an ongoing concern of the government, citizens, and service providers. Therefore, in this study, we focus on investigating the technical efficiency of public service organizations under the influence of public management reforms at times of tightening budget and demographic changes. Accordingly, the context of the study is confined to typical public services in the countries that experienced those striking features, including the hospital services in the United Kingdom (England, more specifically), healthcare service in New Zealand, and the public higher education in Japan in the period after the global financial crisis in 2008.

Like many European countries, public spending on the healthcare sector in the United Kingdom (UK) has decreased markedly after 2009 (Ongaroet *et al.*, 2015). Also, in the UK, the population has increased recently due to the growing numbers of immigrants and birth rates; however, it seems that the benefits obtained from the population growth might not compensate for the increase in corresponding demands (Latham and Prowle, 2012). Adopting pro-cyclical approaches[1], the UK's Coalition Government, formed in 2010, set a

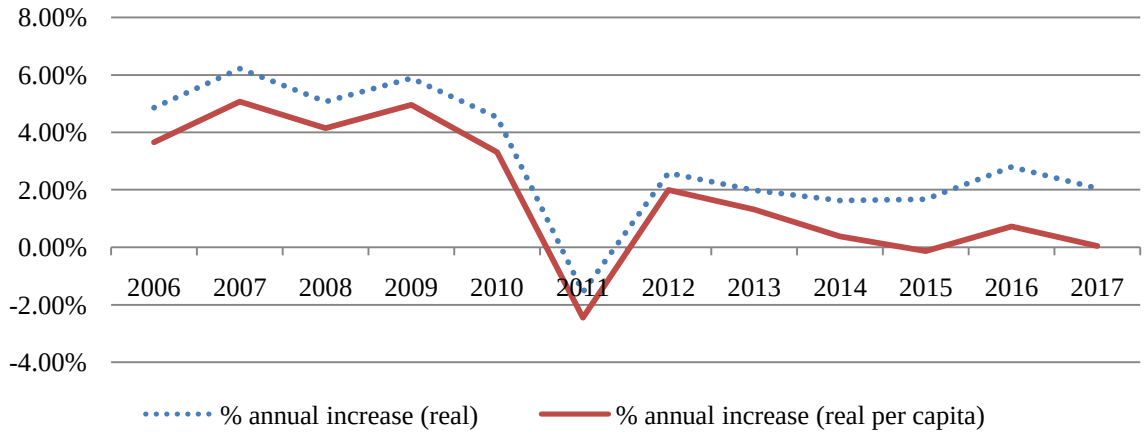
[1] In response to economic recession, a government might adopt counter-cyclical approach (increase spending) or pro-cyclical approach (reduce spending).

goal to eradicate the budget deficit in four years through a package wherein increasing tax and reducing public expenditure. More particularly, in response to the funding reduction and growing demands, NHS England was required to make unprecedented savings of £20 billion in four years from 2011 to 2014, which mainly relied on the efficiency improvement of hospitals - the biggest spenders. Although the Coalition Government assured that funding for healthcare would increase in the real term, it could not meet the rising demand due to technological and demographic change (see Hurst and Williams, 2012). Indeed, we found that while the average annual rate of healthcare expenditure per person between 2004 and 2009 had grown at 6.6%, this rate had reduced to just 2.4% from 2010 to 2015. The Coalition Government proposed and undertook the reforms amid the fiscal austerity, especially the structural change following the introduction of the Health and Social Care Act 2012. The approaches that lie behind the reforms were the use of targets and performance management, inspection and regulation, and competition and choice, which are expected to further motivate the performance of NHS England. However, it is argued that the main approaches to deliver efficiency, such as freezing pay (limiting the increases of staff salary), reducing hospital tariffs, and cutting back-office costs, have reached their limit, and it has become more difficult to achieve efficiency. Also, the reforms might exacerbate the pressure on NHS England; moreover, there was little evidence on the effectiveness of the approaches applied in the reforms (Applyby *et al.*, 2015; Ham *et al.*, 2015). Therefore, since hospitals play a vital role in delivering efficiency targets, it is worth examining how the efficiency of English hospitals changed over the period of reforming and budget austerity.

From the perspective of health policy and healthcare system, New Zealand can be compared to England (e.g. structural similarities - universal health system, public hospitals are key providers). Similar to England, improving efficiency became a central issue of New Zealand in the post–2008 agenda when the National government found noticeable

overlaps in management and transaction costs within the 20 district health boards (Gauld, 2016). Given that this was not as massive as those in England, reforms were also undertaken when the New Zealand National government came into power in 2008. For example, the mergers of primary health organizations (PHOs), the requirements of collaboration among DHBs following amendment to the Public and Health Service Act in 2010, and the alliance between DHBs and PHOs (Gauld, 2016). Like elsewhere, following the financial crisis, New Zealand's healthcare sector has experienced financial squeeze. Health spending increased in the real term but at a much lower growth rate when compared with prior 2009 (see Figure 1). Moreover, it is argued that the New Zealand healthcare sector is underfunded and the government funding did not keep up with the rising costs and growing demand due to population growth and aging (Keene *et al.*, 2016; Rosenberg and Keene, 2016; The Treasury, 2017a). Indeed, it is estimated that when compared with 2009/10, accumulated funding shortfall in government health expenditure for 2016/17 financial years was $1.20 billion (Rosenberg and Keene, 2016).

Figure 1. Vote Health[2] (operating) expenditure growth



Source: The Treasury (2017); Real expenditure means was adjusted for CPI inflation; per capita numbers are based on a simple count of total population without adjustment for demographic factors, such as aging, that could tend to increase costs *per capita*.

[2]Vote Health is the primary source of funding for New Zealand's health and disability system, which is administered by the Ministry of Health; about three-quarters of Vote Health goes to fund the 20 District Health Boards.

Although NHS England and New Zealand have similar contexts, the trajectories are likely to be different, especially the policy tools, underlying philosophy, and resulting structures (Gauld, 2016). In addition, while English hospitals are key factors in delivering efficiency targets, 20 New Zealand DHBs consume three-quarters of the total government funding for healthcare. Moreover, unlike English hospitals, DHBs not only provide hospital services but are also responsible for funding the provision of health services in their district (e.g. primary care). Therefore, New Zealand DHBs may provide different aspects in measuring the efficiency of the healthcare services.

Similar to the healthcare sector, higher education in many countries have experienced reforms following the wake of NPM (Christensen, 2011a). Accordingly, efficiency has become a growing concern for politicians, teachers, and other education stakeholders (De Witte and López-Torres, 2017) since universities are also responsible for utilizing scarce resources efficiently like any other public organization. It is suggested that universities should be given more autonomy and adopt modern management principles to improve efficiency (Paradeise *et al.*, 2009a cited by Christensen, 2011). Although it occurs later than the university reform in other countries, the incorporation of Japanese national universities in 2004 (to grant universities the corporate status) also has the features of NPM, which is considered the most radical reform in higher education after World War II. The key idea behind this transformation was to provide national universities with more autonomy to renovate education and research, improve efficiency, reduce reliance on government grants, and enhance international competitiveness (see Christensen, 2011; MEXT, 2010). Basically, key changes in the governance of national universities included the abolishment of the public servant system, the allocation of lump-sum budgets rather than line-item budgets, and the adoption of private-type management (MEXT, 2003; Mirozumi, 2019). Thus, improving efficiency was among the important goals of the

incorporation of Japanese national universities. The pressure on achieving efficiency has also been intensified due to the reduction in operating grants. For instance, the first midterm plan (2004 - 2010) set an annual efficiency target of 1% and the funding for national universities was therefore reduced by 1% per year; in the second six-year plan, universities were required to reorganize or else the basic operating grants would be decreased by 1%. Indeed, the accumulated reduction of operating grants from 2004 to 2013 was JPY 162.2 billion. Therefore, national universities struggled with seeking measures (e.g. using part-time staff, reducing academic staff, competing for research funding) to cope with budget restraints (see Mirozumi, 2019; Mizobata and Yoshii, 2015). In addition, higher education in Japan has encountered with decline in the number of students due to the rapid aging population and constant shrinking young population, which also threatens the efficiency and viability of universities. The Japanese national universities following the incorporation with the pressures of fiscal hardship and shrinking student numbers, therefore, provide a good context for investigating efficiency.

1.3 Research motivation

There has been an increasing concern in improving the efficiency of public service organizations. In this regard, English hospitals, New Zealand DHBs, and Japanese national universities under the structural reform, the pressure of budget restraints, and the impact of other environmental factors provide good contexts for studying efficiency. However, it is interesting that there is a scarcity of recent studies on measuring efficiency in the contexts mentioned above; especially those that employ the Data Envelopment Analysis (DEA) approach. For instance, Valdmanis *et al.* (2016) observed little DEA research was conducted on UK hospitals. Moreover, Hurst and Williams (2012, p. 11) argued that "studies of the determinants of efficiency are largely confined to observational studies, with an absence of controlled trials or longitudinal studies that link efficiency to its supposed determinants over time". Indeed, we found that most of the extant DEA

literature on the UK hospitals was conducted before 2006 and a few applied approaches such as second-stage analysis to identify the factors that impact efficiency. Similarly, through an investigation of the efforts in measuring the efficiency of the New Zealand healthcare sector, Knopf (2017, p. 5) concluded, "Attempts to measure efficiency/productivity in the health sector has been tough going. There are data gaps, missing paradigms, and communication issues. The analytical capacity and capability across the sector appear to be in short supply. Meaningful succinct measures to populate performance frameworks have been elusive". Regarding the Japanese public universities, little is known about the technical efficiency of national universities despite a large corpus of DEA literature on universities in many other developed countries (see Agasisti and Wolszczak-Derlacz, 2016; Lee, 2011; Thanassoulis *et al.,* 2011). Therefore, the lack of studies on the efficiency of these organizations has motivated this research.

The contexts and objectives of the research are also expected to complement the extant literature on efficiency measurement of public service organizations in several ways. While English hospitals have been well examined previously, this study clearly confines to acute foundation trusts, which make the comparison more comparable. Regarding methodology, while some extant literature uses crude hospitals' outputs (e.g. number of inpatients, outpatient visits), this study carefully adjusts hospital activities (case-mix) to capture the complexities and diversities of services among hospital and, thus, hospital outputs are more accurately calculated; the reliability of efficiency scores estimated might be improved. In the context of New Zealand, as discussed previously, measuring the efficiency of public service as well as the healthcare sector has been elusive. This research introduces an additional approach to comprehensively evaluate the efficiency of healthcare organizations and complement the current methods used to measure the performance of the public sector. Likewise, the research on measuring the technical efficiency of Japanese

national universities with striking features (e.g. reforms, declining population), may provide useful aspects for studies in efficiency measurement of public higher education.

As the concerns of any government, policymakers, managers, and stakeholders are not restricted to the benchmarking, how to improve efficiency is very likely to be more of their interest. Therefore, this study proposes possible policies to better improve the efficiency of organizations under evaluation. More particularly, apart from the environmental factors that might be out of control, this study also focuses on the controllable factors that have an influence on efficiency, such as patients' characteristics, asset utilization, scale, and scope. From the policy perspective, we also make some recommendations for better efficiency improvement.

1.4 Purposes

The key goal of this research is to measure the technical efficiency of public service organizations in the healthcare and education sector. It specifically aims to identify both the efficient and inefficient units and determinants of efficiency. In accordance with each specified context of the study, the research seeks to explore the three following key questions:

First, did the expected improvement to technical efficiency, indicated by reforms and fiscal distress, indeed occur over the examined period?

Second, are there factors that explain any variability in efficiency across the public service organizations, especially the impacts of the internal factors?

Third, what are the policy implications to improve technical efficiency?

1.5 Thesis outline

This thesis comprises 7 chapters.

The first chapter provides an overview of the background information on the necessity of performance measurement in the public sector (efficiency measurement, more particularly), briefly describes the context of the study, and clarifies the motivation as well as objectives of the study.

The second chapter reviews the key movements in the public management reform since the 1980s, including New Public Management, Neo-Weberian state, and New Public Governance. Also, reforms in the context of the study (the UK healthcare, New Zealand healthcare, and Japanese national universities) are more elaborated. In addition, we also summarize the key aspects of financial management reform, including performance budgeting, the modernization accounting system, and the responses to the financial crisis. While there have been different approaches to the reform, improving the performance of the public sector is considered as the ultimate goal, and the performance measurement has been a growing concern. Therefore, this chapter also provides a literature review closely related to performance measurement and performance management in the public sector (e.g. merits and dysfunctions). The last section of this chapter primarily focuses on the approaches applied to measure efficiency, an aspect of performance measurement, in the public sector. Along with quantitative analysis used in measuring efficiency, we also introduce qualitative content analysis to complement the findings and policy implications.

Chapter 3 focuses on the issues related to performance measurement in the public sector. Accordingly, the study sets the conceptual framework on the performance paradox and investigates its manifestations. We conduct semi-structured interviews to get information on the implementation of performance measurement in the New Zealand DHBs, Australian hospitals, and Vietnamese hospitals. The recorded interviews are then transcribed and the transcripts used for content analysis. Similar to the factors that are well examined in the extant literature (intended errors, unintended errors, and synecdoche), this study explores the existence of performance paradox in the given context. It also sets up

the argument that efficiency scores, which can be estimated from the DEA technique, can considerably complement the current performance measurement framework.

Chapter 4 presents empirical research on measuring the technical efficiency of English acute foundation trusts from 2009 to 2016. Basically, this chapter employs a two-stage DEA technique to both measure efficiency and identify the determinants of efficiency. Accordingly, we calculate the technical efficiency score of each hospital as well as the entire hospitals for each year to examine the changes in efficiency scores. We also attempt to link the efficiency trends with the factors, events, or policies that may cause changes in the efficiency of the whole hospitals examined. Moreover, we aim at identifying the internal factors (e.g. old patients, asset utilization) that significantly affect efficiency and may provide meaningful policy implications for hospital managers to improve efficiency.

Similar to Chapter 4, Chapter 5 investigates the technical efficiency of New Zealand DHBs from 2013 to 2016. Data envelopment analysis is also employed as the technique to estimate the relative technical efficiency of each DHB in addition to monitoring the trend in efficiency over the period of analysis. Although Chapter 5 is also about measuring the efficiency of the healthcare sector, it should be noted that, unlike English acute foundation trusts, New Zealand DHBs operate as a combination of local authorities and hospitals. Therefore, the set of variables used should capture not only hospital services but also other activities that aim at improving the health condition of people within the defined regions. In addition to estimate efficiency trends, a second stage bootstrapped truncated regression model is used to explain and identify the determinants of efficiency. In explaining the efficiency trend, we not only link to the public policies or government reports but also analyze the cost structure to identify the main sources of decline or improvement in efficiency. We still put more emphasis on internal factors such as assets utilization and budget management. Finally, public policy recommendations are derived, aiming to improve the efficiency of DHBs.

Chapter 6 presents another empirical research on measuring the efficiency of public service organizations in public higher education, particularly Japanese national universities between 2010 and 2016. Apart from structural reform and budget restraints, we highlight the unique features of Japanese national universities (e.g. shrinking student enrollments, overexpansion of private universities, language barriers), which set an important case study to investigate how technical efficiency may have been affected by these challenges. In addition to using DEA as a benchmarking tool, a second-stage analysis is also used to provide more insights on the determinants of efficiency scores. While there have been a variety of explanatory factors found in the extant literature, we particularly focus on funding related factors (e.g. share of government funding), characteristics of universities (e.g. attached hospitals, scale). In addition, we also conduct semi-structured interviews with senior managers to explore their assessment of the achievements of the incorporation of national universities and the performance following structural changes. Based on the interviews, the content analysis approach is employed to augment the findings and fortify the policy implications.

The final chapter summarizes the key findings from the study in each context, concerning the efficiency trends and the impacts of explanatory factors. Also, policy recommendations to enhance efficiency are more emphasized, especially those under the control of the managers or policymakers. Besides, we also clearly specify the limitations of the research and suggest scope for future studies.

CHAPTER 2: LITERATURE REVIEW

In the Introduction, the public services organizations under analysis were characterized by a common feature of reforming, coupled with the pressure of financial distress and social-economic impacts. Therefore, this chapter delineates the reforms in public management, especially the emergence and evolvement of the New Public Management. In addition, as reforms ultimately aim to improve the performance of the public sector, it is necessary to discuss how performance measurement and performance management have become one of the central concerns. More particularly, to measure efficiency, an aspect of performance measurement, this chapter also describes approaches in measuring the technical efficiency of public services organizations.

2.1 Models of public management reform

Reform means "change in a direction advocated by some groups or individuals. It does not necessarily mean improvement" (Rubin, 1992, p. 20 as cited in Pollitt and Bouckaert, 2017); Public management reform is defined as the "deliberate attempts to change the structures, processes, and/or cultures of public sector organizations with the objective of getting them (in some sense) to run better." (Pollitt and Bouckaert, 2017, p.2).

Traditionally, public administration is characterized by hierarchy top-down system with standardized procedures stipulated by policies and regulations. However, this paradigm was criticized for inefficiency and lack of responsiveness (see Hyndman and Liguori, 2016; Lampropoulou and Oikonomou, 2018). Therefore, since the 1980's, three key approaches have been proposed for public management reforms, namely New Public Management (NPM), Neo-Weberian State (NWS), and New Public Governance (NPG).

2.1.1 New Public Management

New public management is a term first used by Hood (1991) to label the wide range of reforms at Anglophone countries (Australia, New Zealand, the UK, and the USA) during

the 1980's and 1990's, when principles of management in the private sector were advocated to be applied in public administration context. This movement then received huge support from international organizations such as OECD and World Bank and spread to industrialized and developing countries in Asia and Africa (Pollitt and Bouckaert, 2017); it is also considered as the most notably popular movement in public management reforms.

In the seminal work, Hood (1991) described seven elements to stylize NPM model as follows:

● 'Hands-on professional management': More decision autonomy for managers to control the operations of organizations. The accountability is therefore achieved through a clearer assignment of responsibility.

● 'Explicit standards and measures of performance': Goals, targets, and indicators of success are clearly defined and normally expressed in quantitative terms, which help to evaluate the achievement of the objectives set and explains their accountability.

● 'Greater emphasis on output controls': Rather than input controls and bureaucratic procedures, the emphasis is placed on results, which are the basis for funding allocation and rewards;

● 'Shift to disaggregation units': The conventional management systems (monolithic units, unitary structure) are replaced by disaggregated and decentralized units. The purpose is to create manageable units and separate provision and production to improve efficiency through the advantage of contracts or franchises.

● 'Shift to greater competition': Contract and tendering mechanisms are used to inspire the competitions within the public sector and between the public sector and private sector. The underlying reason is that rivalry is expected to reduce costs and improve the quality of service provided.

● ‘Stress on private-sector-style management practices’: More flexibility in human resources management (hiring, rewarding) and greater use of public relations techniques.
● ‘Stress on greater discipline and parsimony in resource use’: Cost-cutting, raising labor discipline, and limiting compliance costs. The aim is to improve efficiency by “doing more with less” – proving more services with fewer resources consumed.

Apart from the initial generalization, other attempts were made to refine the key components or characteristics that distinguish NPM with other reforms (Hyndman and Lapsley, 2016). For instance, Pollitt and Bouckaert (2017, p. 10) classified NPM into two tiers: (i) the “high level” that mainly focuses on the theory or doctrine wherein NPM is the adoption of businesslike management styles and market principles and (ii) the “normal level” that comprises of more specific concepts and practices, including (1) greater emphasis on ‘performance’, especially through the measurement of outputs; (2) a preference for lean, flat, small, specialized (disaggregated) organizational forms over large, multi-functional forms; (3) a widespread substitution of contracts for hierarchical relations as the principal coordinating device; (4) a widespread injection of market-type mechanisms including competitive tendering, public sector league tables, and performance-related pay; and, (5) an emphasis on treating service users as ‘customers’ and on the application of generic quality improvement techniques.

Despite different conceptualizations, in essence, the NPM model is characterized by “minimization” and “marketization” or “market-like incentives”, in which the reformers expect to downsize governments’ functions by facilitating/motivating the participation of private sector in providing public services where possible and encourage implementing the businesslike principles for the services provided by the publicly owned sector to reduce costs and improve their performance (Kettl, 2005; Pollitt and Bouckaert, 2017).

The questions about the viability of the NPM model have been a debatable topic, drawing the attention to academics in the public sector. There exists a school of thought claiming that the era of NPM is over. For instance, in a broad sense, Drechsler and Randma-liiv (2014) argue that NPM has been no longer a dominant theory since 2005. More particularly, Levy (2010, p. 234) claimed that "NPM is arguably as much a casualty of the global economic crisis as are the markets and market mechanisms which underpin it" and proposed four alternative scenarios to NPM that can be applied in the UK context and elsewhere.

In contrast to the arguments that NPM is obsolete, proponents confirm the persistent existence of NPM. Employing the virus theory[3], Hyndman and Lapsley (2016) explained the trajectory of NPM in the UK context through different phases and argued that the key elements of NPM have been continuously augmented with new ideas through the mechanism of adaption and mutation, which are more deeply embedded in the UK public services. Systematically review of literature on a growing trend of post-New Public Management emerging in the early 2000's, Reiter and Klenk (2018) suggested that post-NPM has been very effective in identifying the shortcomings of the NPM model. However, the authors also noted that post-NPM only focuses on some aspects of NPM (decentralization and fragmentation versus re-centralization and reintegration; market-driven versus coordinative/collaborative) and some elements of the NPM are strengthened (e.g. efficiency). Therefore, it is still early to affirm the clear signs of post-NPM period; rather, it is a process of layering that incorporates new features into the old model. Similarly, De Vries and Nemec (2013) asserted that it is far from certain to state that NPM is passé´; NPM key doctrines (e.g. minimum state and privatization) are still employed given that some countries have followed alternative approaches.

[3] The evolvement of a management reform is explained by six corresponding stages that applied in the study of virology: 1. Infectiousness - Adoption; 2. Immunity - Non-adoption; Replication - Entrenchment; 4. Incubation - Maturation; 5. Mutation – Translation; 6. Dormancy - Inactive/Reactive.

2.1.2 New Public Governance

New Public Governance is the concept developed by Osborne (2006), which is classified as a successor of NPM combining the strengths of both traditional public administration (legitimacy and inter-relation) and the NPM (service delivery process). The rise of NPG probably departs from the need to strengthen the cooperation among the providers in delivering public services (Dickinson, 2016). Indeed, NPG was expected to be a new approach for the limitation of NPM such as organizational fragmentation or over-emphasis on the intra-organizational process. Accordingly, on the basis of network theory, NPG emphasizes the "inter-organizational relationships and the governance of processes and stresses on service effectiveness and outcomes" (Osborne, 2006, p. 384).

However, without a universal definition of "governance" *per se* is also problematic; "most descriptions of governance-networks, inter-organizational and inter-jurisdictional cooperation, power-sharing federations, public-private partnerships, and contracting out are forms of institutional adaptation in the face of increasing interdependence" (Frederickson *et al.,* 2012, p. 222). Similarly, as indicated by Hughes (2010, p. 102), "The definitions of governance as networks, as new public management, as socio-cybernetic systems, as the new political economy, to name but a few, do not add to the understanding of governance as a word, although they might have other utility". Therefore, NPG has been criticized for its broad and abstract definition and lacking empirical evidence, which makes it far from an adequate theoretical framework (Pollitt and Bouckaert, 2017). In addition, Klijn and Koppenjan (2012) argued that the rise of NPG might not an alternative candidate for hierarchy and NPM, but rather hybrid practices will emerge.

2.1.3 Neo-Weberian State

While the NPM model was pervasive in the Anglophone countries, Neo-Weberian State (NWS) is considered as an alternative paradigm to NPM, labeling for the public management reforms that occurred in continental and Northern European countries (e.g.

France, Germany) (Byrkjeflot, du Gay, and Greve, 2018). The term Neo-Weberian State itself is the combination of the two features: (i) Weberian - traditional bureaucracy and (ii) Neo - innovative ideas of more professional, efficient, and citizen-friendly, as clearly described by Pollitt and Bouckaert (2017, pp. 121–122) as follows.

Table 1. Elements of Neo-Weberian State

Weberian' elements	'Neo' elements
Reaffirmation of the role of the state as the main facilitator of solutions to the new problems (e.g. globalization, demographic changes).	Shift from internal orientation toward bureaucratic rule-following to an external orientation toward meeting citizens' needs and wishes through the creation of a professional culture of quality and service.
Reaffirmation of the role of representative democracy as the legitimate element within the state apparatus.	Supplementation (not replacement) of the role of representative democracy by a range of devices for consultation and the direct representation of citizens' views
Reaffirmation of the role of administrative law in preserving the basic principles pertaining to the citizen–state relationship	Greater orientation to the achievement of results rather than merely following procedures; shift from ex-ante to ex-post controls (not completely abolish the former).
Preservation of the idea of a public service with a distinctive status, culture and, to some extent, terms and conditions.	Professionalization of the public service; the 'bureaucrats' are not simply experts in the government policies but also professional managers oriented to meeting citizens' needs.

According to Drechsler (2014), the Neo-Weberian model is superior to NPM, much suitable for an innovation-based society (e-governance is compatible with NWS) and for times of crisis. However, Byrkjeflot *et al.* (2018) indicated that while the neo-elements are expected to make the Weberian paradigm work better, the founders of NWS (Pollitt and Bouckaert, 2004) did not clearly clarify whether these new elements are compatible or undermine the structures and functions of Weberian bureaucracy. The authors also questioned the possibility that such a hybrid of traditional style and modern ideas might end up with the worst for both and suggested that it should be better to discuss the magnitudes of Weberian state than to separate between Neo-Weberian states and those that are not.

2.2 Performance measurement and performance management in the public sector

The ultimate goal of ongoing reforms in public sector in the last decades was to improve the performance (Moynihan, 2008). In parallel, performance measurement has been pervasively employed by public management and public policy and has become an integral part of the public management reforms (Dooren *et al.*, 2015). Especially, following the emergence of NPM model with the shift to "results-based" orientation[4], principals (politicians) set the objectives and targets while agencies (managers) are given more autonomy on running public service organizations, the performance of which should be measured and aligned with the objectives and targets set. As a result, in the performance management aspect, the public sector has adopted and adapted the private-type management approaches or developed its own models (Talbot, 2010) to support policymakers, managers in making internal decisions, giving account to the external

[4] Traditionally, public management system emphasized controls on inputs through budget appropriations (e.g. specific line items) and personnel control (centralized civil services rules) or process management (emphasizes on compliance and consistency with rules) by which the managers had little flexibility in managing their resources. Following the rise of NPM, financial and personnel management were decentralized to increase the managerial autonomy (e.g. maintaining unspent budget and recruiting employees) (see Boyne, 2010; Moynihan, 2008)

stakeholders, and improving public performance (Moynihan, 2008). While there exists abundant literature, we have confined this study to the core concepts, merits, and dysfunctions of performance measurement in the public sector.

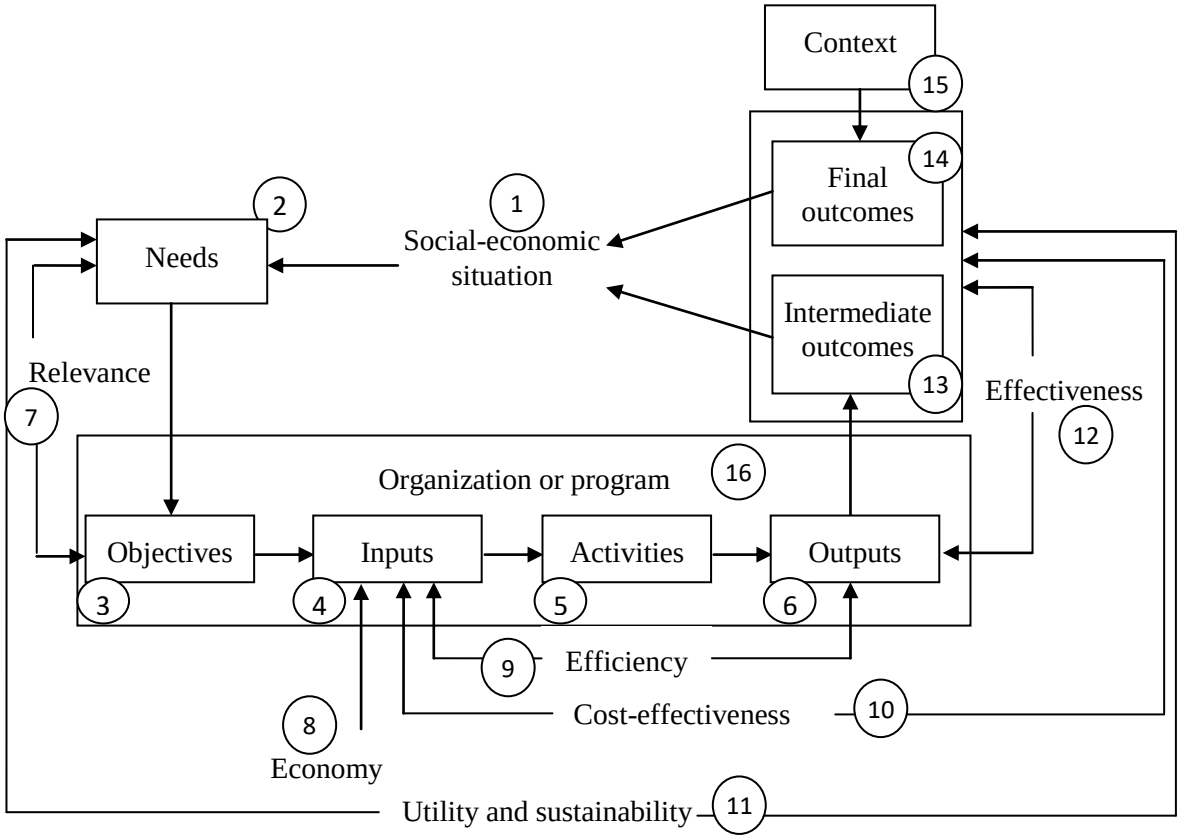
2.2.1 Basic concepts

2.2.1.1 Performance

Although the concepts of performance vary by disciplines, based on the work of Dubnick (2005), Dooren *et al.* (2015) infer that performance is a deliberate action executed by individuals or organizations. Conventionally, on the foundation of the production process, performance in the private sector puts more emphasis on the inputs, activities, and outputs, which inadequately capture the distinctive features of the public sector (e.g. multiple goals[5] and principals). Therefore, performance in the public sector can be understood through four perspectives: (1) performance is conceptualized as production, focusing on actions executed (e.g. medical treatment, teaching a course), regardless of whether they are successful; (2) performance is viewed as competence or capacity - a highly competent performer is likely to produce more output with higher quality; (3) performance can also be equal to the results in which the quality of the results (both outputs and outcomes) is underlined rather the quality of actions; and, (4) when the quality of actions and the quality of results are both emphasized, performance is illustrated as a sustainable result (Dooren *et al.,* 2015). The following figure illustrates the extension of performance grounded in the production model, reflecting the entire chain from inputs to outcomes.

[5]Public service organizations are often required to improve efficiency without compromising equity and quality of the services provided

Figure 2. The production model of performance



Sources: Adapted from Dooren *et al.* (2015)

2.2.1.2 Performance measurement

According to Poister *et al*. (2014, p. 7), performance measurement is defined as "the systematic, orderly collection of quantitative data along with a set of key indicators of organizational (or program) performance". More particularly, performance is delineated as the process of acquiring performance information that includes the following five stages: prioritizing (what to measure), selecting indicators (how to measure), collecting data (including internal and external sources), analyzing (transformation of data into information used for making a decision), and reporting (formats and contents vary by the users) (see Dooren *et al.*, 2015, pp. 39, 63). In a simple approach, Bruijn (2007) suggests the key idea of performance measurement at the public professional organizations (e.g. hospitals, universities, court service) is using the performance indicators to track their envisaged performance (whether achieved or not) and the cost incurred. Thus, in short,

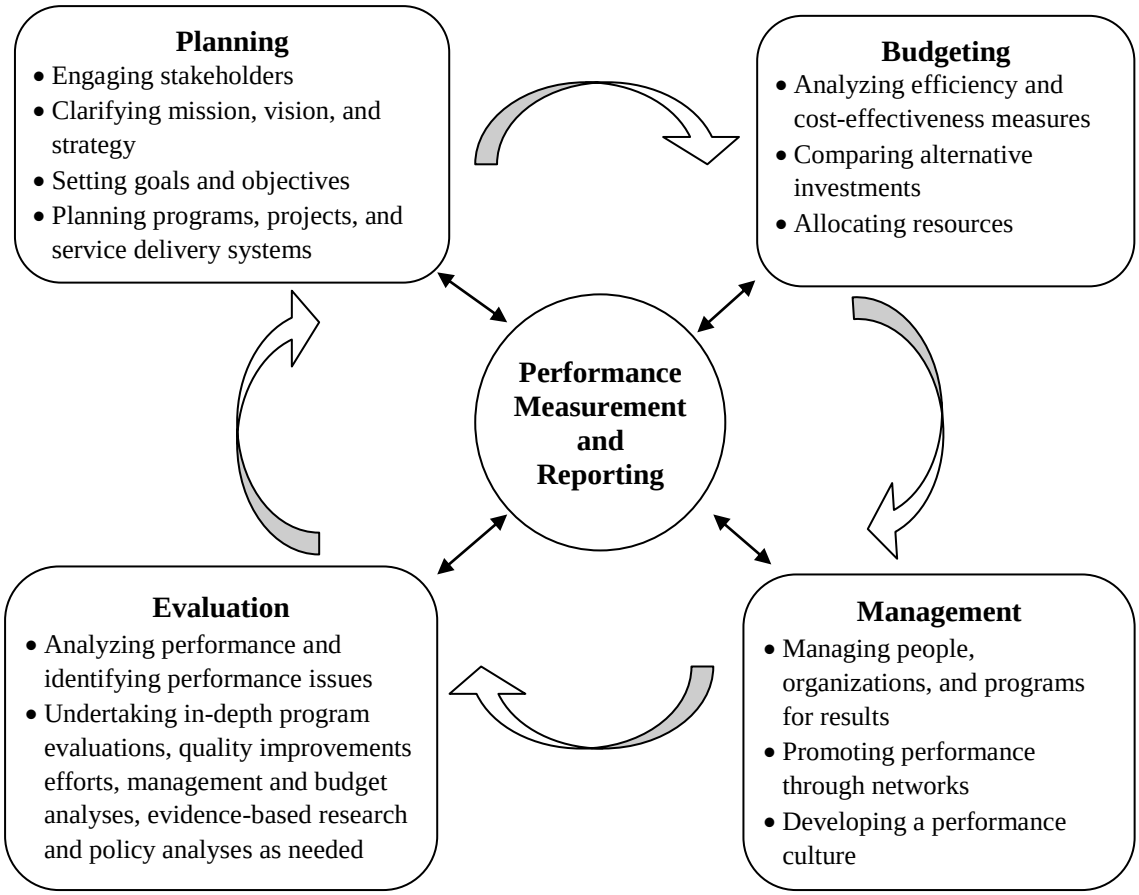
performance measurement refers to the steps and processes required to create performance information, which is assumed to inform for making the decision.

2.2.1.3 Performance management

Performance measurement itself does not guarantee the use of the performance information produced. Therefore, "the incorporation and use of performance information in decision-making" refers to performance management in a broad sense (Dooren *et al.*, 2015, p. 39). Similarly, Moynihan (2008, p. 5) views performance management as "a system that generates performance information through strategic planning and performance measurement routines that connects this information to decision venue, where, ideally, the information influences ranges of possible decision". In another definition, performance management is more elaborately conceptualized as "the collection and purposive use of quantitative performance information to support management decisions that advance the accomplishment of organizational (or program) strategic goals" (Poister *et al.*, 2014, p. 7)

Overall, the above definitions delineate connotations of performance and clearly distinguish the performance measurement and performance management in public sector management. However, it should be noted that given the separation in meaning, they are closely interrelated in an ongoing management system. The information generated from the process of performance measurement is considered the "lifeblood" for managers to make decisions that may require revisions of performance measurement (e.g. adding new scope, updating the indicators, changing the analysis method). In addition, the performance data should not be isolated but integrated with other management processes, including human resources, budgeting, and general management, which make performance management different from simple performance measurement (see Poister *et al.*, 2014; Moynihan, 2008), as illustrated in the following figure.

Figure 3. Performance management framework



Sources: Adapted from Poister *et al.* (2014)

2.2.2 Merits and dysfunctions

2.2.2.1 The benefits

To identify the benefits, we alternatively find the rationales and expectations for adopting performance management and performance measurement in the public sector. The idea of incentives for performance improvement behind the NPM assumes central agencies devolve more operational authority to lower units and focus on controlling the outputs and outcomes; as a result, the managers have more autonomy in utilizing their resources to achieve and account for the clear goals and objectives. Therefore, basically, the performance management doctrine is grounded in the key premise that the use of performance information will better support decision-makers in improving on efficiency

(allocative efficiency and technical efficiency) and be accountable to the public, policy-makers, and other stakeholders (see Poister, 2003; Moynihan, 2008). For instance, performance information can help to explain to the citizens/voters how well public organizations are performing; in return, the public has a greater opportunity to participate in the process of setting goals and evaluating performance. Also, it provides transparent and necessary information for managers and policy-makers to rectify the deficiencies and modify their actions through learning forums (especially double-loop learning), which facilitate a top-down pressure to reform and improve efficiency.

The usefulness of performance management and performance measurement can be observed through its specific purposes. Although there are more than 40 potential uses of performance information, by setting three questions, Dooren *et al.* (2015) categorized them into three corresponding purposes: to learn (how to improve the performance), to steer and control (whether performance aligns with targets), and to provide account (how to explain the performance). From the managerial perspective, in answering the question of why public managers measure performance, Behn (2003) listed eight significant roles of performance measurement, out of which the ultimate goal is to improve performance (see Table 2). In a general approach, its merits can be stratified into two levels: while the common level revolves to 3Es (economy, efficiency, effectiveness), maintaining equity, and improving the quality of services, the deeper level includes accountability to the public, justifying increased resource, enhancing competitiveness (e.g. to attract residents and businesses) (see Poister *et al.*, 2014).

Table 2. Purposes that public managers have for performance measurement

Evaluate	How well is my public agency performing?
Control	How can I ensure my subordinates do the right thing?

Budget	On what programs, people, or projects should my agency spend the public's money?
Motivate	How can I motivate line staff, middle managers, non-profit and for-profit collaborators, stakeholders, and citizens to do the things necessary to improve performance?
Promote	How can I convince political superiors, legislators, stakeholders, journalists, and citizens that my agency is doing a good job?
Celebrate	What accomplishments are worthy of the important organizational ritual of celebrating success?
Learn	Why is what working or not working?
Improve	What exactly should who do differently to improve performance?

Sources: Adapted from Behn (2003, p. 588)

Thus, the possible benefits suggest the indispensable role of performance management and performance measurement. This can also be partly explained by their proliferation and expansion in the public sector in recent decades where the refrain, "If you can't measure it, you can't manage it," has become popular.

2.2.2.2 Undesirable effects

Although the beneficial effects are recognizable and irrefutable, it seems that only if the quality of performance measurement is assured and performance information is properly used can those merits be realized. Notably, the act of measurement is mainly executed by humans whose behaviors change when being observed, making it difficult to measure. In addition, based on the argument that humans are rational utility maximizers, in response performance imperatives, public sector staff might positively react and/or play the system

for maximum advantage (Talbot, 2010). As suggested by Bevan and Hood (2006), the governance by targets can turn "knights" to "knaves" such as reactive gamers and rational maniacs. Therefore, on these conditions, performance management and performance measurement are inevitably free from problematic issues.

Indeed, the dysfunctional effects have been well examined in the literature, manifesting in different guises. In the most cited paper, Smith (1995) found eight unintended effects of publishing performance data in the UK public sector as follows:

Table 3. Unintended effects

Unintended consequences	Description
Tunnel vision	Managers only emphasize what is being measured (quantified aspects) but neglect unquantifiable objectives. Although, it is impossible to measure all activities, failure to capture important features could not provide sufficient information as expected.
Sub-optimization	The pursuit of narrow objectives at the expense of the organizations' overall objectives.
Myopia	To achieve the performance indicators, the managers may mainly focus on short-term targets than long-term objectives.
Measure fixation	The emphasis is on the measure of success rather than the underlying objective. This happens when a measure cannot capture all dimensions of the objective; managers then just pay attention to the reported measures than the objective.

Misrepresentation	The deliberate manipulation of performance data making reported performance different from actual performance, which could be manifested in creative reporting (e.g. accounting choices) or fraud.
Misinterpretation	The possibility to misinterpret the performance information due to the complexity of the public sector and the lack of objective, independent, and professional analysis.
Gaming	The deliberate action to secure strategic advantage (e.g. ratchet effect) in which the managers have lesser incentives to improve the performance since it will affect the targets in the coming years.
Ossification	An organization is paralyzed due to the intensive use of the performance measurement system.

Following the work of Smith (1995), other academic scholars have found analogous evidence of problems in performance regime. VanThiel and Leeuw (2002) investigated in-depth one of the unintended consequences, performance paradox, which refers to the loose coupling between performance indicators and the performance itself. In other words, the performance indicators did not accurately reflect the actual performance. The performance paradox could be unintended or intended, and the reasons for these two types are different. In particular, the former could be the result of minimal accountability requirements, the ambiguity of objectives, and the intensive use of performance indicators. On the contrary, the latter is caused by manipulating (misrepresentation and misinterpretation), restricting the performance report on the most efficient scopes, and "cherry picking"[6]. In examining the system of governance by targets applied in English public healthcare systems, Bevan

[6] Cherry picking refers to the tendency to provide services that cost less or improve the performance of public organizations (e.g. hospitals might exclude patients with severe morbidity).

and Hood (2006) found problems related to synecdoche (take a part for the whole) and gaming. A typical example was that in response to the 4-hour target at the Accident and Emergency Department, patients were required to wait in queues outside until the staff question to ensure that patients could be seen within four hours. Similarly, in response to the "trolley-wait" target that required hospitals to admit patients to hospital beds within 12 hours of emergency admission, the patients were put in the hallways. Moreover, the evidence of gaming is not only restricted to the healthcare sector but can also be found in other public services (see Hood, 2006; Bird *et al.*, 2005).

In sum, performance measurement and performance management are the key principles of NPM, which claims to have important roles in enabling stakeholders to make informed decisions related to performance improvement of public service organizations. However, attention should also be paid for dysfunctions that might undermine the achievements of desirable effects. Patchy data could not provide useful information until being analyzed (Dooren *et al.*, 2015). Therefore, data analysis is among the steps in the process of performance measurement to transform the data into information used for making the decision. The final section of this chapter will summarize techniques related to measuring the technical efficiency of public service organizations, especially Data Envelopment Analysis, the method for performance assessment and benchmarking against best practice.

2.3 Overview and results of public management reforms

2.3.1 The layering process of models in public management reforms

Public management reform is an unceasing movement in which governments around the world try to achieve greater value for tax money and redefine the relationship with its citizens. Also, given the broad contexts with various versions through the times, the reform in public management share the following six common features (Kettl, 2005):

Productivity: Growing demand and finite resources require the government to "do more with less," providing more services with the same or lower level of resources consumed.

Marketization: By privatization and contracting out to the private sector for service provision, the governments expect the market mechanism will change the behavior of public servants and replace the traditional command-and-control regime.

Service orientation: Citizen are central to service delivery system in which citizens have more choices from alternative providers.

Decentralization: The central government devolves and transfers more the responsibility in providing public services to local governments, which makes the government more responsive to the citizens' preferences.

Policy: Splitting the purchasers (policy-making function) and providers (service provision function) to improve the efficiency in delivering services.

Accountability: The bottom-up and result-driven systems replace top-down and rule-based systems to help the government to deliver what they have promised.

Additionally, although the models proposed are various, it should be noted that there have been many reforms among countries that cannot be exclusively attributed (ascribed) to any of those models. Even in a model, there may be different versions because of the diversity in culture, political, and legal systems. Particularly, while NPM appears to dominate over the 1980's–1990's (especially at Anglophone countries), no dominant model can be found in the subsequent period rather a complex mixture of different concepts such as Neo Weberian State, New Public Governance, Joint-Up Governance (JUG)[7], and Whole-of-Governance (WoG)[8] (Drechsler and Randma-liiv, 2014; Pollitt and Bouckaert, 2017).

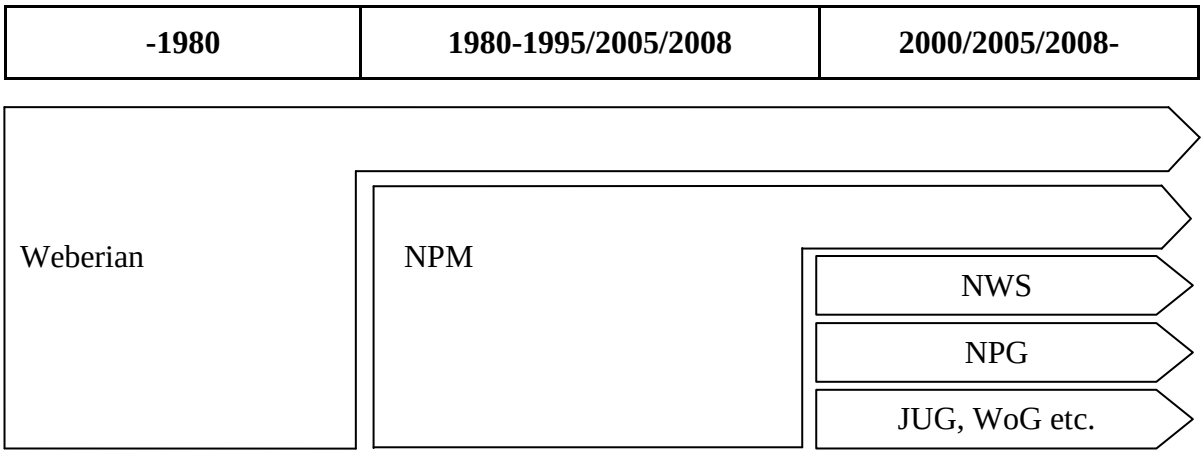
Moreover, there have been evidence of "pendulum swing" patterns when new governments or political leaders take the power and propose new agendas for public

[7] "Joint-up Government" refers to the horizontal and vertical co-ordination among public organizations to remove the contradictions and duplications among policies, programs to improve the efficiency in using resources and proving public services (see Pollitt, 2003).

[8] "Whole-of-Governance" implies the integration and co-ordination across public agencies might enhance efficiency and effectiveness in policy designs and the public service provision (see Christensen and Laegreid, 2011; OECD, 2006).

service reform. For example, in the UK context, the balance between "succession" and "reversals" in policies related to management reform was particularly observed in four areas: "from 'marketization' to performance management in local government", "from reduction of public sector workforce to investment in public service and back to an increasing emphasis on efficiency", "from fragmentation to reintegration at various levels of organizational design", and "from controlling individual organizations to emphasizing co-operation across organizational boundaries" (Wegrich, 2009, p. 149). Similarly, for New Zealand, it is likely that while some original features of NPM have been maintained, some were adjusted or improved, and some have been terminated of even reversed (Grant and Chapman, 2010)

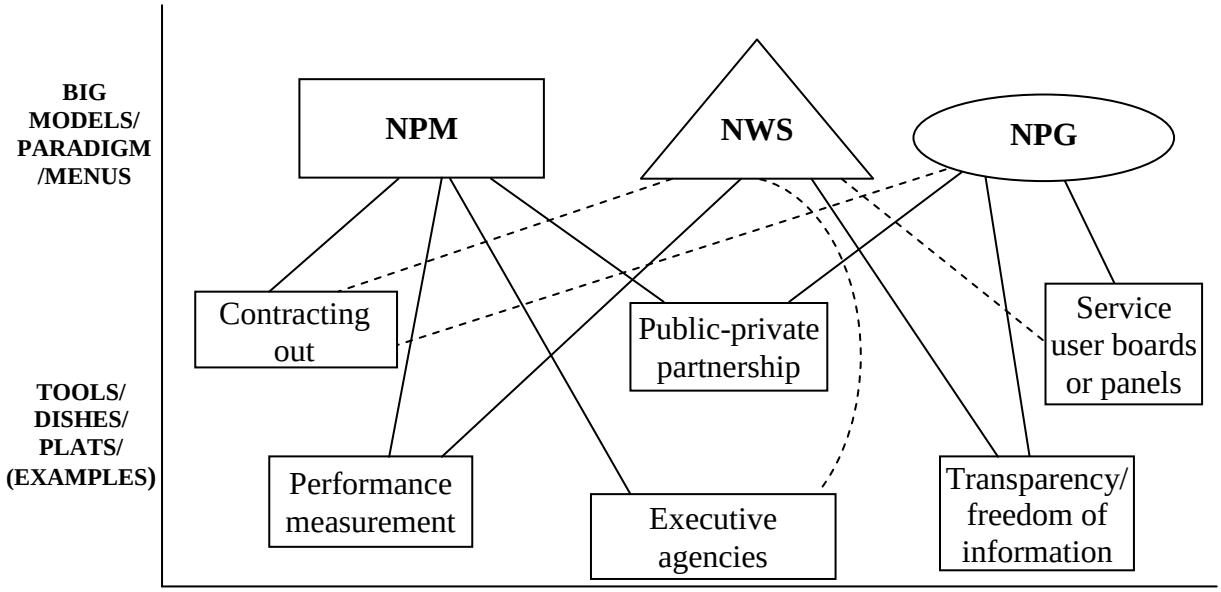
Figure 4. The evolvement of models overtime



Sources: Adapted from Drechsler and Randma-liiv (2014).

Thus, the emergence of new models does not necessarily mean that the successor completely negates the predecessors. In other words, these models are not mutually exclusive but provide additional features. Also, the new models share some common features that make the hybrids more likely to occur (see Figure 5).

Figure 5. Plats and paradigms



Source: Adapted from Pollitt and Bouckaert (2017)

2.3.2 Assessment on results of public management reforms

Since the reforms have globally spread and spanned for decades, a fundamental question arising is how successfully have these movements been or, more basically, what were the results delivered by the reforms.

However, measuring the performance or results of the large-scale reforms is a challenging task and the achievements (e.g. improvement in outputs, outcomes) hardly attribute to reforming programs (Pollitt and Bouckaert, 2017). There are a variety of reasons for this; for example, the problem in identifying the main contributors to the outputs or outcomes achieved, aspects to be measured and the interpretation of reforms depends on different stakeholders' viewpoints, the lack of baseline performance and constant changes in policies applied, and the absence of evaluations to simply avoid critics. In a similar vein, when discussing the difficulties in evaluating the cost savings of management reforms, the reasons ascribed might be because of the inadequacy of baseline for comparison, insufficiency in tracking the inputs and outputs, and mixture between ongoing reforms and political changes (Kettl, 2005).

Given the difficulties mentioned above and although there have been no complete pictures, there exists anecdotal evidence with cross-national comparisons suggesting certain success in many countries (e.g. producing more services with lower resources consumed) (Kettl, 2005). More comprehensively, Pollitt and Bouckaert (2017) reviewed the reforming results from different aspects, including saving money, improving efficiency, increasing effectiveness, and enhancing citizen satisfaction and trust. However, what has been summarized is fragmented and insufficient to make a straightforward assessment; rather, the authors viewed the results of reforms as a "half-empty and half-full wineglass". More particularly, the "empty part" manifests in the absence of adequate information (outcomes, efficiency, and even outputs) and the patchy evaluation of reforms. On the contrary, "the full part" can be seen through the performance data available and the growing concerns of the international community about the public management reforms.

In general, complex realities, dynamic evolvements, and divergence in viewpoints have not allowed scholars, policymakers, and other stakeholders to make a unanimous assessment. It is likely that the extant literature reveals as much as it conceals.

2.4 Public services reform - specific evidence from the United Kingdom, New Zealand, and Japan

Public management reforms involve many sectors with a long history; therefore, we confine to the services (healthcare and education) and key changes closely relevant to the context and time frame examined in the empirical analysis.

2.4.1 Reform in the UK healthcare service

Since the UK is often considered as a typical example of the New Public Management (see Wegrich, 2009), healthcare service was one of the central areas in this movement. Especially, unlike other countries where NPM was partly replaced, the principles of NPM in the UK are embedded and further institutionalized (Dooren *et al.*, 2015). While there

are four healthcare systems in the UK, including England, Scotland, Wales, and Northern Ireland, key reforms in English healthcare are more emphasized since they provide services for over 80% of the UK's population.

As a part of the quasi-market, the "purchaser-provider split" introduced in 1990 was expected to stimulate the competition among hospitals for funding and rewards for efficiency and quality (R. Lewis *et al.*, 2009). The introduction of the "Free choice" policy in 2008 or Any Qualified Provider scheme in 2012 aimed the patients to choose hospitals from a list of qualified providers provided it met the NHS service quality requirements, prices, and normal contractual obligations. As a result, patients had the power to make decisions and choices to meet their needs, which is expected to use public funds more efficiently through the competition among the providers. Also, with the same price list (tariff), the hospital that patients chose to treat will get paid more when providing more services (see Department of Health, 2011; Timmins, 2018; Vizard and Obolenskaya, 2015). Just after the Coalition came to power in 2010, the White Paper - Equity and excellence: Liberating the NHS - was published and which outlined approaches for health reforms, such as extending patient choice and competition, improving health outcomes, empowering professionals and providers with more autonomy and making them more accountable for the results, and increasing efficiency (Department of Health, 2010a). The apogee of reform in healthcare sector under the Coalition government was the enactment of Health and Social Care Act 2012 (came into effect on April 1st, 2013), which is viewed as the most broad and controversial reform of the NHS since its foundation in 1948 (T. Powell, 2016). The basic ideas behind the new Act were the reinforcement of the quasi-market (strengthening commission of services; freeing up providers to innovate; enhancing competition), the enhancement of using external incentives (e.g. targets and performance management, inspection and regulation, and competition and choice), and the promotion of transformation of public hospitals

(trusts) into business-like organizations (foundation trusts) (see Department of Health, 2012; Ham, 2014; M. Powell and Exworthy, 2016).

Generally, the main principle of healthcare reform in England was decentralization (e.g. since 2004, English hospitals have been encouraged to transform into foundation trusts with more autonomy) and reinforcement of the internal market to motivate the cooperation with the private sector and internal competition, coupled with performance standards, which are expected to improve health care quality (Cylus *et al.,* 2015). In a similar view, Ham (2014) indicated that performance management (e.g. targets, service standards), overseeing (inspection and regulation), and competition have been the three main approaches to NHS reform. Thus, the reform of NHS England has mainly focused on system restructuring and adopted the tenets of NPM (M. Powell and Exworthy, 2016).

2.4.2 Reform in New Zealand healthcare sector

Similar to the UK, New Zealand was a renowned leading country of New Public Management reform. In the same vein, the New Zealand healthcare sector has experienced a series of reforms since the 1990's. The period from 1991 to 1996 featured by separation of purchasing and provision of health services to motivate competition among providers; the principle of reform in the following period (1996 - 1999) supported the idea of cooperation rather competition; from 2000 onwards, the reform was characterized with the devolution of funding and planning to local governments (see Gauld, 2016; World Health Organization, 2014).

Thus, currently, the New Zealand healthcare sector is characterized by slightly different features. For example, while the UK still adopts a formal purchaser-provider split principle, the introduction of the New Zealand Public Health and Disability Act in 2000 has abolished this regime. Accordingly, 20 District Health Boards were established in 2001 to not only provide services (through their hospitals) but also act as funders and cover primary care, which was expected to be more responsive to the people's demand through local planning and needs assessment. However, "the principle of keeping policy

and operational functions, and funder and provider, separate was preserved since the purchasing of services, and in-house service delivery (public hospitals) remained separate, 'ring-fenced' budget items" (Grant and Chapman, 2010, p. 305). Also, the Ministry of Health also maintains overseeing its responsibility through strategic plans, targets, or even intervenes on the operation of DHBs wherever applicable. The recent reforms, especially following the formation of the coalition government in 2008, have been evolutionary rather than revolutionary and put more emphasis on the performance improvement of service providers than structural changes. For instance, in 2009, the National Health Board was established as a unit under the Ministry of Health rather than an independent unit to mitigate possible legislative and structural adjustments (Ashton and Tenbensel, 2012).

Thus, there have been pendulum swings in New Zealand healthcare reforms, and the reform in 2000 appears to have shifted away quasi-market and businesslike approach (Starke, 2010). In the case of New Zealand public sector generally, while Grant and Chapman (2010) argued that some of the NPM principles still exist, some have been modified and even abandoned or reversed, Lodge and Gill (2011) found little evidence of a shift from NPM to post-NPM and suggested that the diversified changes have been *ad hoc* and politically motivated.

2.4.3 Reform in Japanese national universities

Unlike the UK or New Zealand, the notion of New Public Management was first mentioned in Japan in 2001, when Prime Minister Junichiro Koizumi and his cabinet proposed NPM as an approach to promote reform in the public sector. Accordingly, they advocated the implementation of key principles of NPM such as free competition, result-oriented evaluation, and separation of policy and execution (Yamamoto, 2009).

As a part of larger public sector reform package, although driven by political and economic factors, the scheme in incorporating national universities in 2004 was basically characterized by NPM principles (see Chan-Tiberghien, 2010; Kaneko, 2012), which can found in the directions of the reform (MEXT, 2003):

- Incorporation along with deregulation (e.g. budget, personnel affairs) gear at providing more autonomous management and generating a competitive environment for national universities.

- The introduction of business-like management styles.

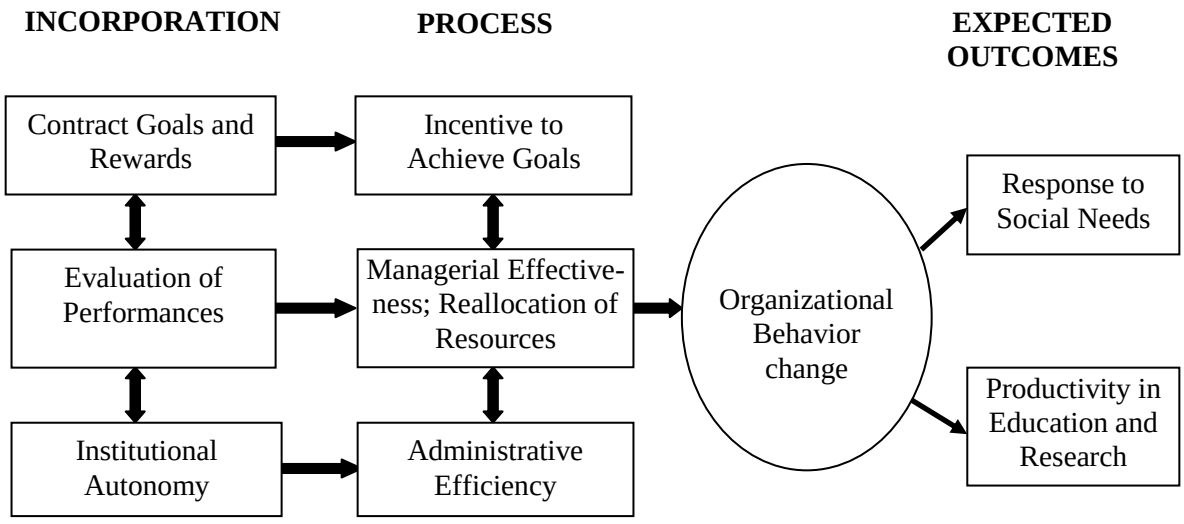
- The institutionalization of the management system, including the participation of external representatives in the board of directors and deliberations on management policy.

- The public servant system will be replaced by a flexible personnel system that is not bounded by the National Public Service Law.

- Application of ex-post check system based on performance measurement to allocate resources and improve transparency.

Thus, the underlying ideas of this transformation were to "introduce market principles in university governance, fund-raising, academic labor management, performance evaluation, and university-industry cooperation in order to make Japanese universities globally competitive on the one hand and locally responsive to rapidly changing social and economic needs on the other" (Chan-Tiberghien, 2010, p. 44).

Figure 6. Design and expected consequences of incorporation



Source: Adapted from Kaneko (2012)

However, it should be noted that despite sharing some common characteristics with the international reforms of universities, restructuring Japanese national universities have certain peculiarities. For example, while the idea of devolution is adopted, the reluctance of MEXT in abandoning its authorities (e.g. approval of plans) makes the Japanese higher education system more centralized than European ones. Therefore, the expected autonomy of Japanese national universities may be limited due to higher control and oversight (Christensen, 2011a).

Thus, although there exist certain variations among those three countries in the approaches and times to reform, the core principles of NPM can be found as a common feature, and improving the performance of the public service providers has been an ongoing theme.

2.5 Financial management reforms in the public sector

In broader terms, "Public financial management (PFM) is concerned with the laws, organizations, systems, and procedures available to governments wanting to secure and use resources effectively, efficiently, and transparently. While PFM encompasses taxes and other government revenue, borrowing, and debt management, its main focus is expenditure management, especially in the context of public budgeting" (North, 1991 cited by Allen *et al.*, 2013, p. 2). Finance is one of the four main components of reforms in public management, along with personnel, organization, and performance measurement (Pollitt and Bouckaert, 2017). The innovations and reforms in PFM are diversified, such as fiscal rules, medium-term budget frameworks, risk management, and performance budgeting, which aim at promoting public management and improving operational efficiency (Schick, 2013). However, this section is only concerned with the key aspects of financial management reform, including performance budgeting, the modernization accounting system, and the responses to the financial crisis.

2.5.1 Reforms in budgeting

Budgeting plays a vital role in PFM since it covers most of the financial decisions. Also, unlike the private sector, budget is not only a tool of planning (medium, long-term), controlling, and measuring performance but also a legal approval of any expenditures in the short-term (usually from legislative levels) (Bergmann, 2009).

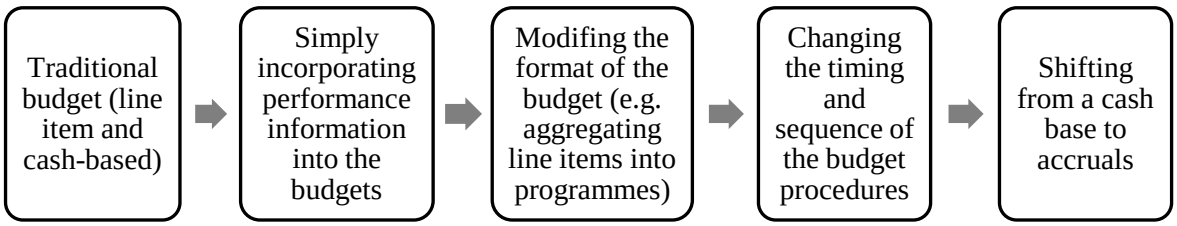
Traditionally, in preparing budgets, expenses and incomes are classified by characteristics that provide little information on the expected objectives or efficiency of services provision (Bergmann, 2009). In addition, under the pressure to contain the growth in public expenditure and the requirement to change the budget to improve public performance, budgeting has been integrated with other management processes such as planning, operational management, and performance measurement (Pollitt and Bouckaert, 2017). Therefore, as a significant aspect of NPM, performance budgeting was developed to capture more detailed information on performance in budget preparation and decision making, aiming at performance improvement. More particularly, the amount of budget being allocated is linked with the level of achievement in outputs or outcomes, implying public agencies compete for funding, which provides incentives for performance improvement (Moynihan, 2008). Along with the introduction of performance-based budgets, other related forms that also provide greater flexibility for managers in return for higher accountability include (i) allowing managers to spend on budgets by approving the budgets' total value (e.g. expenses, revenues) rather than in-detail line items, (ii) allowing carrying over unused funds, and (iii) allowing for using retained earnings (see Bergmann, 2009).

Based on the links between budget allocation and measurable results, performance budgeting can be classified into three models: presentational, performance-informed, and direct performance budgeting. In the first model, performance information (targets,

results) is presented as background information, which is mainly used for accountability and discussion but not for formal decisions. Unlike the first one, in the second model, budget allocation is indirectly associated with past or forecasted performance. Although performance information plays an important role in informing the decision-making process, it does not determine the amount of funding, but is often combined with other social-economic information. Finally, the third model requires budget allocation to be directly based on the performance achieved. In the context of OECD countries, performance-informed budgeting is more widely applied while the third model is only restricted to specific sectors rather than for a wide system (Curristine and Flynn, 2013).

More comprehensively, Pollitt and Bouckaert (2017) view the evolvement of reform in performance budgeting through the following four stages:

Figure 7. The evolvement of reform in performance budgeting



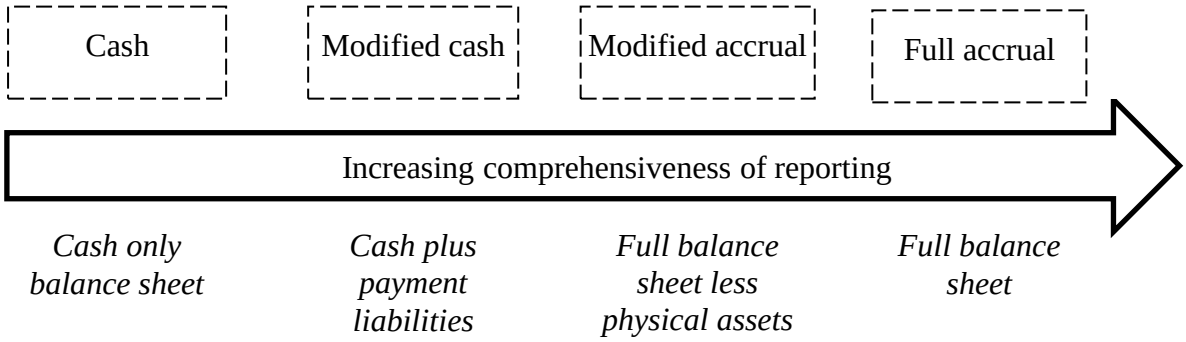
2.5.2 Accounting reforms

Similar to budgeting, as an aspect of NPM, the shift from cash-based to an accrual-based system has been a central part of financial management reforms in the public sector since the early 1990's. Given the merits of simplicity and understandability, the cash basis is criticized for considerable weaknesses. For example, in presenting the financial position and performance perspective, cash-based accounting cannot present all assets and liabilities (e.g. capital assets like land, buildings are normally treated as free goods) or separate between expenditure for investment and consumption. Also, since the full cost of using assets is not recorded, it is unable to connect between expenditure and cost as well

as between cost and performance (see Bergmann, 2009; Hyndman and Connolly, 2011; Pollitt and Bouckaert, 2017). Therefore, the introduction of accrual accounting is expected to overcome the shortcomings provoked by the cash basis approach for making better decisions. Particularly, accrual basis provides a more comprehensive fiscal position and financial performance, improves resource allocation, motivates the mangers to operate efficiently, and enhances accountability to the public (Flynn *et al*., 2016; Schick, 2013).

According to Flynn *et al.* (2016), globally, up to 2015, 41 governments have fully adopted accruals, 16 governments utilized a modified accrual basis, 28 governments are on a modified cash basis, and 114 governments (57 percent) still use cash accounting. The pathway to a full implementation of accrual basis is as follows:

Figure 8. The pathway to full implementation of accrual basis



Sources: Adopted from Blondy *et al.* (2013)

2.5.3 Responses to the financial crisis

The global financial crisis in 2008 led to economic recessions across many developed countries with an increase in unemployment rates and budget deficits. Therefore, it is essential to investigate the responses of the public sector to the financial crisis and its impacts on the reforms. From the crisis decision-making perspective, Peters (2011) suggested that in response to the financial crisis, the government have a tendency to centralize and politicize their decision-making. Also, they tend to make the decision more public when the governments have to take over the economic functions that used to be

managed by the free markets (e.g. nationalizations of companies). In addition, in times of crisis, the governments might take these as opportunities for innovations and learning; in contrast, some resist to change but simply reinforce the existing approaches. Similarly, in an analysis of OECD countries under austerity, Lodge and Hood (2012, p.86) found four possible reactions of government to the financial crisis: (i) "directing state" – state enterprise and planning replace for the business style management, (ii) "hollow state" – public services are provided by the private sector, (iii) "local communitarian state" – voluntary and community organizations are involved in providing more public services, and (iv) "coping state" – abandon part of the public services or require to provide more services with fewer resources.

As a consequence of the financial crisis, many countries experienced significant retrenchment in public spending and austerity, coupling with the pressure of ongoing reforms. Pollitt (2010) argued that although financial austerity may inhibit reforms, it can stimulate more radical changes. In particular, the author suggested that cutback approaches (cheese-slicing, targeted, efficiency gains) will result in management reform. While this cheese-slicing imposes an equal level of cut for all institutions, in targeted cut (or strategic cut) some institutions or areas face a greater cut than others. Efficiency savings normally require providing more services with a restrained budget without sacrificing service quality. Although each approach has both strengths and weaknesses in practice, the government can combine the three strategies and the common target is to achieve savings.

Regarding the extent of financial management reform, it also varies by the context under analysis. New Zealand, the pioneer reformer in the PFM, has introduced accruals accounting since 1989 and produced accrual budgets since 1992, which emphasized outputs and outcomes rather than inputs and activities. This paved the way for reforms in

other Westminster countries such as the UK, Australia, and Canada with the expectation to change the managers' behavior, provide more financial transparency, and improve performance (see Kettl, 2005; Pollitt and Bouckaert, 2017). In Japan, PFM reform commenced in 2001 as part of the NPM agendas. However, the implementation of performance budgeting and accrual financial reporting has slowly diffused within limited areas and provided insufficient information for decision-making (the amount of budget and the coverage were quite modest; financial reports only provide supplementary information for the budget system) (Yamamoto, 2009). Therefore, along with measuring the efficiency of public services organizations (the UK and New Zealand, the healthcare sector, and the Japanese public higher education, more particularly), the impacts of assets utilization and financial plan will be examined. The idea behind this is that efficiency concerns the ability to minimize the resources used and only on the accrual basis are the costs fully reflected. Therefore, the performance against financial plan and the value of assets can be used as proxies to examine the association with efficiency and, thus, probably imply motivation for further and radical implication of accrual accounting and accrual budgeting in the Japanese public sector.

2.6 Approaches to measuring the efficiency

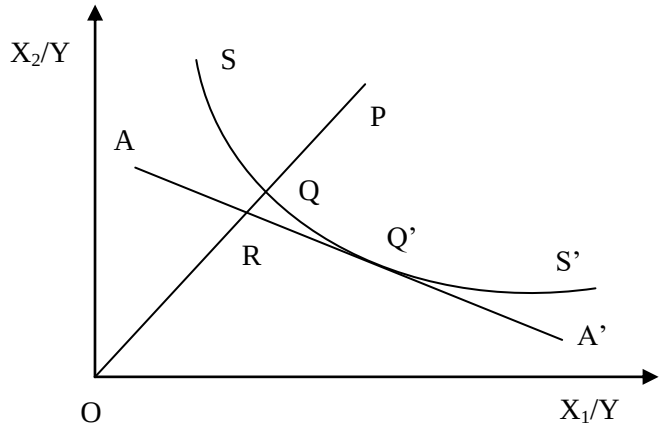
As discussed previously, a holistic assessment of the public management reform is a formidable task. We alternatively focus on the performance measurement of a specific service within a particular time period following the reforms, which might provide more insight. Particularly, measuring technical efficiency and identifying factors influencing the efficiency of the public service organizations is the key objective. Accordingly, there are three aspects that should be taken into consideration before conducting a study on efficiency: "i) definitions of efficiency measures, ii) methods for calculating efficiency measures, and iii) relevant data for inputs and outputs of activity we want to measure

efficiency for" (Førsund, 2018, p. 10). Therefore, the following section will, in turn, discuss and delineate these concerns. Moreover, since the efficiency scores could not explain why some units are more efficient than the others, the two-stage analysis is an additional technique expected to provide more insightful understanding.

2.6.1 Concepts

Efficiency is a key concept pervasively used in economics that refers to the ability to maximize the outputs produced within given resources. The seminal concept of technical efficiency stems from the paper of Farrell (1957), which is described in the following graphical illustration:

Figure 9. Farrell efficiency



Technical efficiency: Assuming a constant return to scale, the SS' curve is the best-practice production frontier (unit isoquant of efficient firms) which shows the different minimum combinations of the two inputs (X1, X2) to produce one output (Y). Accordingly, any points along the frontier SS' are technically efficient. For instance, Q is the efficient firm using inputs with a similar ratio to P. The firm with production at P is technically inefficient since it can generate the same volume of output while consuming lower levels of inputs (point Q). In other words, the technical inefficiency of firm P is the amount of inputs that can be reduced without a reduction in output, which can be

represented by the distance PQ. Thus, technical efficiency (TE) at P is measured by the relative distance to the frontier with ratio: TE = OQ/OP.

Allocative efficiency: In case market prices are available, the AA' line represents the isocost line, which is tangent to the isoquant SS' at point Q', suggesting the minimum cost is achieved at Q'. Supposedly, unit P can improve its efficiency by reducing inputs used to the level at point Q, its cost will still be higher than the minimum cost (imagine that the isocost line crossing the point Q is above the lowest isocost line-AA'). Allocative efficiency or price efficiency (AE) is given by the ratio: AE= OR/OQ.

Overall efficiency: Based on the two mentioned concepts of efficiency, the total operating cost to produce output at Q' in comparison with P is the measure of the overall (economic or productive) efficiency (OE), in which OE = OR/OP = OQ/OP x OR/OQ or OE = TE x AE.

2.6.2 Techniques in measuring efficiency

There is no single method but rather a variety of approaches to measure the efficiency of the public service organizations (the term "decision-making units" (DMUs) is used as a general term to refer any units under assessment for the consistency). While each method has both strengths and shortcomings, we emphasize more on the Data Envelopment Analysis for its pre-eminence compared to its counterparts.

2.6.2.1 Ratio analysis

Using performance indicators is a simple and common method to measure the performance of a public service organization, especially efficiency. Accordingly, efficiency is defined as the ratio of output over input or vice versa.

$$Efficiency = \frac{Output}{Input}$$

For example, to compare the performance of universities, the ratio of teachers over students can be utilized. The problem is that students at different levels consume different resources, suggesting the teacher/student ratio should be disaggregated into multiple ratios to capture the variances. However, this will confuse the managers since there is no unique ratio for benchmarking. Therefore, ratio analysis only works well in the simple production process using single input to produce a single output. In the real world where normally multi-inputs and multi-outputs are simultaneously used, it is impossible to assess the performance of different units through a list of ratios with mixed results (see Thanassoulis, 2001, Ozcan, 2014)

2.6.2.2 Ordinary Least Square (OLS)

On the assumption that all DMUs are efficient, the OLS is a parametric technique that can be applied in the context where a single input is used to produce multiple outputs or alternatively multiple inputs used to generate a single output. This model can be illustrated by the general equation as follows:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ ... + \beta_n x_n + \varepsilon$$

In principle, by adopting the OLS technique, the model assumes that y is a random variable and independent from others, β_i is the set of parameters to be estimated, ε is the proxy for random noise and has a normal distribution with a mean value equal to zero. Accordingly, assuming y denotes for input and x_i is the set of output, with OLS, we can estimate the predicted level of input of a certain DMU, which is used to compare with the actual input to measure input efficiency.

Although OLS can accommodate multiple inputs or outputs and can be used to measure technological changes when time-series data are available, this approach has many weaknesses: (i) it estimates the average value rather than measure efficiency, and (ii) all variance between the predicted value and observed value is attributable to inefficiency. In

addition, the parametric formulation requires a hypothetical type of production function (e.g. linear, quadratic, exponential), which is not always correctly specified (see Ozcan, 2014, Thanassoulis, 2001, Coelli *et al.,* 2005)

2.6.2.3 Stochastic Frontier Analysis (SFA)

SFA is also a parametric method to partly overcome the criticism of OLS approach.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \mu + \alpha$$

Unlike OLS, SFA assumes all DMUs are non-efficient and the random noise (ε) now is disaggregated into two terms: (i) random error (α) and (ii) inefficiency (μ). While α is assumed to be normally distributed, $\mu \geq 0$ and its distributions are assumed to be half-normal or exponential. The set of parameters (β) can be estimated by using OLS or Maximum Likelihood methods and the efficiency of DMU can be identified as the predicted value of μ given the observed value of μ + α at that DMU.

SFA has advantages in accounting for noises and can be used for conventional hypothesis testing. However, its shortcomings include the need to hypothesize the functional form for production function and the form of distribution for the inefficiency term, which is vulnerable to misspecification. Also, similarly to OLS, SFA could not handle the cases that require to incorporate multiple inputs and outputs simultaneously (see Ozcan, 2014, Thanassoulis, 2001, Coelli *et al.,* 2005)

2.6.2.4 Data Envelopment Analysis (DEA)

Basic concepts

On the foundation of technical efficiency concept in the work of Farrell (1957) with single output/input, Charnes, Cooper, and Rhodes (1978) developed DEA, a non-parametric method based on linear programming, to measure relative efficiency in the cases of multiple output/input (Charnes *et al*., 1994). Accordingly, the relative technical efficiency of a certain DMU is estimated by calculating the ratio of the weighted sum of outputs to

the weighted sum of inputs with the weights as unknown variables, as illustrated in the following formula:

$$Max\ h_0(u,v)\ = \frac{\sum_{r=1}^{s} u_r y_{r0}}{\sum_{i=1}^{m} v_i x_{i0}}$$

$$\text{Subject to: } \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}} \leq 1\ for j = 1, \dots, n$$

$$u_r, v_i \geq 0\ for\ all\ i\ and\ r$$

where, h_0 is technical efficiency score, x_{ij} is the amount of input i that DMU_j used to produce amount y_{rj} of output r, and u_r and v_i are non-negative weights assigned to output r and input i, respectively. Linear programming is employed to calculate the optimal weights that maximize the outputs/inputs ratio (efficiency) subject to the constraint that the efficiency score must be less than or equal to unity. The transformation of the ratio form to linear programming form is as follows:

$$\sum_{r=1}^{s} u_r y_{rj}\ \leq 1 * \sum_{i=1}^{m} v_i x_{ij}$$

$$\sum_{r=1}^{s} u_r y_{rj}\ - \sum_{i=1}^{m} v_i x_{ij}\ \leq 0$$

Define $\sum_{i=1}^{m} v_i x_{i0} = 1$, so that

$$Max\ h_0(u,v) = \sum_{r=1}^{s} u_r y_{r0}$$

Accordingly, the mathematical problem of the input-oriented[9] model can be presented as follows:

[9] The model can be reoriented to output-orientation by examining the ratio of weighted sum of inputs to weighted sum outputs (see Cooper *et al*., 2011 for more detailed transformations)

Table 4. Transformations of DEA model

Primal model (Multiplier model)	Dual model (Envelopment model)	
$Max\ h_0(u,v) = \sum_{r=1}^{s} u_r y_{r0}$	$Min\ \theta$	
Subject to	Subject to	
$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0$	$\sum_{\mathrm{j}=1}^{n} \lambda_j x_{ij} \leq \theta x_{i0}$	$i = 1, \ldots, m$
$\sum_{i=1}^{m} v_i x_{i0} = 1$	$\sum_{\mathrm{j}=1}^{n} \lambda_j y_{rj} \geq y_{ro}$	$r = 1, \ldots, s$
$u_r, v_i \geq 0\ for\ all\ i\ and\ r$	$\lambda_j \geq 0$	$j = 1, \ldots, \mathrm{n}$

In essence, DEA identifies a group of best practice DMUs with efficiency scores equal to one, which form a piece-wise linear frontier from the combination of efficient DMUs. Therefore, inefficient units are enveloped by the frontier and the technical efficiency of each DMU, which is then calculated as the distance of each unit from the frontier (see Coelli *et al*., 2005; Ferrari, 2006). Therefore, it should be emphasized that the efficiency estimated by DEA is "relative efficiency," which is defined as "a DMU is to be rated as fully (100%) efficient on the basis of available evidence if and only if the performances of other DMUs does not show that some of its inputs or outputs can be improved without worsening some of its other inputs or outputs" (Cooper *et al.*, 2011, p. 3).

Initially, the DEA model invented by Charnes, Cooper, and Rhodes (1978) assumes the production process is characterized by constant returns to scale (CCR), later extended by Banker, Charnes, and Cooper (1983) to variable returns to scale (BCC). The CCR model assumes aconstant return to scale (CRS), suggesting the inputs and outputs proportionally increase or decrease. On the contrary, the BCC model relaxes this assumption, allowing for variable returns to scale (VRS), which implies the existence of economies or diseconomies of scale.

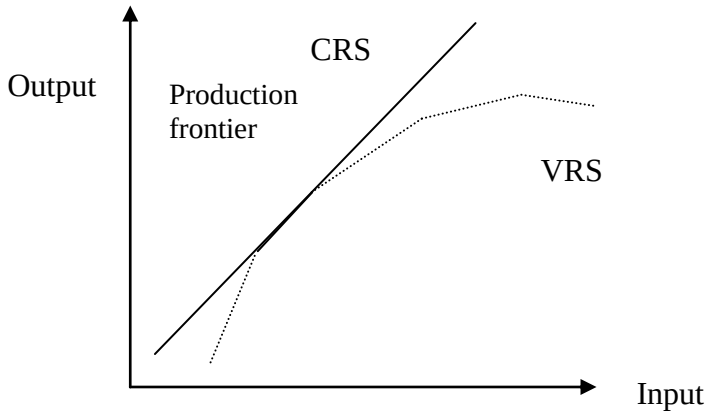
In comparison to the parametric approach, the most advantageous aspect of DEA is the capability to incorporate multiple inputs and outputs at the same time. Also, unlike the parametric method, the piece-wise linear frontier suggests there is no need to specify the production model and thus avoid the risk of model misspecification (Coelli *et al.,* 2005). In addition, public services are characterized by the complexity with multiple inputs and outputs and the market prices are not available to allow straight-forward comparison (using the price to calculated weighted inputs and weighted outputs). Therefore, DEA is particularly suitable for measuring the efficiency of public services such as health and education (G. Johnes and Tone, 2017). Indeed, in an analysis of two most popular approaches used in performance measurement research over the period from 1987 to 2011, Lampe and Hilgers (2015) detected that while DEA was most applied in "Operation Research" with a majority of publications (4,021), SFA is more employed in "Economics" and with a much lower number of publications (761). Likewise, a review of approaches to measure efficiency in 223 empirical papers on education from 1977 to 2015 indicated that DEA was the dominant technique when compared with SFA (De Witte and López-Torres, 2017). Also, a bibliographic collection of DEA literature over the period from 1978 to 2016 indicated that education and healthcare are among the five most pervasive application fields of DEA methodology (Emrouznejad and Yang, 2018). The main drawback of DEA is its deterministic nature, which does not account for noise. However, this shortcoming can be overcome by the bootstrapping (resampling the original data of inputs and outputs) to avoid bias in data used and identify the confidence interval for efficiency scores (Ozcan, 2014).

Model specification

Basically, the use of DEA revolves around the choice between the CCR model and the BCC model (related to returns to scale), model orientation (input-oriented or output-oriented), and the selection of inputs and outputs.

First, a crucial point in DEA is the selection of envelopment surfaces between (i) constant returns-to-scale surface (CCR model) and (ii) the variable returns-to-scale surface (BCC model) (Charnes *et al.*, 1994). As discussed previously, the constant return to scale assumes a constant rate of substitution between inputs and outputs. In contrast, when economies of scale exist, the shape of the frontier is totally different from that in the CRS model. Accordingly, a proportional increase in one or more inputs can cause either a greater or a lower than proportional increase in outputs, suggesting the variation in returns to scales. In practice, there is no common rule for specifying CRS or VRS but normally relies on the analysts' understanding of the context, the purpose of analysis (Jacobs *et al.*, 2006). The following graphical illustration provides more details on the differences between these two surfaces:

Figure 10. Constant return to scale and variable return to scale



Second, as presented in the ratio form of DEA, technical efficiency can be either the maximization of outputs for a given level of inputs (output orientation) or the minimization of inputs for a given level of outputs (input orientation). However, the choice of orientation depends on the factors that DMUs can control (Jacobs *et al.*, 2006). Therefore, if DMUs have discretion in deciding the level of resources used then we should evaluate the ability to minimize inputs keeping outputs fixed; otherwise, the ability to maximize outputs keeping inputs fixed should be evaluated. For example, the input-

oriented approach is mainly applied when measuring the efficiency of hospitals since normally, the emphasis is on controlling cost rather than expanding demand, and the managers have more control over inputs than outputs (O'Neill *et al.*, 2008). In addition, although under the CRS model, there is no difference in efficiency scores whether input-orientation or output-orientation is selected; under the VRS model, these are not always the same. Therefore, the choice of input or output orientation should be carefully considered due to its effect on the efficiency scores under the VRS assumption.

Finally, the selection of inputs and outputs is a decisive step since DEA is a deterministic method; the results of analysis (DEA scores) are sensitive to any mistakes in selecting variables or data errors (e.g. outliers). Although the concepts of inputs and outputs are quite straightforward and understandable, the researchers often neglect in properly choosing the factors to adequately capture the production process under study (Cook *et al*., 2014). Therefore, in principle, the selection of input and output sets should fully cover the resources used and capture all activities (Dyson *et al.*, 2001a). Indeed, both omitting important inputs or outputs or including unrelated factors can distort the result (Jacobs *et al.*, 2006). Also, uncontrollable factors should be excluded from the DEA model (Chilingerian and Sherman, 2011) and avoid mixing up absolute and relative data (Dyson *et al.*, 2001a). In addition, when the number of inputs and outputs increases, the DEA will lose its discrimination power, suggesting most of the units under evaluation will have similar efficiency scores. Therefore, as a rule of thumb, it is suggested that the number of DMUs should be least two times greater than the total number of inputs and outputs to ensure the degrees of freedom (Cook *et al.*, 2014).

Two-stage DEA analysis

As a benchmarking method, DEA can identify efficient and inefficient DMUs but unable to explain why there exist variations in efficiency scores derived. Therefore, the two-stage analysis is expected to provide more insight on the determinants of technical efficiency,

which is likely to be of key interest to managers and policy-makers (more in-depth information for making the decision is obtained). Although the two-stage DEA analysis is an exploratory approach rather than being based closely on theory, it is one of the two important evolutions of DEA that attempts to identify the determinants of efficiency despite certain limitations (Førsund, 2018).

The basic idea in the two-stage analysis is that the efficiency scores estimated by DEA exercise in the first stage (dependent variable) will be regressed against explanatory factors (contextual variables) in the second stage to identify the factors that have a statistically significant association with efficiency scores. Accordingly, different types of regression models have been proposed to relate efficiency scores to possible influencing factors. Since DEA scores are bounded by one and zero, making them similar to corner solution variables, Tobit regression is considered a suitable approach. However, McDonald (2009) argued that DEA scores are not censored values or conner solution data but fractional data; Tobit is therefore inappropriate and asserts that ordinary least square (OLS) is a more consistent estimator. Although advocating OLS, Hoff (2007) concluded that the Tobit model is still acceptable. Lovell *et al.* (1994) suggested to use super efficiency scores (can be greater than one) as a dependent variable to remove the upper-bound problem (thus justify for Tobit model). Similarly, Banker and Natarajan (2008) confirmed that ordinary least square (OLS), maximum likelihood (ML), or even Tobit can be applied in the second stage analysis. Especially, they proposed statistical foundation to justify for the applications of regression model used in earlier DEA studies, in which the error term or deviation from the production frontier includes three factors: "(i) a linear function of multiple, possibly correlated, contextual variables; (ii) a one-sided inefficiency term; and (iii) a two-sided random noise term bounded above" (Banker and Natarajan, 2008, p. 49). They also run simulations and found that the two-stage method works well

when the levels of correlation between the contextual variables and the inputs are low, and vice versa. Thus, since the Tobit model is not an appropriate approach due to its mis-specifications of DEA scores, more scholars advocate using OLS in the second stage DEA instead. Accordingly, a simple regression model using OLS can be specified as follows:

$$\hat{\theta}_i = \beta_0 + \beta Z_i + \varepsilon_i$$

Where: $\hat{\theta}_i$ is DEA technical efficiency scores obtained the first stage; β is parameter estimated from the regression model; Z_i denotes contextual factor; ε_i proxies for two-sided random noise.

In a different approach, Simar and Wilson (2007, 2011) criticized the conventional two-stage DEA for two issues: (i) lacking a coherent data generating process (DGP) to provide a rationale for the two-stage procedure; and (ii) possibility of serial correlation of efficiency scores since they are estimated from a common sample. Therefore, in their model, the data generating process is assumed that efficiency of DMUi is directly influenced by the contextual variables as follows:

$$\theta_i = \beta Z_i + \gamma_i$$

Where: β proxies for coefficients estimated from the model; γ_i denotes the part of inefficiency that cannot be explained by contextual variables and are assumed to be independently and normally distributed as $N\ (0, \sigma_\gamma^2)$) with left truncated at $1 - Z_i\beta$.

In addition, they argued that the dependent variable (θ_i) is unobserved as clearly explained by Badunenko & Tauchmann (2018, p. 5) that: "the estimated efficiency score $\widehat{\theta_i}$ one obtains from running a DEA is not θ_i. In other words, $\widehat{\theta_i}$ is not the distance of ($\mathbf{Y}_i$, $\mathbf{X}_i$) to the true production-possibility frontier but the distances to an estimate of the latter". Also, in a small sample, the DEA scores estimated ($\widehat{\theta_i}$) are biased toward the unity (value of one). Therefore, Simar and Wilson (2007) developed two bootstrap procedures (single

bootstrap - Algorithm #1 and doubled bootstrap - Algorithm #2) to overcome problems of serial correlation and finite sample bias. Basically, in the bootstrapping procedure in Algorithm #1, the confidence intervals and standard errors for $\hat{\beta}$ and $\hat{\sigma}$ are calculated from the bootstrap distribution of $\widehat{\beta^b}$ and $\widehat{\sigma^b}$ that are obtained by repeating B times following three steps:

(1) For each DMU, artificial error ($\tilde{\gamma}$) is drawn from N $(0, \hat{\sigma})$ distribution with left truncated at $1 - Z_i\hat{\beta}$;

(2) Artificial efficiency score ($\widetilde{\theta_i}$) is calculated for each DMU with the formula: $\widetilde{\theta_i} = \hat{\beta}Z_i + \widetilde{\gamma_i}$ ($\hat{\beta}$ is obtained from truncated regression of efficiency scores initially estimated on Z_i);

(3) Run a truncated regression $\widetilde{\theta_i}$ on Z_i to obtain maximum likelihood, bootstrap estimates $\widehat{\beta^b}$ and $\widehat{\sigma^b}$.

Unlike Algorithm #1 that only uses conventional efficiency scores, the doubled bootstrap procedure (Algorithm #2) calculates the bias-corrected efficiency scores ($\widehat{\theta_i^{bc}}$) through artificial DMUs. A truncated regression of $\widehat{\theta_i^{bc}}$ on Z_i is then conducted to obtain $\hat{\hat{\beta}}$ and $\hat{\hat{\sigma}}$. The confidence intervals and standard errors for $\hat{\hat{\beta}}$ and $\hat{\hat{\sigma}}$ are calculated from the bootstrap distribution of $\widehat{\hat{\beta}^b}$ and $\widehat{\hat{\sigma}^b}$ with similar steps as described in Algorithm #1 (see Simar and Wilson, 2007 and Badunenko & Tauchmann, 2018 for detailed steps and explanations).

Given that two-stage DEA analysis has been widely applied, selecting the regression model is an ongoing debate, particularly between the authors of two influential papers (Simar and Wilson, 2007 and Banker and Natarajan, 2008) that lay the statistical foundation to justify for the applications of the regression model. Simar and Wilson (2011, p. 205) claimed that "second-stage OLS estimation is consistent only under very peculiar

and unusual assumptions on the data-generating process that limit its applicability. In addition, we show that in either case, bootstrap methods provide the only feasible means for inference in the second stage". On the contrary, even though the publications using Simar and Wilson approach have significantly increased after 2007, it is argued that "Simar and Wilson completely ignore the role of two-sided noise in the production environment which raises serious doubts about the applicability of their method in stochastic environments" and asserted that DEA+OLS significantly outperforms Simar and Wilson truncated regression model (Banker *et al*., 2019, p. 369). Indeed, in relation to the controversy around the two-stage analysis, Liu *et al*. (2016, p. 39) indicated that: "From a practical point of view, the current state of development, nevertheless, still leaves some confusion to practitioners whose true need is a clear guidance on what methodology to use."

In our research, we use both two key approaches to estimate the impacts of contextual factors on efficiency. In particular, in the case of the English hospitals, we basically follow DEA + OLS approach. However, it should be noted that the super efficiency scores are employed as a dependent variable rather than conventional DEA scores to avoid boundary problems and thus justify the Tobit model. On the other hand, the double bootstrapped truncated regression model (Algorithm # 2) is applied to identify the determinants of efficiency for the cases of New Zealand DHBs and Japanese national universities. As mention previously, one of the key underlying purposes of doubled bootstrapped truncated regression is to address the correlation and finite sample. In this regard, unlike English acute foundation trusts of which outputs are only related to hospital activities, outputs for New Zealand DHBs also include activities in primary healthcare services related to health targets (e.g. immunization, advice to quit smoking) mandated by the Ministry of Health. Similarly, a wide range of activities is specified for Japanese national universities

including education, research, and hospital services. It seems that the possibility of auto correlation of outputs between different years might be relatively higher in the case of New Zealand DHBs (e.g. the next year target might rely on the performance of the preceding year) and Japanese national universities (e.g. the number of students). In addition, the sizes of sample to conduct the two-stage analysis in both New Zealand DHBs (67 observations) and Japanese national universities (421 observations) are smaller than that of English acute foundation trusts (483 observations), suggesting that the bootstrap procedure is more suitable to avoid possible bias due to small sample. Moreover, technically, the regression model specified for English acute foundation trusts includes a large number of variables (seven inputs, four outputs, and ten contextual variables), which is beyond the capacity of our available tools for analysis with the bootstrapping procedure.

Overview of the DEA development and application

It has been more than 40 years since its invention in 1978; DEA has been developed into many paths with an exponential increase in the number of publications widely applied in various subfields. In particular, in a survey of DEA literature from 1978 to 2010 (4,936 papers), Liu *et al*. (2013) identified that the development of DEA literature can be divided into two phases: the first phase began at 1978 and saturated around 2009 while the second phase started in 2001 and forecasted to reach the saturation in the 2020s. The authors suggested that the latter phase had faster growth than the earlier period; probably due to the advantage of DEA and the availability of data and software tools. In addition, they classified recent DEA literature into five major branches, including: ‘‘two-stage contextual factor evaluation framework’’, ‘‘extending models’’, ‘‘handling special data’’, ‘‘examining the internal structure’’, and ‘‘measuring environmental performance’’, in which the ‘‘two-stage contextual factor evaluation framework’’ draw relatively more attention. Similarly, Liu et al. (2016) categorized DEA literature from 2000 to 2014

through a citation network and found that the four main subareas of DEA include: (i) "bootstrapping and two-stage analysis": DEA scores are estimated with confidence intervals and regressing DEA scores on explanatory factors; (ii) "undesirable factors": approaches to estimate efficiency with undesirable outputs or inputs (e.g. wastes, pollutants, mortality rate), (iii) "cross-efficiency and ranking": an extension of DEA model to increase the discriminating power of DEA, and (iv) "network DEA, dynamic DEA, and slacked-based model (SBM)": network DEA and dynamic DEA consider the internal production process and the interdependence between periods while SBM considers the slacks (excesses in inputs and shortfalls in outputs) when measuring efficiency. Further analysis, they also found there exists association between DEA application on a specific field and a methodology. For example, undesirable factor models were well examined in the cases of energy, environment, and agriculture; network DEA and SBM models are popular in the studies banking and finance sector whilst two-stage analysis was widely used in DEA applications for education and healthcare. Recently, a bibliometric analysis of DEA literature conducted by Emrouznejad and Yang (2018) indicates that, by the end of 2016, there were 10,300 DEA-related journal articles with 11,975 authors (except for conference proceedings, working paper, and book chapter). Also, the authors identified that the DEA application is most popular in the five areas including energy, industry, banking, educations, and healthcare.

Aside from the review of general DEA literature, several studies also systematically summarize the existing DEA literature applied for a specific field. For instance, Sueyoshi *et al*. (2017) reviewed 693 DEA articles on energy and environmental issues (e.g. electricity industry, energy saving, environmental efficiency) and found a surge in the number of publications after 2000s. In addition, history of DEA and collection of DEA applications in a wide range of fields in both public and private sector (e.g. banking,

transportation, country and regional performances) are also well reflected in several books specifically to DEA method (see Cooper *et al*., 2011; Zhu, 2016)

Thus, DEA has become a versatile methodology for efficiency measurement and benchmarking. An overview of existing DEA literature also suggests that two-stage DEA analysis is popular in the DEA applications on healthcare and education, which is selected as the key approach in our study to measure the efficiency of the public service organizations in the contexts specified.

2.7 Qualitative content analysis

Along with the quantitative approach used to measure technical efficiency and identify the determinants of efficiency, content analysis using interview transcripts is also employed to complement and provide a more insightful understanding.

According to Krippendorff (2004, p. 18), "Content analysis is a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the context of their use". More specifically, the author suggests that the same results should be obtained when the same technique is applied (reliability) and that the results should be carefully scrutinized to ensure validity. Also, texts are not only exclusively confined to documents, but also include a wide range of communication media such as images, maps, audio recordings, television shows, social media, and even numerical records (Drisko and Maschi, 2016; Krippendorff, 2004).

Based on the purposes and techniques applied, content analysis is categorized into (i) basic content analysis, (ii) interpretative content analysis, and (iii) qualitative content analysis. While the first approach focuses on quantitative methods (e.g. word counts) to analyze data, the second aims to uncover the latent contents (inferences) than merely counting the word frequency. Finally, as a recent approach, qualitative content analysis is defined by

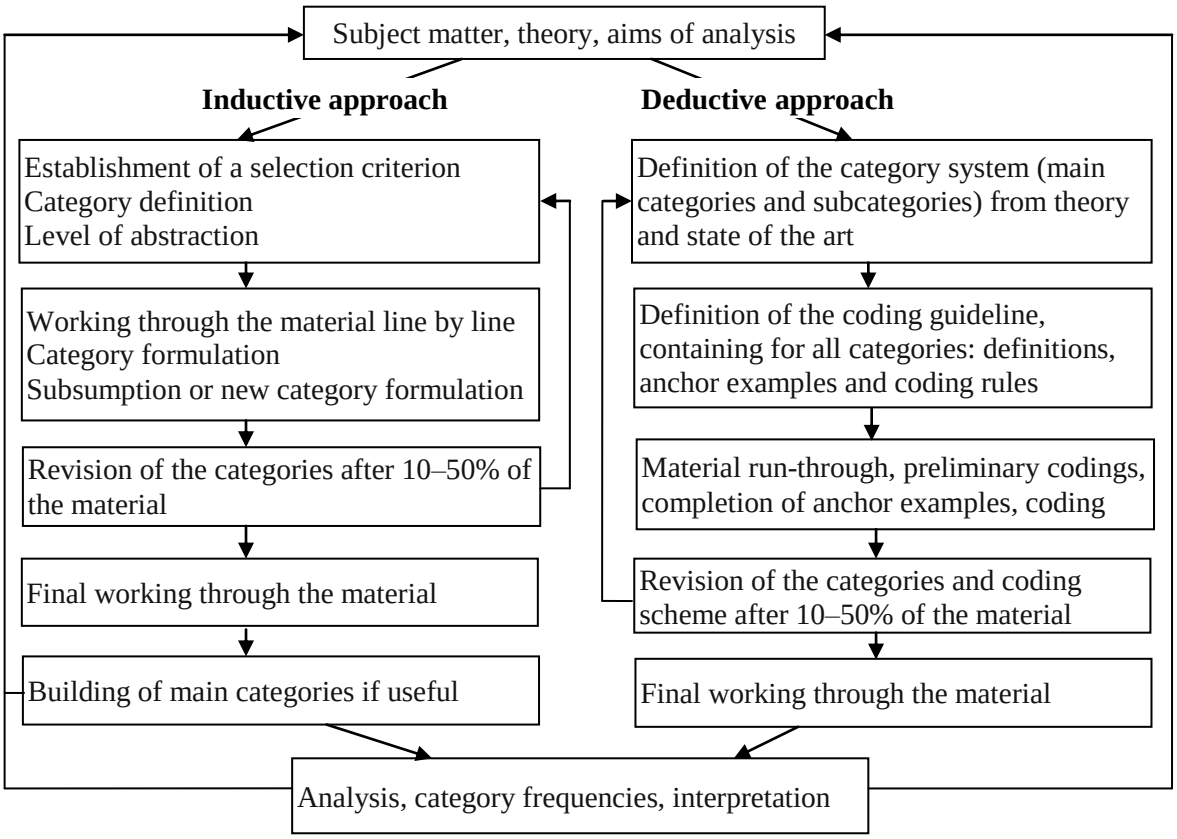
Mayring (2000, paragraph 5) as “an approach of empirical, methodological controlled analysis of texts within their context of communication, following content analytical rules and step-by-step models, without rash quantification”.

Regarding data collection for analysis, apart from existing texts, newly generated interviews are common data sets used in qualitative content analysis (Elo *et al.*, 2014). There are three types of interviews: structured, semi-structured, and unstructured. In structured interviews, a list of predetermined questions is used and there is no room for probing follow-up questions. On the contrary, unstructured interviews do not specify any question in advance, which is normally very time-consuming and can be difficult to manage. More flexibly, semi-structured interviews include both key questions and allow for expanding more detailed ideas through interactive discussions between interviewers and interviewees during the interviews (Gill *et al.*, 2008).

Qualitative content analysis can be conducted inductively, deductively, or as a combination of two ways subject to the purpose of the study. The former approach is more suitable when the researchers have insufficient and fragmented knowledge about the phenomenon; on the contrary, the latter is recommended when the study aims at theory testing (testing categories, concepts, models, or hypotheses) based on previous knowledge (theories, model, literature review) (see Drisko and Maschi, 2016; Mayring, 2015).

Although there is no step-wise procedure applied, in principle, qualitative content analysis is to systematically summarize a large amount of raw texts into concise key themes. Four basic steps in conducting qualitative data analysis include (i) familiarizing oneself with the data and the hermeneutic spiral, (ii) dividing the text into meaning units and condensing meaning units, (iii) formulating codes, and (iv) developing categories and themes (Erlingsson and Brysiewicz, 2017). More particularly, the detailed procedures may vary depending on the deductive or inductive approach as follows:

Figure 11. Procedures of content analysis



Sources: Adapted from Mayring (2015)

Accordingly, qualitative content analysis is employed to investigate the implementation of performance measurement at New Zealand DHBs (complemented by the evidence from Australian and Vietnamese hospitals) and the perceptions of managers at Japanese national universities on the changes following the incorporations.

In summary, while this chapter has reviewed the three key movements in the public management reform since the 1980's, it is likely that the core elements of NPM have been maintained and employed. The emergence of new reform waves such as Neo-Weberian state and New Public Governance do not totally break with NPM but rather complement additional features or modify certain aspects. Key aspects of reforms related to financial management were also summarized. In addition, improving the performance of the public sector has been the crucial goal of any reform; we also elaborated performance measurement and performance management. This chapter provided the methods to measure efficiency, especially two-stage DEA which is applied in future empirical studies.

CHAPTER 3: PERFORMANCE PARADOX: AN INVESTIGATION OF PERFORMANCE MEASUREMENT AT NEW ZEALAND DISTRICT HEALTH BOARDS

Performance measurement is often perceived as a vital tool for daily management practices and public policies to improve the efficiency, accountability, and quality of public services. However, the utility of performance measurement regimes is sometimes hampered by the performance paradox which is the difference between performance on paper and the actual performance perceived by various stakeholders. We employed a content analysis technique using semi-structured interviews with senior managers at New Zealand District Health Boards (DHBs), supplemented by interviews at Vietnamese and Australian hospitals to confirm the robustness of results obtained, and found evidence of paradox in the performance measurement system. The main implication arising from our work is that more patient-focused indicators should be developed to capture the healthcare outcomes and to counter possible manipulation of performance information. We also identified the need for a single holistic measure – such as technical efficiency – which might help end-users gain additional utility from the performance measurement regime.

3.1 Introduction

As a key feature of New Public Management, performance management and performance measurement have become core principles, being adopted in many countries (e.g. the US, the UK), as well as advocated by a number of international organizations such as World Bank and OECD (Boyne, 2010; J. M. Lewis, 2015). As a leading country in the NPM movement, New Zealand is a key example of early and continued use of performance-based management throughout its public service reforms with an emphasis on outputs or outcomes rather than inputs (B. Allen and Eppel, 2019; Scott and Boyd, 2016)

In recent times, performance management in New Zealand healthcare services has placed increasing importance on the use of measurable targets as a policy tool. Accordingly,

health targets have been utilized as a cornerstone of the performance framework since 2007 (Knopf, 2017). The targets are established by the Ministry of Health and District Health Boards (DHBs) which have a responsibility to report their progress against the targets to the Ministry of Health quarterly. Moreover, the Ministry of Health publishes performance achieved by DHBs for each target on its website (Chalmers *et al.*, 2017; Knopf, 2017). These health targets comprise "a set of national performance measures specifically designed to improve the performance of health services that reflect significant public and government priorities" (Ministry of Health, 2019, p.1). Over time the targets have been revised or replaced[10], resulting in the current six targets: shorter stays in emergency departments, improved access to elective surgery, faster cancer treatment, increased immunization, raising healthy kids, and better help for smokers to quit (see the Appendix 1 for further details).

In addition to the targets mandated by the Ministry of Health, DHBs also internally develop a range of their own performance indicators for managing health care in their corresponding regions. For example, as the largest DHB in New Zealand, Auckland DHB classifies its outputs into four categories: prevention services, detection and management, intensive assessment and treatment, and rehabilitation and support services. Each output class is then divided into sub-categories with detailed indicators (output measures). For instance, "intensive assessment and treatment" includes acute services, mental health, maternity, elective surgery, and quality and patient safety. A typical metric for the elective sub-category is the "percentage of people receiving urgent diagnostic colonoscopy in 14 days". Based on the performance relative to the individual targets set, the rating for each target can be categorized into "achieved", "substantially achieved", "not achieved but progress made", and "not achieved" (Auckland District Health Board, 2017). However, it

[10] Initially, there were 10 targets: improving immunization coverage; improving oral health; improving elective services; reducing cancer waiting times; reducing ambulatory sensitive (avoidable) hospital admissions; improving diabetes services; improving mental health services; improving nutrition, increasing physical activity, and educing obesity; reducing the harm caused by tobacco; reducing the percentage of the health budget spent on the Ministry of Health (see Ministry of Health, 2008)

should be noted that the internal target-setting and evaluation might vary by DHB subject to their priorities, capabilities, and objectives. Thus, performance-oriented management is extensively applied in New Zealand DHBs for the purpose of not only giving accountability to the public but also improving the performance of DHBs as well as the entire healthcare sector.

In addition, to strengthen our findings and arguments, we extend our analysis to Vietnamese hospitals and Australian public hospitals where targets or performance indicators are also implemented. The Australian case was used to augment the New Zealand evidence in view of the fact that the hospital and health systems are quite similar in the two countries. Vietnam, on the other hand, has different health, economic and political systems which might provide some interesting light on the implementation of performance measurement. The framework for quality assessment of Vietnamese hospitals is stipulated by the Ministry of Health (Decision No. 6858/QD-BYT dated November 18, 2016). Accordingly, 83 indicators are used to evaluate hospitals' quality (both public and private), classifying into five key areas: patient-oriented activities (19 indicators), human resources development (14 indicators), professional activities (35 indicators), quality improvement (11 indicators), and specialist area (4 indicators). Each area is also categorized into sub-criteria with performance being measured on a five-point scale. For instance, the patient-oriented activities include a sub-criterion: One hospital bed is used for only one inpatient (no more than two inpatients stay in the same bed); a hospital will be scored one (lowest level) if three or more patients staying in one bed (except for natural disasters, catastrophes, and epidemics) (Ministry of Health, 2016a). In the case of Australia, under the Performance and Accountability Framework outlined by Council of Australian Governments, hospital performance is evaluated through 17 indicators, covering various aspects such as effectiveness, quality, access, efficiency, equity, competence, capability, continuity, responsiveness, and sustainability (Braithwaite *et al.*,

2017). For example, with regard to waiting times for emergency department care—proportion completed within 4 hours, unplanned readmissions, waiting times for elective surgery, or potentially preventable hospitalizations - are among the performance indicators that annually reported by the Australian Institute of Health and Welfare (AIHW) (Australian Institute of Health and Welfare, 2018a).

Performance measurement has been employed to better support decision-makers in improving efficiency and providing accountability to the public, policy-makers, and other stakeholders (see Poister, 2003; Moynihan, 2008). However, dysfunctions might distort the real picture of performance and undermine the merits of performance measurement tools. Dysfunctions of performance measurement in the public services have been well documented and, indeed, the literature on dysfunctional effects exceeds those on functions of performance measurement (Dooren *et al.*, 2015). Therefore, it is critical that the dysfunctions are identified and mitigated to minimize undesirable effects. In the context of New Zealand healthcare services, it seems that there is a lack of general assessment on the undesirable consequences of target-based management. Therefore, in this study, we employ a content analysis technique to identify the problems, especially with respect to the performance paradox through an investigation of performance measurement in health boards in New Zealand supplemented with evidence from Vietnamese and Australian hospitals.

In the next section, we develop the theoretical framework explaining the components of the performance paradox and review the extant literature on the dysfunctions of performance measurement. This is followed by a content analysis of semi-structured interviews conducted in each nation, with a discussion on the manifestation of performance paradox. We conclude our article with a review of the public policy implications and suggestions for further research.

3.2 Theoretical framework and literature

The performance paradox is a well-established concept in the scholarly literature on performance management (see, for example, De Bruijn and Van Helden, 2006; Pollitt, 2013). Essentially it refers to the gap between the performance on paper, as measured by various metrics, and the actual performance perceived by different stakeholders. Notably, the existence and magnitude of the performance paradox is mediated through the experience and knowledge of specific stakeholders – for instance, the performance paradox perceived by senior management of a DHB may well be quite different from the performance paradox perceived by its patients.

Extant literature suggests that performance paradox is derived from three principal factors (Drew and Gamage, 2018). First, intended error may result from deliberate actions such as gaming or deliberate misinterpretation, in which "the former is the manipulation of actual behavior while the latter leads to distortion in reported behavior" (Smith, 1995, p. 298). Moreover, Drew and Gamage (2018) argue that gaming is not an example of fraud or cheating, but rather examples of people exploiting grey areas to provide a more favorable impression on performance. More specifically, ratchet effects, threshold effects, and output distortions are the three main types of gaming of which extent is subject to the motive and opportunity of service providers (Bevan and Hood, 2006). Detection and explication of intended error is a major part of the extant literature (see, for example, Bevan and Hood, 2006; Drew and Grant, 2017).

The second source of the performance paradox is unintended error. Compared with the intended error, this source has largely been ignored by performance management scholars (for a notable exception see Drew and Gamage, 2018). Particularly in the cases where dichotomous benchmarks are employed, or time-series data is used for the interpretation of performance metrics, small errors in the source data can result in misleading conceptions of performance (a performance paradox). It is therefore important to exert appropriate effort into auditing all of the data employed for the purposes of constructing

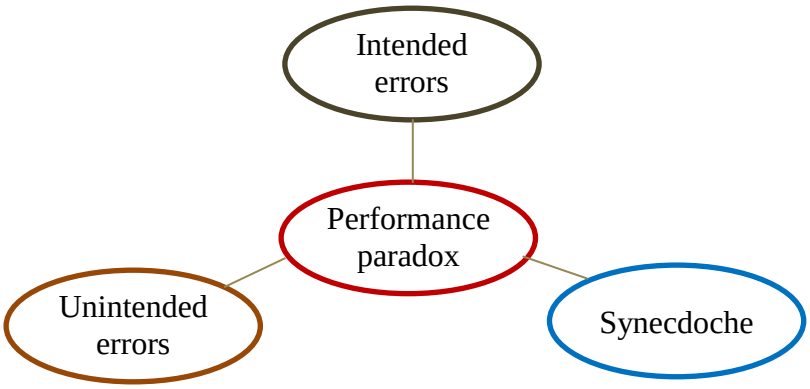
performance metrics – and the effort exerted should be increased as the importance to which the performance management regime increases (Bird *et al.*, 2005).

The third source of performance paradox is the problem of synecdoche. Synecdoche is the rhetorical trope of taking a part to stand for the whole, or the whole to stand for the part (Drew *et al*., 2018). This source of the performance paradox and the problems engendered has also been subject to scholarly inquiry recently (see, for example, Bevan and Hood, 2006; Pollit, 2013). Essentially the literature had argued that the gap between what was measured (the part) and what was not (the whole), had a direct association with the size of the performance paradox. Moreover, the literature noted that a synecdochical gap was almost certain to arise given that not all aspects of performance are amenable to measurement and also because too many metrics are not advisable given that large amounts of data can make it difficult for end-users to form conclusions on overall performance and thus reduce the utility of the performance measurement regime (Bevan and Hood, 2006; Smith, 1995).

More recently attention has been brought to the salience of whole for the part synecdoche (see Drew *et al.*, 2018; Drew and O'Flynn, 2019), particularly with respect to performance paradox. Stakeholders view performance metric data through a lens closely approximating the whole and it is therefore not surprising that they might disagree with the impression painted by the parts emphasized by the performance management regime. It has been argued that this synecdoche is an important reason for stakeholder perceptions of the performance paradox. Indeed, in a study of the manipulation of performance management regimes by Australian schools, Drew and O'Flynn (2019) demonstrated that much of what had hitherto been labelled as gaming was, in fact, an attempt by stakeholders to better reflect their knowledge of the whole in the part chosen by regulators for performance reporting purposes.

Figure 12 summarises extant accounts of the performance paradox in the corpus of scholarly literature.

Figure 12. The three causes of performance paradox



However, we believe that the extant literature has largely neglected an important element of the performance paradox – certainly with respect to performance management regimes that are operated over multiple years. Exogenous factors may conspire to alter the gap between performance on paper and performance assessments of various stakeholders. We refer to these as exogenous factors as they are largely beyond the control of both those conducting performance management regimes and those being monitored. It is important to understand that we are not referring to exogenous influences on the performance itself – but rather to exogenous factors that may change the size of the gap between performance on paper and performance being attained or perceived by specific stakeholders. Examples of key exogenous influences on the performance paradox include specific events, changes in the performance of key associates, changes to the policy, and definitional drift.

The occurrence of certain events is likely to have a large bearing on both actual performance *and* the performance paradox of health organizations. For instance, health epidemics will increase the demand for medical services that may exacerbate gaps between perceived capacity and capacity as measured by performance indicators. Moreover, a sudden increase in patient numbers due to natural disasters would make it difficult for end-users of performance management data to reliably interpret time trends. Similarly, increases to population, especially non-organic growth through migration, may exacerbate any discordance between the synecdochical conceptions of responsiveness in treatment, and metrics designed to measure responsiveness. For instance, language

difficulties and complex and unusual pathologies may consume additional resources and create an impression in performance metrics of reduced efficiency that is due to the circumstances of the patient rather than the actions of the health staff.

Changes to the performance of key associates can also contribute to perceptions of the performance paradox. For example, new treatments for disease, arising from the actions of medical training institutions and pharmaceutical companies, could alter the perception of the gap between actual performance and performance on paper. This could arise because new treatments might increase the cost of serving patients and be erroneously perceived as a decrease in technical efficiency (when it is, in fact, a change in dynamic efficiency, which is rarely measured). The actions of other close associates– such as ambulance services and allied health professionals – could also alter the size of the performance paradox. For instance, if more ambulances are staffed with higher trained paramedics then it is possible that patients might receive more extensive treatment on route to DHBs. This might act to reduce the treatment time at the DHB and thus give a better impression of technical efficiency than might be strictly warranted.

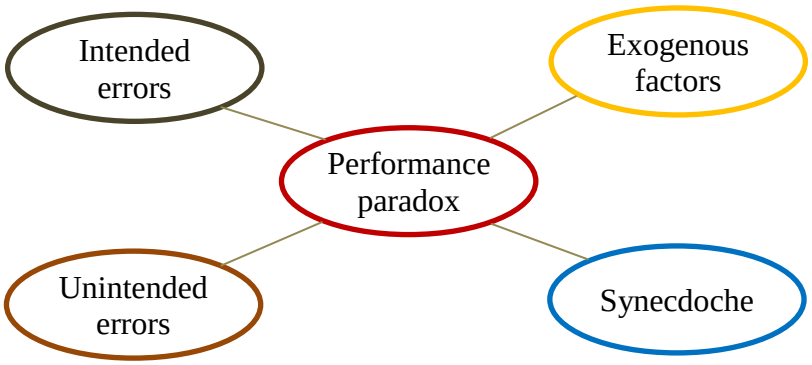
Changes to public policy might also be expected to have an effect on the performance paradox. For instance, mandating new immunization programs, providing stronger incentives, or making immunization programs compulsory could increase the volume of services which might exacerbate existing discrepancies between recorded and actual performance. Similarly, changes to funding models might also elicit behaviors (such as altered depreciation accrual practice or reduced inventories of medical supplies), that might exacerbate gaps between perceived performance and performance on paper.

Finally, definitional drift (which has been identified in the scholarly literature, but not classified as an exogenous cause of the performance paradox (see, Bird *et al.*, 2005)) can change the recording of performance and hence alter the performance paradox. For example, how wait times are defined will have a bearing on how they are recorded and presented in the performance measurement regime. If wait times change from, say, arrival

at hospital (including arrival in an ambulance) to arrival at the emergency room then this will alter end-users' perceptions of trends over time. The need to carefully define key inputs to the performance monitoring regime and exert appropriate audit effort has long been argued necessary to prevent this kind of exogenous influence on the performance paradox (Bird *et al.*, 2005).

Figure 13 illustrates the performance paradox after the addition of the previously neglected exogenous effects.

Figure 13. Performance paradox and the influence of exogenous factors



The scholarly literature provides a number of examples regarding the existence and consequences of the performance paradox (Bird *et al.*, 2005; VanThiel and Leeuw, 2002). For instance, Smith (1995) identifies eight unintended consequences in the implementation of performance indicators in the UK public services, including tunnel vision, sub-optimization, myopia, measure fixation, misrepresentation, misinterpretation, gaming, and ossification. A typical example given in the English hospitals was that the waiting time for surgery is counted from the day patients are put on the waiting list by a consultant surgeon. In responding to this target, surgeons can improve their performance by delaying consultation with the patients or delaying placing the patients on the waiting list. In another analysis of governance by targets applied in English public healthcare systems, Bevan and Hood (2006) observed problems of synecdoche, measurement errors, unintended consequences, and gaming (e.g. output distortions). For example, the

publication of mortality rate might lead to the possibility that the surgeon avoids operating on high-risk cases. Similarly, in response to the targets for the star rating system applied at NHS England, data were adjusted to meet the ambulance response-time targets - within 8 minutes, in which times between 8 and 9 minutes were reclassified to be less than 8 minutes. In addition, patients were also categorized at a lower urgency level to avoid strict time response benchmark (Hood, 2006).

In a different context, Dooren *et al.* (2015) use the term "performance targets paradox" to describe the situation in which an indicator is gradually unable to single out the good and bad performers since the organizations will adjust their performance to the targets. In other words, while under-performers try to hit the target by functional or dysfunctional ways, those already exceeding the targets might reduce their performance and thus bring about what has been termed the 'threshold effect'.

In the case of the New Zealand public sector, Allen and Eppel (2019) investigate the implementation of the Performance Improvement Framework (PIF)[11] and found evidence of persistence of gaming, synecdoche, and reputation-protecting behavior. As for healthcare services, research has been done to investigate the delivery of targets or the impacts of a specific target. For instance, studies have been conducted to examine whether emergency targets do in fact reduce the length of stay at emergency departments (Ardagh, 2015; Tenbensel *et al.,* 2017), as well as the impacts of the emergency target on patient outcomes (P. Jones *et al.,* 2017). Similarly, Blackett *et al*. (2014) investigated the impact of the waiting target for elective surgery to estimate the number of patients who were declined for surgery at two hospitals over one year and found that 25% of all patients with hip or knee osteoarthritis were declined due to insufficient capacity. In a different study Willing (2016) investigated the immunization health target and suggested that health

[11] PIF is "an analytical framework and a change management process to which is used to measure and lift the performance of the state services to deliver outstanding results for New Zealanders and help senior leaders lift the performance of the agencies they lead" (Allen *et al*., 2017, p. viii)

targets might be an effective tool if more attention was paid to the potential for performance paradox at the outset.

We now proceed to an explication of our methodology for assessing the performance paradox that operates at DHBs in New Zealand, Australian hospitals, and Vietnamese hospitals.

3.3 Methodology

Initially, we conducted a content analysis using semi-structured interviews to investigate the manifestations of performance paradox in the context of the New Zealand healthcare sector. We first developed a list of open-ended questions relating to the key aspects of performance measurement and performance management at New Zealand District Health Boards such as the scope of activities to be measured, target setting, and the process of collecting data, analyzing, and reporting. We then classified these questions into sub-themes according to the conceptual framework including synecdoche, unintended errors, intended errors, exogenous factors, using performance information, and additional questions (see Appendix 2 for the detailed questions).

Following this, we then selected the sites and participants for the interviews by employing a relevance sampling method Krippendorff (2004). There are 20 DHBs in New Zealand, providing healthcare services for a total population (in 2016) of 4.7 million ranging from 33 thousand to 600 thousand per DHB (Ministry of Health, 2018). The population is generally characterized as being multi-ethnic, in which Maori typically experience relatively poorer health outcomes such as lower life expectancy (The Ministry of Health, 2014). DHBs consume around three-quarters of public health funding which is allocated based on a Population-Based Funding Formula (PBFF)[12]. Taking these factors into account, and incorporating information from recent reports issued by the Treasury on the

[12] This formula takes into account the number of people served along with other factors such as age, gender, ethnicity, population density, deprivation, and tertiary cost structures to ensure the equality in funding allocation.

performance of DHBs (The Treasury, 2016; 2017), we focus on DHBs that differed in population characteristics, financial status and the quality of tracking hospital productivity. As a result, two DHBs were purposively selected. One DHB has a large population with a low proportion of Maori and consistent financial performance; the other serves a medium-sized population with a higher Maori population, relatively steady financial performance, and a good productivity track record on hospital services. Moreover, as the research principally concerned managerial perspectives, senior managers responsible for performance management were selected rather than professional practitioners. We sent the list of questions (via email) to the potential participants in advance and asked for their willingness to be an interviewee or recommendation for other people that might be more suitable (after examining the questions, some potential interviewees indicated that they were not the most appropriate person to be interviewed).

In order to augment our data, we also conducted a similar approach to select interviewees from Australian hospitals and Vietnamese hospitals. More specifically, two hospitals were chosen from a total of twenty-two hospitals (including private hospitals) operating at Da Nang city, Vietnam (Health Department of Da Nang, 2020). Moreover, from more than 220 public hospitals and health services in New South Wales, Australia (Australian Institute of Health and Welfare, 2018b), a supplementary interview with a typical small-scale rural hospital was undertaken[13].

The interviews were separately conducted at the office of each interviewee, audio recorded, and later transcribed (the transcripts in Vietnamese were translated to English by the authors). The interview transcripts were then imported to the Nvivo software package which facilitates the process of coding and categorizing themes (see Miles and Huberman, 1994).

[13] Whilst additional interviews were initially planned, the restrictions imposed in Australia to contain the COVID-19 epidemic (limiting entry into hospitals and restricting non-urgent meetings to a maximum of 30 minutes) and the increased stress placed on hospitals in Australia which limited availability of management and precluded additional interviews from being conducted.

3.4 Results

3.4.1 An overview of interviews

In total, we interviewed five healthcare organizations. The summary of the interviews is as follows:

Table 5. Background information on the interviews

	Healthcare organization/country	Position	Participant	Place	Date	Duration
1	DHB/ New Zealand	Chief Financial Officer	1	On-site	16th December 2019	1h:04m:32s
2	DHB / New Zealand	Chief Data Officer	1	On-site	17th December 2019	1h:04m:54s
3	Hospital / Australia	Nurse manager	1	On-site	16th March 2020	0h:24m:04s
4	Hospital / Vietnam	Head of Quality Management Department	1	On-site	05th February 2020	1h:34m:39s
5	Hospital / Vietnam	Head of Quality Management Department	1	On-site	06th February 2020	0h:32m:34s

While the participants come from different divisions and countries, their main concerns revolve around the performance measurement as illustrated in the following word frequency cloud[14].

Figure 14. The one hundred most frequent words used in the interviews

[14] The word cloud was generated by Nvivo tool based on the content of interview transcripts.

3.4.2 Manifestation of performance paradox

Synecdoche

Performance of District Health Boards is principally assessed with reference to health targets set by the Ministry of Health. In addition to these Ministry targets, each DHB also internally develops indicators that respond to their specific priorities. Generally, the indicators are clearly defined and understandable. Most of these internal metrics are developed by senior management or leadership teams. Although there is a wide range of indicators, they can be classified into process measures (e.g. patient flow, the timing of discharge length of stay in the emergency department), output measures (e.g. case-weighted discharges, elective surgery volume), and outcome measures (e.g. patient safety). Because it is not generally possible to measure all aspects of complex organizations like DHBs, performance monitoring architects generally implicitly invoke the trope of synecdoche and assert that the part which is measured is representative of the whole. However, it seems that the measures which have been used in health performance monitoring regimes mainly focus on outputs (neglecting outcomes) and thus create a rather large synecdochical gap. Evidence of this could be garnered from the interviews, with general concerns about the sufficiency of the measures employed:

"There's a good chance a lot of the data understates the complexity. I think there probably is not enough of a view at the center around outcomes, the quality of what we deliver. Not just the measures like infection rates, but actually how services compare, their mortality rates, their outcomes. I don't think there is enough of an overview on the outcomes by specialty nationally" (DHB 1)

"It is less specific around quality, infection rates, readmissions, a lot of other things that are related to the quality of care. It is qualitative analysis rather than going by the numbers… And [whilst] people respond to numbers, an equal amount of people respond to

pictures, or to stories. Having a couple of those stories encapsulated in the patient feedback provides for a much stronger reason to change, than a number that might be red, because someone set a target" (DHB 2)

In addition, it was suggested that the Ministry of Health should devise additional metrics or alter the current indicators to make more fulsome comparisons of performance among DHBs:

"We share and have our own peer reviews, and there is a national dataset that we can get information from. But it strikes us that the Ministry does not really look at that, maybe they have other focuses" (DHB 1).

However, despite the criticism, this sentiment may already be shared by external agencies or policymakers, given the recent efforts by the Ministry of Health to develop a new set of targets focusing on population health outcomes in addition to the existing suite of measures already in place (Ministry of Health, 2019).

Similar to New Zealand DHBs, although performance indicators seem to cover key aspects of hospital activities in Vietnam, there exist indicators that might not be realistic, applicable, or could not capture the complexity, which requires adjusting for other factors before making any judgments. The following two quotations reflect well the sentiments of the interviewees:

"Most of the current set of criteria reflects the hospital performance pretty well; many criteria reflect the reality. However, we found that the criterion for patient satisfaction is not very realistic. Because it is not easy to accurately reflect the patients' satisfaction when doing the surveys." (VN1)

"We realize that there are points that need to be adjusted or it takes time to achieve. For example, it is difficult to assess the performance of an individual. Whether to judge by the number of operations or by the severity of surgeries? A surgeon might

perform five operations per day, but one operation per day will be different. […]. Therefore, how to do such evaluation is problematic". (VN1)

In addition, whilst the average score could potentially be used to compare the quality of hospitals in Vietnam, this approach appears problematic since the same set of criteria is applied to all types of hospitals with different functions, scales, and complexities (VN1).

Intended errors

On the basis of the argument that humans are rational utility maximizers, it is reasonable to expect that in response to performance imperatives, people might react by seeking to manipulate or game the system (Talbot, 2010). Intended errors result from deliberate actions of manipulating measurements or manipulating the outputs (e.g. underrepresentation, misinterpretation). The interviewees were asked about the likelihood that manipulation occurs at their organizations or other organizations with the intent to create a more favorable picture of performance than might be strictly warranted. A common characteristic in the performance frameworks under investigation was that there were no rewards or penalties for delivering health targets or quality indicators, suggesting that there was less likelihood for systematic deliberate distortion of performance information. Also, although the participants did not clearly indicate specific examples of gaming or manipulation, they did seem to concede that some isolated cases of manipulation were inevitable.

"I think whatever [measure] you choose there's always a way people will look at to try and manipulate it. … There have just been anecdotal stories about people saying this starts here or stops here …with the waiting times, we've had weird things where people have ended up not on waiting lists because someone's kept information, we call it hidden waiting lists, and that's been a problem in the past" (DHB 1)

"I can't really comment on any others but there's always been the noise around emergency targets. What you classify as an emergency becomes a topic of concern, so "is short stay emergency or not?" … If all the patients that come into through the emergency department door, but need a bit more diagnostics, and then end up in a different unit that then get discounted because the system allows you to do that, is that fair or not?" (DHB 2)

In relation to the potential for manipulation, the two interviewees in Vietnam asserted that their hospitals are not motivated to distort the performance data. However, a typical example of gaming by individuals was provided as follows:

"We have to check whether staff execute their missions properly. For example, a nurse was assigned to work on nightshift but he/she went to sleep rather than cared for the patients. However, in the next morning, he/she can still provide a full record (fabricated data). Another nurse worked all night, did not rest, obeyed his/her doctor's orders but forgot to record the information; therefore there are no reports in the next morning. [..]"

Unintended errors

Interviewees noted that data was systematically collected (through a national minimum dataset or other programs), and also subject to audit, which should provide reasonable assurance of its accuracy. Nevertheless, interviewees did identify some systematic problems that might result in unintended errors that could contribute to the performance paradox:

"There have been examples I know in other hospitals where they've [asked] people "what percentage do you treat [that] have this co-morbidity?" and the doctors have said "about 15 percent" and they've [replied] "based on your coding no one does, because you're not coding it". …We are anecdotally seeing older, sicker people, and people with more co-morbidities, but our case weights are not quite telling us that, and so I think there

is something to do with the system actually measuring what's really going on with the population." (DHB1).

"…I think the systems are probably too relaxed, I find that the data entry systems, the transactional systems allow for users to not fill in a lot of data, which period, for instance, that just creates bad data" (DHB 2).

In addition, the interviewees suggested that the reliability of information might be influenced by the process of data collection (especially when data are manually collected), the sample size, as well as the diligence of the collectors.

"…as I said there was an issue some time ago with the transfer of care documentation that was coming through from the data… we set a time limit with accepting patients from an ambulance service, and being a small facility we only had two staff on night duty, so unless the computer is actually accessed, it can be entered retrospectively, so it doesn't actually look like the staff are seeing the patient upon presentation which is not an accurate reflection because staff go immediately into providing clinical care… and the computer can sometimes not be turned on" (AUS)

Moreover, while interviewees did not clearly articulate the problem of unintended error, both the absence of clarity in the indicators' setting and insufficient training were cited to be likely to lead to undesirable and unintended distortions:

"In terms of understandability, it is not really easy to understand. In my opinion, firstly, understandability means having a consistent understanding suggesting that people should have the same understanding of a definition or a problem. However, the reality now is that with the same issue one might think it is A, the other thinks it is B. There are still contents that we do not completely understand but still have to do; even it is impossible. The reason is that there has been no training so we did not know what to do. For instance, it is required to write a process or a plan; If not trained how do they

write? Employees are not properly trained in quality assessment. Therefore, it is not easy to understand and certainly, it is not easy to implement." (VN1)

Exogenous effects

We suppose that the exogenous factors such as changes in target setting (definitional drift), changes in policies, or specific events might also contribute to the performance paradox. With respect to the New Zealand health targets, the interviewees identified that these metrics are quite stable even with the changes in government. However, since the population of New Zealand is quite small, they suggested that any epidemics or incidents might have significant impacts on data collected and hence could contribute to the perception of a performance paradox. For example, a measles epidemic or natural disaster (such as the 2019 White Island eruption):

"To have a whole cohort of people puts a strain on the whole system, and things like electives get displaced because you are using the same staff and resources, they need the theatre space" (DHB 1).

More generally, the magnitude of the event on the performance paradox was also highlighted:

"For this matter, I would like to elaborate that it depends on the magnitude of the events. For example, with the same epidemic, a disaster, or an event but on a national scale is different from the international level. If we cater to a sporting event, for example, the international or city scale will have different impacts. Naturally, at any level it will also affect the ordinary operation of the hospital while hospital resources are limited and the hospital is overload; also many staff absent due to training, sick leave, maternity …Therefore, we do not always have sufficient resources to meet daily activities and provide services for such events." (VN1)

In addition, it was suggested also that when politicians and media scrutinized the daily

activities of DHBs that also exerted an exogenous effect on the public perception of performance paradox (DHB 2). Notably, in performance assessment, these external influences are reflected in the reports and recognized by the authorities such as the Ministry of Health or the Department of Health.

3.5 Discussion and conclusion

This study examines the performance paradox through an investigation of the implementation of performance measurement at New Zealand District Health Boards, expanding to the cases of Australian and Vietnamese hospitals. In addition to the factors that are well examined in the extant literature (intended errors, unintended errors, and synecdoche), we attempt to identify the impacts of exogenous factors (e.g. changes in policies, special events) on the disparity between actual performance and performance reported.

Consistent with the corpus of scholarly literature on unintended consequences of target-based management in the healthcare sector, our results obtained from interviews with senior managers at two DHBs and three other hospitals in Vietnam and Australia confirm the existence of the performance paradox. Synecdochical gap seems to have contributed to some of the performance paradox given that the current design of targets set by the Ministry of Health do not completely capture the key functions and expectations of DHBs. On the other hand, it was difficult to evaluate the magnitude of intended and unintended errors through interviews alone – although it is clear from our analysis that both of these sources are likely to have contributed to the performance paradox. Our major contribution to the literature was to identify the salience of exogenous factors to the performance paradox – specifically, we noted how the perception and interpretation of performance of DHBs against the targets might be affected by the epidemic and natural disasters, media, and political scrutiny.

As well as identifying shortcomings of the existing performance management systems in place for New Zealand DHBs, the interviewees also suggested solutions to mitigate perverse outcomes and to improve the effectiveness of the current measurement framework. Key recommendations include the need for more patient-focused metrics to capture the quality of care and reflect the voices of patients as expressed in this comment:

"If anywhere I would want more measures, it's around the outcomes. The best of that was the verbal commentary, yes we can score on a scale, but the commentary was actually more insightful because if you were going to do a service improvement, you were going to go into the portal and look at the commentary from people who have experienced the survey to hear what's working and what is not" (DHB 2).

This sort of sentiment is consistent with the extensive work of De Bruijn on the importance of dialogue to complement numbers (see, for example, De Bruijn and Van Helden, 2006).

In addition, interviewees in New Zealand also noted that it might be helpful if a balanced measure (e.g. patient experience, health outcomes) was employed to counter potential manipulation and complement the quantified metrics. Moreover, it was suggested that data should be linked to the clinical record because it was felt that this data source has the highest level of integrity. Interviewees were adamant that whilst performance metrics and associated targets provide some information value, it was by itself insufficient to evaluate the performance of a DHB as a whole or for benchmarking against peers.

Specifically, interviewees were united in noting the need for additional metrics. Indeed, given the acknowledgment that efficiency is crucial for a high-performance healthcare system, it seems surprising that investment and development of succinct efficiency measures in New Zealand have been largely overlooked (Knopf, 2017). We, therefore,

recommend that an additional measure, such as a comprehensive efficiency score, is employed in the future to mitigate the fragmentation of health targets.

As a limitation of this study, it should be noted that the interviews were conducted at only a small number of healthcare organizations in New Zealand, Australia, and Vietnam. It might be fruitful to extend the study to additional units in the future and also include people in charge at the policy-maker levels such as the Ministry of Health or the Department of Health at local governments.

Appendix 1: New Zealand national health targets

Target	Description
Shorter stays in emergency departments (Q)	95% of patients will be admitted, discharged, or transferred from an emergency department within six hours.
Improved access to elective surgery (V)	The national volume of elective surgery will be increased by at least 4,000 discharges per year
Faster cancer treatment (T)	Before 1st October 2014: All patients, ready-for-treatment, will wait less than four weeks for radiotherapy or chemotherapy Since 2014: 85% of patients referred with a high suspicion of cancer receive their first cancer treatment within 62 days. (The target was increased to 90% from July 2017)
Increased immunization (C)	95% of infants aged eight months will have completed their primary course of immunization on time (the threshold set for 2013 was 90%)
Raising healthy kids (Q)	By December 2017, 95% of obese children identified in the B4 School Check programme will be offered a referral to a health professional (Replaced for "Heart and Diabetes Checks" target since 2016)
Better help for smokers to quit (Q)	Secondary Care: 95% of smoking patients are offered advice and support to quit smoking (This target has been abolished since 2016) Primary Care: 90% seen in primary care are offered advice and support to quit smoking Maternity Care (since 2015): 90% of newly registered pregnant women provided with advice to quit smoking.

Sources: Ministry of Health (2019). The letters in brackets (C, Q, V, and T) denote the type of measure for each target, in which: C-coverage, Q- Quality, V-Volume, and T-Timeliness.

Appendix 2: Semi-structured interview questions for New Zeland DHBs[15]

1. Introductory questions

- The information collected will be used for doing research only. Respondent's information is anonymous and confidential. Please ask the interviewer to clarify if the respondent does not understand any questions.

- Could you please first state your position and responsibility?

- How long have you been working in the healthcare sector?

2. Main questions

* Synecdoche

Q1. What main activities are being measured? On what basis do DHBs identify which activities need to be measured?

Q2. Is there a large gap between the health targets/current performance indicators and what you think are the most important things your DHB does?

Can you provide examples of important metrics that should be probably included that they are not? For these examples, why do you consider this very important?

Q3. What are the key metrics/indicators to measure the performance of DHBs? Who sets the indicators? Is there an indicator that captures the whole activity of your DHB to compare with other DHBs?

Q4. Are these indicators clearly defined and easy to understand? Are staff involved in the process of designing performance indicators?

[15] The questions are slightly modified accordingly to the context of interviews in Vietnam and Australia.

Q5. There might be a performance paradox when performance indicators are not strongly correlated with performance. What is your assessment of this manifestation at your organization?

* Unintended error

Q6. How is the data collected for performance measurement? What is your assessment of the adequacy, timeliness, and reliability of data collected? What are the main difficulties in collecting data for performance measurement?

- *Regarding health targets, which metrics are most likely to have measurement error?*
- *Can you provide examples of how measurement error might be accidentally introduced?*
- *What could be done to reduce measurement error?*
- *Is there an audit process for any of the data to provide assurance?*

Q7. What extent do you think that the performance information can clearly explain the results of performance (accountability)?

Q8. How is the information on performance useful in improving or redesigning the performance indicators? How often are the results of performance indicators discussed among the staff, top managers, or between managers and staff?

Q9. What are the possible unintended (negative) effects of performance measurement at your DHB? Why?

* Deliberate performance Paradox

Q10. At what extent do you think the chances for manipulating the performance measurement at DHBs occur?

- *In relation to heath targets, do you believe there are many opportunities to manipulate the metrics in order to provide a more favorable picture of performance?*
- *Can you provide us with an example of how this might be done?*
- *What procedures might be implemented to reduce intended error?*

* Exogenous factors

Q11. Is there any (5 years) policy change that may have made the performance paradox worse?

Have any of the key definitions associated with metrics changed in recent years (definition drift)

Q12. Is there a document listing precise definitions and guidance for calculating the metrics?

Q13. Has there been any health epidemic/crisis that made the gap between actual performance and performance on paper worse?

* Using performance information

Q14. How frequently do DHBs analyze and report the results of performance measurement? Who do you report to?

Q15. Who is the primary audience for performance information produced by DHBs (external organizations or internal divisions)? Why? Please provide more detailed examples.

Q16. How is the performance information used for steering and controlling the operation at your DHB?

Q17. What are the incentives for performance measurement at DHBs? In your opinion, who plays a vital role in the success of performance measurement? What are the

prerequisite conditions for adopting and implementing performance measures in your organization?

Q18. What are the main obstacles in the implementation of performance measurement? How to resolve these issues?

Q19. Under the circumstances of aging populations, rising healthcare costs, and restrained budget, in what aspects do you think performance measurement should be designed to best utilize the resources available?

* Additional questions

Q20. In a review by the Controller and Auditor-General in 2016, health assets had the lowest condition ratings in the public sector. Based on the financial data from 2012 to 2016, we observe that actual capital expenditure was only 65% of the forecast, on average.

Can you provide us possible reasons why DHBs did not sufficiently invest in fixed assets as budgeted? Is there any possibility that DHBs more emphasized on the staff rather than fixed assets?

Q21. It appears that staff productivity at DHBs has been increased. In your opinion, what are the underlying factors that contributed to this increase?

CHAPTER 4: EVALUATING EFFICIENCY OF ENGLISH ACUTE FOUNDATION TRUSTS UNDER SYSTEM REFORM: A TWO-STAGE DEA APPROACH

The English healthcare sector underwent extensive system reform over the period from 2010 to 2015, aimed principally at improving technical efficiency. This study examines the effect of the reforms on foundation trusts in England with particular emphasis on technical efficiency. By employing Data Envelopment Analysis and a second-stage regression, we found evidence of an overall improvement in efficiency, notwithstanding some fluctuations. Specifically, we found that bed utilization had a positive and statistically significant association with the efficiency of acute foundation trusts; suggesting that better management of patient flows and bed utilization might be expected to improve hospital efficiency. We also found evidence to suggest that efficiency might also be improved through better management of staff numbers, optimizing liquidity, and better utilization of assets such as buildings and information technology.

4.1 Introduction

The National Health Service (NHS) England is the largest part of the United Kingdom (UK) healthcare system and employs around 1.2 million people to serve a population of 54.3 million citizens. Since its establishment in 1948, three pillar principles have guided health service delivery: (i) “meet the needs of everyone”, (ii) “free at the point of delivery”, and (iii) “based on clinical need, not ability to pay” (NHS Choice, 2016). The main revenue source (98.8%) of the NHS budget is derived from taxation and National Insurance (TheKing’sFund, 2017). As a result of the economic downturn in 2007 and 2008, and the subsequent period of austerity, the NHS England was exposed to considerable financial pressure thus prompting widespread reform.

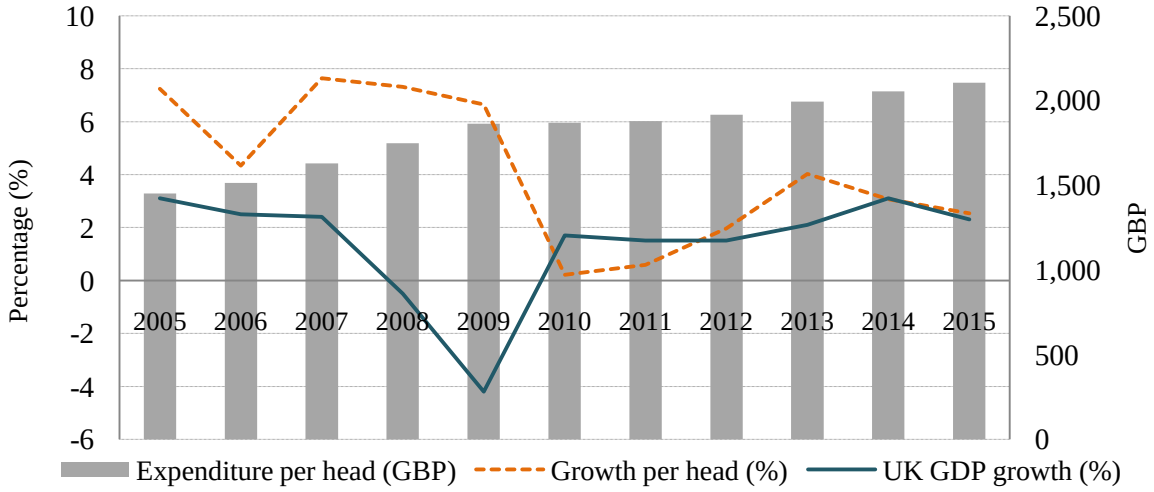
Reform of the English healthcare system under the Health and Social Care Act 2012 was

considered “the most wide-ranging and controversial restructure” (Powell, 2016). As explained by Department of Health (2012b), the three main reasons for enacting the Act were to address the pressures arising from increased demand and treatment cost; to instigate improvements designed to avoid falling behind comparable countries; and to relieve pressure on public finances. On the basis of these three rationales, five major reforms were conducted: (i) formation of a new clinical commissioning group where clinicians’ role would be enhanced; (ii) providing greater choice to patients; (iii) enabling providers to deliver quality services; (iv) increasing accountability at both national and local level; and, (v) streamlining the arms-length healthcare bodies. A key driver of the policies was a belief that a quasi-market model and decentralization would yield greater efficiencies.

Alongside organizational changes, a striking feature during the period from 2010 to 2015 was the general atmosphere of budget austerity in the United Kingdom. As spending on health was the second largest component of the national budget - accounting for about 23% of all spending on public services in England over the 2009-2015 period (HM Treasury, 2017) – savings from this area were deemed to be an important part of the national budget repair effort. Indeed, the savings target set for the NHS in England by 2014 was £15 to £20 billion (Roberts *et al.*, 2012). Leading up to this time public health expenditure in real terms had grown at an average of 3.7%, but from 2009 to 2013 growth was reduced to just 0.7% per annum (Lanfond, 2015). In per capita terms, the average rate of healthcare expenditure per person between 2004 and 2009 had grown at 6.6% per annum, but this rate was reduced to just 2.4% per annum for the period from 2010 to 2015 (see Figure 15). Moreover, from 2011 onwards, the tariff (price paid to hospitals for services) was reduced in an attempt to realize the above savings target by 2014. The adjustment to the tariff was made in advance and was calculated by subtracting imposed efficiency dividends from the total increase attributable to pay and price inflation. As a

result, providers faced a nett cut of 7.7% between 2010 and 2015; after accounting for the increases in the total efficiency of 23% and price inflation of 15.3% (see Table 6).

Figure 15. Healthcare expenditure



Source: Authors' calculation using Public Expenditure Statistical Analyses 2010, 2013, and 2017 (HM Treasury, 2017).

Table 6. Changes in the tariff (2009 – 2015)

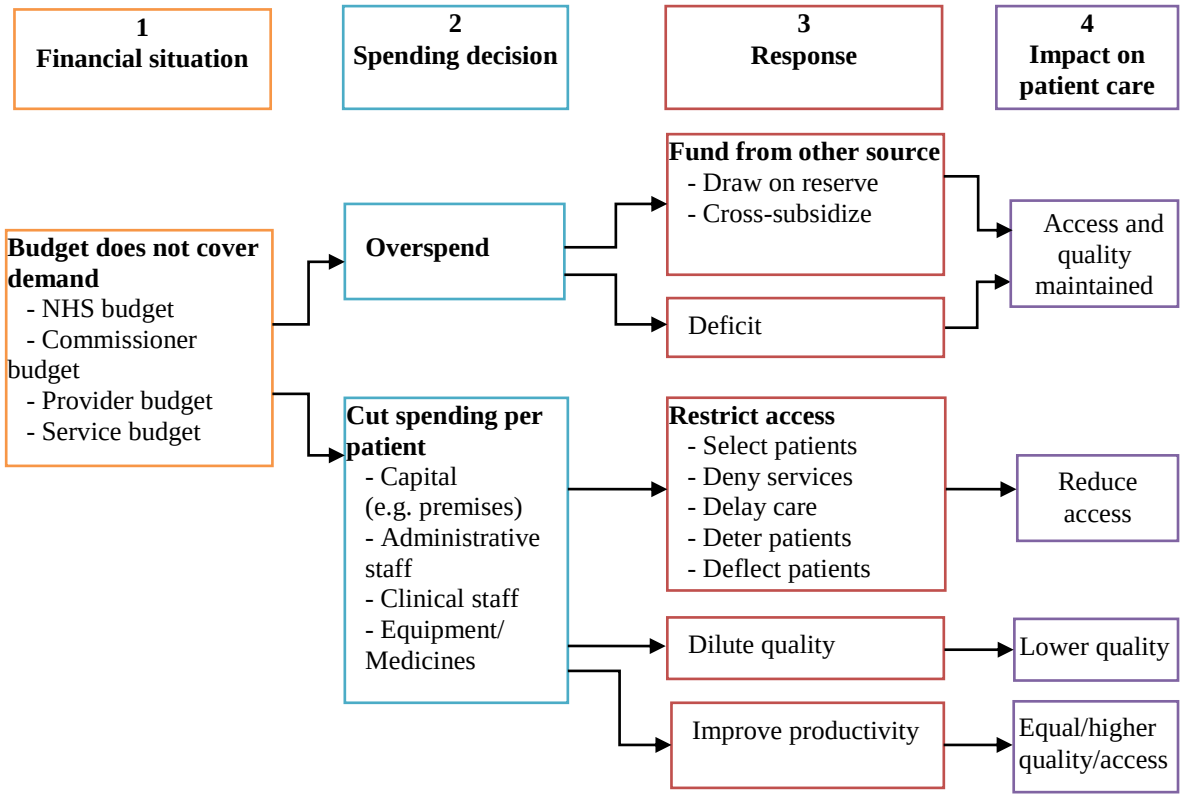
Tariff year	Pay and price inflation (%)	Efficiency requirement (%)	Net tariff uplift (%)
2009-10	4.7	-3.0	1.7
2010-11	3.5	-3.5	0.0
2011-12	2.5	-4.0	-1.5
2012-13	2.2	-4.0	-1.8
2013-14	2.7	-4.0	-1.3
2014-15	2.5	-4.0	-1.5
2015-16	1.9	-3.5	-1.6

Source: Department of Health (2012a); Marshall *et al.* (2014); Monitor (2015a, 2015b)

In addition to the policy-based changes, hospital services also encountered challenges such as increase in demand owing to growth in the size and aging of the population, increases in morbidity, and rising health costs for equipment and materials (Crawford and Stoye, 2015; Licchetta and Stelmach, 2016). Reactions by hospital managers to the changed

budget and policy environment were varied but one clear response was to improve productivity (see Figure 16). One way that this might occur would be, similar to other countries, for the number of staff in English hospitals to gradually increase but the total number of hospital beds (a frequently used indicator of hospital capacity) to decline. According to Baker (2017), "[s]ince 2011, the number of beds available overnight has fallen by 7,400 (a fall of 5.4%)", which seems to support this contention. Indeed, over the period 2009 to 2016, in comparison with other European Union (EU) countries, the UK had one of the lowest number of hospital beds per capita with around 3 per 1,000 population, lower than the EU average level (5 beds) and far lower than the highest (Germany) which had around 8.3 beds per 1,000 population (OECD, 2017). The reduction in the number of beds might also be associated with a shorter length of patients' stay, which is typically considered as a sign of efficiency (OECD/EU, 2016).

Figure 16. Possible responses to the funding pressures



Source: Adapted from Robertson *et al.* (2017)

Thus, in responding to the economic and policy environment, hospitals were expected to use resources more efficiently to meet pressure from both increasing demand and higher quality expectations. This prompts at least three salient research questions. First, did the expected improvement to technical efficiency indicated by policy changes and budgetary constraints indeed occur over the period 2009 to 2016? Second, did reduction in beds contributed to any efficiency gained? Third, are there factors that explain any variability in efficiency across the hospitals which make up the foundation trusts in England?

To answer our three research questions we employed DEA on recent longitudinal data to estimate the technical efficiency of NHS foundation trusts. Although trusts and foundation trusts are both publicly owned hospitals, we focused on the latter as they make up about 66% of the total NHS hospitals and have more autonomy in governance and financial freedom than their peers. Foundation trusts were established as a new type of NHS hospital in 2004 and must meet strict criteria (e.g. well managed, legally constituted, financially viable) to attain the foundation trust status (Verzulli *et al.*, 2018). Thus, foundation trusts and trusts are different in terms of financial and management perspectives and it might not be appropriate to combine them into one group to make a comparison. As of 31 March 2017, based on their principal services, 157 foundation trusts are categorized into five different types including acute (85), mental health (44) specialist (17), ambulance (5), and remaining community (6) (NHS Improvement 2017b)

DEA has been employed a number of times in relation to analyses of healthcare sectors, although there is a paucity of DEA research on UK hospitals (Valdmanis *et al.*, 2016). Specifically, of 262 DEA papers reviewed over the period from 2005 to 2016, only one investigated UK hospitals (Kohl *et al.*, 2018). We found that most DEA literature on the UK hospitals was published prior to 2006 (typically examining periods before 2000) and used crude output proxies (without adjusting for case-mix). Moreover, few of the extant

works applied techniques such as second-stage analysis required to identify the determinants of efficiency (DEA measures efficiency, but second-stage analysis is important to identify the determinants of efficiency which is an entirely different matter). To remedy this gap, especially with reference to the absence of *UK* work on the determinants of efficiency, we used two-stage analysis with panel data regression to uncover factors associated with differences in efficiency across providers.

The remainder of the paper is organized as follows: In the next section, we summarize recent studies on evaluating hospital efficiency with DEA models. Following this, we explain our model specification and the reasons for selecting inputs, outputs, explanatory factors and the sources of data obtained. In the subsequent Result section, we present our findings and discuss the efficiency scores and the regression outcomes. We conclude with a brief discussion of the public policy findings suggested by our work.

4.2 Literature review

4.2.1 Application of DEA in measuring hospital efficiency

Efficiency is a key concept in economics, normally referring to the ability to maximize the outputs produced within a given level of inputs or alternatively minimize the uses of inputs to produce a chosen level of outputs. In the extant literature, there are two distinguished approaches to evaluate efficiency, Stochastic Frontier Analysis (SFA) and Data Envelopment Analysis (DEA). The former is a parametric method developed by Aigner, Lovell, and Schmidt in 1977 which assumes that the residuals are decomposed into inefficiency and random errors and also that these two components are distributed differently. The latter is a deterministic method developed by Charnes, Cooper, and Rhodes in 1978 to measure the relative technical efficiency by using linear programming to identify an efficient frontier from the piece-wise linear combination of best practice units. Accordingly, in DEA inefficient units are enveloped by the efficiency frontier and

the technical inefficiency of each unit is measured by the distance of each unit from the frontier (Coelli *et al.*, 2005). There are inherent strengths and drawbacks for each methodology. SFA accounts for noise and can be used to conduct tests of hypothesis but requires specification of a functional form and assumptions about the distribution of the inefficiency term (Coelli *et al.*, 2005). On the other hand, DEA does not require a functional form or assumptions about the inefficiency distributional properties but is sometimes criticized for its deterministic nature and the lack of tests to verify the most appropriate model specification (Ozcan, 2014).

Since the Charnes, Cooper, and Rhodes (1978) constant returns to scale (CCR) model, later extended by Banker, Charnes, and Cooper (1983) to variable returns to scale (BCC), DEA has been increasingly applied which suggests its general acceptance by academics and practitioners alike, in a wide variety of situations (Chilingerian and Sherman, 2011). Indeed, bibliographic collection of DEA literature over the period from 1978 to 2016 indicates a dramatic growth in publications which employ DEA techniques, of which healthcare is among the five most pervasive fields (Emrouznejad and Yang, 2018). A possible reason for increasing application of DEA could be due to its ability to incorporate multiple inputs and outputs simultaneously with different unit measurements and without any requirement for prior weights or prices (Charnes *et al.*, 1994).

Given that there are different types and levels of healthcare entities (e.g. primary care and secondary care; hospitals and nursing homes; specific services at hospitals), we emphasize DEA studies that examine hospital efficiency. In a summary of 317 publications up to mid-2006 that measure efficiency in healthcare, Hollingsworth (2008) indicated that a majority (80%) of research used DEA analysis and more than half of applications were with respect to hospitals. More specifically, Hollingsworth (2008) detailed that most research used inpatient days or discharges as hospital outputs while staff and capital

employed were the main inputs. Similarly, O'Neill *et al.* (2008) systematically reviewed 79 studies on hospital efficiency using DEA published during the period from 1984 to 2004. They investigated different characteristics (e.g. types of model, choice of variables) to make a comparison between Europe and the US and found that European studies had a tendency to use panel data, and use a fewer number of inputs but a higher number of outputs. As an update to the research conducted by O'Neill *et al.* (2008), a comprehensive review of DEA literature on hospitals was conducted by Kohl *et al.* (2018) including 262 papers from 2005 and 2016, summarized below:

Research topics

Most of the DEA research (100 papers) were motivated by examining the association between efficiency and other salient factors (quality, ownership type, specialization, regulations) while others (99 papers) just simply conduct DEA, which can be classified as "pure DEA efficiency analysis". The next group (48 papers) focused on developing advanced and new DEA models (using the Malmquist index, comparing multifactor efficiency and non-radial super-efficiency). The last category comprised "surveys on the effects of reforms" (36 papers) which mainly made comparisons between efficiency before and after certain reforms to evaluate how the policy had influenced efficiency.

Additional techniques

DEA scores tended to be a starting point with extensive attention paid to further analysis, including Panel Data Analysis with Malmquist Index (47 publications) and Window Analysis (five papers); Bootstrapping was used in 20% of the papers. A particularly important task from a public policy perspective is, as we have noted, to identify the determinants of efficiency.

DEA measures efficiency with variables employed to proxy inputs and outputs. However, to find the determinants of efficiency it is necessary to employ a different suite of

variables which might be expected to explain the change in inputs and outputs (measured in DEA). Regression analysis is often employed for the purposes and was used in about 25% to 30% of the publications. However, the authors noted that no model was superior to or more reliable than the others. They suggested discussing theories and practical evidence with the stakeholders (hospital managers, policymakers, and economists) to find causes of inefficiency and enhance the robustness and the reliability of the results.

4.2.2 DEA literature on UK hospitals

Turning to studies on UK hospitals, we examined research up to two decades prior. Except for a few recent studies, it seems that most were conducted before 2006. Although NHS England is the largest constituent, serving more than 80% of the UK population, Scottish hospitals were examined more frequently. The purposes of these studies also varied, such as comparing the efficiency between small and large hospitals (Mccallion *et al.*, 2000), evaluating the impacts of market reform on changes in efficiency or estimating efficiency changes over time (Ferrari, 2006; Maniadakis and Thanassoulis, 2000; Valdmanis *et al.*, 2016). Other researchers investigated efficiency scores to verify the validity and potential application of DEA in the healthcare sector or comparing DEA results with other approaches (Hollingsworth and Parkin, 2003; Jacobs, 2001). Concerning the variables used to measure efficiency, beds, staff numbers, and costs were common inputs; whilst, inpatients and outpatients were the main outputs employed (see Table 7).

Using a different method, by combining different outputs and inputs with explicit weights and Ordinary Least Square (OLS), Castelli *et al.* (2015) examined 166 English trusts and found that productivity was positively associated with higher bed occupancy and that foundation trusts tended to be less productive than their counterparts. Similarly, Aragon Aragon *et al.* (2017) explored factors influencing the productivity of English trusts in the period 2010-2012 and suggested that trusts were more productive than foundation trusts

with diseconomies of scale in larger trusts.

4.2.3 Determinants of hospital technical efficiency

A number of existing studies (57 studies) in the healthcare sector have combined statistical methodologies and techniques with conventional DEA (Cantor and Poh, 2018). These statistical methodologies included regression models, statistical tests, productivity change analyses, bootstrapping, and correlation analysis although regression analysis was the most common approach. The general objective was to help explain the variation in organizational performance.

Reviewing several recent papers, where DEA has been used to measure efficiency and second-stage analysis have been employed to find the determinants of efficiency, it appears that external and internal drivers can be classified into socio-economic factors, patient and hospital characteristics, and economies of scale and scope:

Socio-economic factors

These included population density, population over 65, youth-unemployment, full-time employment, elderly dependency rate, education, income (household income, income inequality), hospital density, expenditure on health, hospital (private, public), gross domestic product, life expectancy, infant mortality, and competition.

In their investigation of the hospital sector in OECD countries, Varabyova and Schreyögg (2013) found that countries with good health outcomes, higher income inequality and a longer average length of stay tend to be technically inefficient while countries with higher health expenditures per capita were positively correlated with hospital efficiency. Similarly, Kaya Samut and Cafrı (2016) suggest that both GDP and educational attainment have positive links with hospital efficiency. In an Italian hospitals' context, Matranga *et al.* (2014) found that while unemployment rates in young males and the average length of stay had a negative impact on hospital efficiency, improvement in

socio-economic conditions could have a positive impact.

Patient characteristics

These included age structure (older patients), percentage of Medicare and Medicaid patients, length of stay, and bed occupancy ratio.

Czypionka *et al.* (2013) found that patients aged over 80 years have a negative influence on efficiency in Austrian acute hospitals. An inverse relationship between inpatient age and hospital efficiency was also observed for Canadian acute care hospitals (Fixler *et al.*, 2014). For a specific group of US hospitals, Chou *et al.* (2012) and Nedelea and Fannin (2013) found that the percent of Medicaid admissions had a positive and significant effect on technical efficiency. A positive association between occupancy ratio and efficiency were indicated in the cases of Greek hospitals (Kounetas and Papathanassopoulos 2013), Austrian hospitals (Czypionka *et al.,* 2013), and Canadian acute hospitals (Fixler *et al.*, 2014).

Hospital characteristics

These included size/capacity (based on the number of beds), region (rural, urban), ownership (public, private), for-profit and non-profit hospital, teaching and non-teaching, advanced technology adoption, numbers of operation years, degree of specialization (Herfindahl-Hirschman Index).

When comparing hospitals with different ownership types, public hospitals were negatively correlated with efficiency, and private hospitals had a positive association (Czypionka *et al.,* 2013; Kaya Samut and Cafrı, 2016).

Economies of scale and scope

Giancotti et.al (2017) speculated that hospitals' scale and scope might be expected to have a considerable effect on efficiency. In a review of studies over 45 years (1969-2014)

investigating the optimal size of hospitals, they found that while hospitals with 200–300 beds reaped economies of scale, diseconomies of scale occurred above 600 beds. Regarding economies of scope, a study of Portuguese hospitals by Ferreira *et al.* (2018) suggests that, generally, hospitals can exploit economies of scope, however, this is unlikely to happen in larger hospitals (those with more than 6,000 discharges and/or 7,500 medical appointments).

Generally, DEA and its integrated models have been a preferred method in measuring the efficiency of hospitals and explaining the variation in efficiency scores. Also, it should be noted that the number of inputs and outputs, as well as explanatory factors, varies by study depending on research aim and data availability. However, given the wide application in many countries, it appears surprising that there is a scarcity of recent DEA applications on measuring the efficiency of English hospitals.

Table 7. DEA literature on the UK hospitals

Author	Sample	Method	Input	Output	Finding
1. Mccallion, Glass, Jackson, Kerr, and Mckillop (2000)	23 Northern Ireland hospitals (1986 - 1992)	DEA window Analysis and Malmquist Index	• Nursing • Administration • Ancillary • Specialists • Beds	• General surgery • General medicine • Maternity • A&E	• For smaller hospitals, an increase in technical change offset the decline in efficiency change which is due to a decrease in scale efficiency. • Larger hospitals are more efficient in providing healthcare services.
2. Maniadakis and Thanassoulis, (2000)	75 Scottish acute hospital (1991 – 1995)	Malmquist Index	• Doctors (whole time equivalent - WTE) • Nurses (WTE) • Other personnel (WTE) • Number of hospital beds • Cubic meters of hospital buildings (per 100)	• Accident and Emergency attendances • Inpatients (case-mix) • Day cases (case-mix) • Outpatients (case-mix)	• Given the moderate improvement in productivity was achieved, there still scope for further gains. • Productivity change could not attribute to the internal market reform • Productivity gained mainly stemmed from the allocative efficiency.
3. Ferrari (2006)	53 acute Scottish hospitals (1991- 1996)	Malmquist Index	• Total capital charges • Medical staff (Full Time Equivalent – FTE) • Nursing staff (FTE) • Other staff (FTE) • Total number of beds.	• Inpatients, surgery • Inpatients, medical • Inpatients, others; • Outpatients, day cases, and day patients.	• Frontier shift and technical efficiency changed in an opposite direction and unclear trend. • Overall, technical efficiency reduces by one percent suggesting that the introduction of competition for hospital services had no significant effect on efficiency.
4. Valdmanis, Rosko, Mancuso, Tavakoli, and Farrar (2016)	43 general acute Scottish hospitals (2003–2007)	Malmquist Index (with bootstrap procedure) and OLS time-series for trend analysis	• Staffed beds • Doctors (physicians and dentists) • Nurses (including nurse trainee) • Other labors	• Inpatients (case-mix adjusted) • Outpatients and short stay patients	• The authors could not found consistent patterns in technical change and efficiency change. • A continuous decline in technical change implying that technical changes in previous periods are negatively associated with changes in the subsequent period.

4.3 Models selected and data

4.3.1 DEA model specification

DEA employs linear programming techniques to calculate relative technical efficiency scores by optimizing the ratio of weighted sum of outputs to weighted sum of inputs when weights are unknown. The CCR and BCC DEA models involve choices around "orientation" and "returns to scale". O'Neill *et al.* (2008) argue that most studies use an input-oriented model since "hospital managers and policy-makers generally have more control over their inputs than their outputs, and, in a majority of countries, the emphasis is on controlling costs rather than increasing demand for healthcare". Over the period examined, English hospitals were under severe funding pressure with value for money being a key concern, making it likely that managers were motivated to use resources more efficiently to provide a given volume of services. Therefore, an input-oriented model was selected for this study.

The CCR model assumes a constant return to scale (CRS), suggesting that inputs and outputs increase or decrease proportionally or that all DMUs operate at an optimal scale. By contrast, the BCC model relaxes this assumption, allowing for variable returns to scale (VRS) which implies the existence of economies or diseconomies of scale. In the context of healthcare services, hospitals might not always perform at an efficient scale due to the restraints on available resources, legal framework and the market characteristics (Jacobs *et al.*, 2006). More particularly, Kirigia and Asbu (2013) suggest that increasing returns to scale happen when resources are not completely divisible (e.g. medical equipment, operating theatre). On the other hand, a further increase beyond the optimal level might reduce efficiency in management due to growing complexity, difficulties in communication and distraction executing organizational strategies (i.e. decreasing returns to scale). It is expected that hospitals could exhibit variable returns to scale, however,

Chilingerian and Sherman (2011) argue that hospitals are expected to operate at constant returns to scale. More than 50% of studies used CRS in the period from 1984 to 2004 (O'Neill *et al.*, 2008), though recent studies tend to use the VRS model. Since no model has been proven to be superior, in the first stage, we will estimate efficiency under both CRS and VRS assumptions. However, in the two-stage analysis, as some of the explanatory variables explicitly reflect hospital size, scores under CRS have been employed to avoiding potential bias.

Thus, both CRS and VRS DEA modeling (input-oriented) are used in this study to estimate the efficiency of the foundation trusts. However, unlike cross-section DEA analysis, as our study uses longitudinal data (8 years), it is unsuitable to simply compare normal DEA scores between different time periods because they are calculated on different reference groups. Therefore we adopt global inter-temporal DEA as a measure to investigate the changes in technical efficiency. The notion of this method is that all data across time are first pooled into a single DEA analysis, and then the DEA scores are regrouped in each year. However, the DEA scores used in the two-stage analysis (regression model) are calculated for each year separately to avoid being serially correlated.

Selecting inputs and outputs is an important task in specifying the DEA model. Based on prior literature, inputs are categorized into three groups - capital investment, labor, and other operating expenses, while patient numbers are the main constituents of the outputs (O'Neill *et al.*, 2008; Ozcan 2014). Over the period examined from 2005 to 2016, the most popular inputs used by studies were beds, numbers and type of staff (medical staff, nurses, and nonmedical staff), supplies, equipment, and infrastructure, whilst outpatients, inpatients, and surgery were the most frequent outputs (Kohl *et al.*, 2018), although this varied a little according to the authors' research questions and data availability. In relation to variable selection, Barnum *et al.* (2011) contend that DEA scores might be incorrect

when using non-substitutable inputs and non-substitutable outputs. Therefore, the authors proposed solutions such as using prices or reasonable weights to aggregate non-substitutable variables or incorporating just one of the non-substitutable variables as a proxy for input and output, respectively. However, such measures might be inapplicable for use in this study given the absence of available prices, the arbitrary weights chosen and the inability for a proxy to capture the multiple aspects of hospital production. In addition, the exclusion of high correlation inputs or outputs might also distort the efficiency estimated (Dyson, 2001).

Accordingly, in this study, seven inputs were used including the number of available beds (a proxy for capital investment), staff numbers, assigned to 5 distinct categories (representing labor resources; see Table 8), and other operating expenses. With respect to the output side, since each patient has a different magnitude of complexity and requires different resources, "crude outputs" might not account for the patient heterogeneity (a heart transplant cannot reasonably be counted the same as a broken leg). Therefore, we used case-weighted volumes for outputs. Using reference cost data (NHS Improvement, 2017a), we classified hospital activities into four main categories, namely, inpatients, outpatients, emergency, and other services. Each activity is associated with HRGs (Healthcare Resource Group) with information about volumes and national average unit cost. The case-weighted volume for each output category in a hospital was calculated using the following formula:

$$\text{Case-weighted volume} = \frac{\sum_{j=1}^{n} x_{jh} c_j}{\bar{c}}$$

Where:

x_{jh} is the amount of activity or HRGj= 1…n at hospital h with $h = 1…m$

c_j is the mean unit cost of each HRG or activity.

$\overline{C}$ is the average cost of all HRGs at the hospitals studied and calculated by the formula: $\bar{c} = \frac{\sum_{h=1}^{m}\sum_{j=1}^{n} x_{jh}c_j}{\sum_{h=1}^{m}\sum_{j=1}^{n} x_{jh}}$

Relevant data were retrieved from official databases, including Monitor (2017), NHS Digital (2017) and NHS England (2017a). Since some foundation trusts had been recently transformed and some lacked sufficient data over the study period, 116 out of 157 foundation trusts were collected (70 acute, 32 mental health, and 14 specialist). However, these three types of foundation trusts are likely to have different production functions and characteristics resulting in problems with the homogeneity assumption if they were aggregated into a single sample. Given that the numbers of acute trusts met the rule of thumb[16] for sample size in carrying out DEA analysis, the mental health and specialist trusts were excluded from the research. Moreover, because we were dealing with a sample rather than a census of foundation trust data, we bootstrapped our results with 1,000 replications. Descriptive statistics of inputs and outputs are as follows:

Table 8. Statistical descriptions for variables used in the DEA model

Variable	Description	Mean	Std.Dev.
Inputs			
Number of beds	The average number of beds available at hospitals.	767	305
Medical staff	The number of full-time equivalent (FTE) medical staff	584	329
Nurses, health visitors, and midwives	The number of nurses, health visitors, and midwives (FTE)	1,494	766
Scientific, therapeutic and technical staff	The number of scientific, therapeutic and technical staff (FTE)	625	363
Support for clinical staff	The number of support to clinical staff (FTE)	1,366	640
NHS infrastructure support and others	The number of NHS infrastructure support and others (FTE)	754	403

[16]In order to have adequate numbers of degrees of freedom (adequate discriminatory power for the DEA model), the "n" (number of DMUs) should exceed the number of inputs (m) and outputs (s) by several times. More specifically, a suggested rule of thumb formula is that that "n" should be greater than max {m*s, 3*(m + s)}.

Operating expenses (£1,000)	Other operating costs excluding staff payroll, depreciation and amortization expenses, and impairments (adjusted for inflation)	132,002	94,851
Outputs			
Inpatients	The volume of activities related to patients admitted and treated while staying inside a hospital (elective, non-elective, excess bed days, critical care)	157,550	76,016
Emergency	The volume of activities in Accident and Emergency unit	110,092	44,956
Outpatients	Volume of outpatient services	526,910	266,144
Other services	Volume of other hospital activities, such as other acute services, community services, etc	2,644,675	1,635,744

4.3.2 Regression model

Although DEA allows us to identify efficient and inefficient hospitals, it cannot provide insights on the determinants of the efficiency scores, which is likely to be a major concern for policymakers and hospital managers. Moreover, while hospitals under evaluation are assumed, by DEA, to operate under similar environmental conditions, this assumption is not always applicable in reality. Indeed, various factors might affect hospital performance (e.g. patient characteristics, economic conditions, quality of resources), hence failure to account for these differences might lead to biased judgments (Jacobs *et al.*, 2006). Therefore, regression analysis has been extensively employed in the literature to identify an association between efficiency and possible explanatory factors (Cantor and Poh, 2018), although such analysis has rarely been conducted in recent literature on UK hospitals. We concede that two-stage analysis is an exploratory approach rather than being based closely on theory, and it is challenging to test whether environmental variables are independent of the production function, however, it is one of the two important evolutions of DEA which attempt to identify determinants of efficiency (Førsund 2018). It thus seems an important approach to take in our work. Despite the second stage regressions being

widely utilized, there has been little consensus in choosing the type of regression model for two-stage analysis. Banker and Natarajan (2008, p. 57) argue that "two-stage DEA-based procedures with ordinary least square (OLS), maximum likelihood (ML), or even Tobit estimation in the second stage significantly outperform the parametric methods". Similarly, while Hoff (2007) concluded that either the Tobit model or OLS is acceptable McDonald (2009) argue that since DEA scores are fractional data Tobit is not suitable and OLS should be the preferred method. Following a different approach, Simar and Wilson (2007, 2011) developed a bootstrapped truncated regression model, pointing out that the method outperforms the Tobit model and criticized the use of OLS for its dependence on restrictive assumptions.

Since the debate is ongoing, we have followed Lovell *et al.* (1994) who used super efficiency scores as the dependent variable to remove the upper bound problem (and hence the justification for Tobit regression modeling). Accordingly, we use a standard fixed effects panel data regression for the second stage analysis[17]. Our focus in doing so was to investigate how hospital characteristics, patient characteristics, asset management, and staff satisfaction might be associated with hospital efficiency. The specification for the regression model is given by the following equation:

$$EFF_{i,t} = \beta_0 + \beta_1 LN(SCOPE_{i,t}) + \beta_2 OCCU_{i,t} + \beta_3 OLD_{i,t} + \beta_4 LN(EQUIP_{i,t}) + \beta_5 LN(BUILD_{i,t}) + \beta_6 LN(INFOR_{i,t}) + \beta_7 LN(INVENT_{i,t}) + \beta_8 LN(RECEIV_{i,t}) + \beta_9 LN(SATISFA_{i,t}) + \beta_{10} LIQUID_{i,t} + \varepsilon_{i,t}$$

Where: The set of β_k (k= 1,.., 10) denote the parameters estimated from the panel regression model while $\varepsilon_{i,t}$ represent the error term and $E(\varepsilon_{i,t}) \sim N(0, \sigma^2)$. Descriptive statistics for variables are provided in Table 9.

[17]By way of a robustness check, we did also run regressions employing Tobit random effects, and OLS (with year dummies) and found that the results were quite similar. These results are available from the corresponding author.

Table 9. Statistical description of variables used in the regression model

Variable	Description	Mean	Median	Std. Dev.
EFF	The score(ln) using the super-efficiency model whereby the scores for efficient hospitals can exceed 100%.	4.66	4.62	0.19
LN(SCOPE)	The range of services that a hospital provides based on the numbers of the main groups in the Health Resource Groups (HRGs) – subchapters with three first code characters.	5.19	5.14	0.20
OCCU	The bed occupancy defined as the rate of occupied beds over available beds.	0.87	0.87	0.06
OLD	The proportion of the patients over 60 years divided by total admissions.	0.56	0.56	0.11
LN(EQUIP)	The average net value of medical equipment (including plant and machinery), divided by the total volumes of services provided (unadjusted for case-mix).	-5.43	-5.44	0.64
LN(BUILD)	The average net value of accommodation (building plus dwelling) divided by the total volumes of services provided (unadjusted for case-mix).	-3.13	-3.24	0.62
LN(INFOR)	The average net value of information technology divided by the total volumes of services provided (unadjusted for case-mix).	-6.88	-6.87	0.72
LN(INVENT)	The net value of inventory divided by total volumes of services provided (unadjusted for case-mix).	-6.58	-6.64	0.70
LN(RECEIV)	The net value of trade receivable divided by total volumes of services provided (unadjusted for case-mix).	-5.35	-5.49	0.74
SATISFA	The staff satisfaction defined as the scores given by staff from the annual survey ranging from 0 (lowest) to 5 (highest).	3.67	3.65	0.20
LIQUID	$LIQUID = \frac{\text{Net liquid resources}}{\text{Operating expenses (excluding Depreciation)}} * 365$ Net liquid resources =[Current assets (excluding Inventories, Derivative related assets, Available/held for sale assets and Charitable funds assets)]–[Current liabilities (excluding Charitable funds liabilities)]	3.46	1.10	19.4

Note: Some data use the same sources as DEA model; the satisfaction scores sourced from National NHS Staff Survey Co-ordination Centre (2018).

4.4 Results and discussion

4.4.1 DEA scores

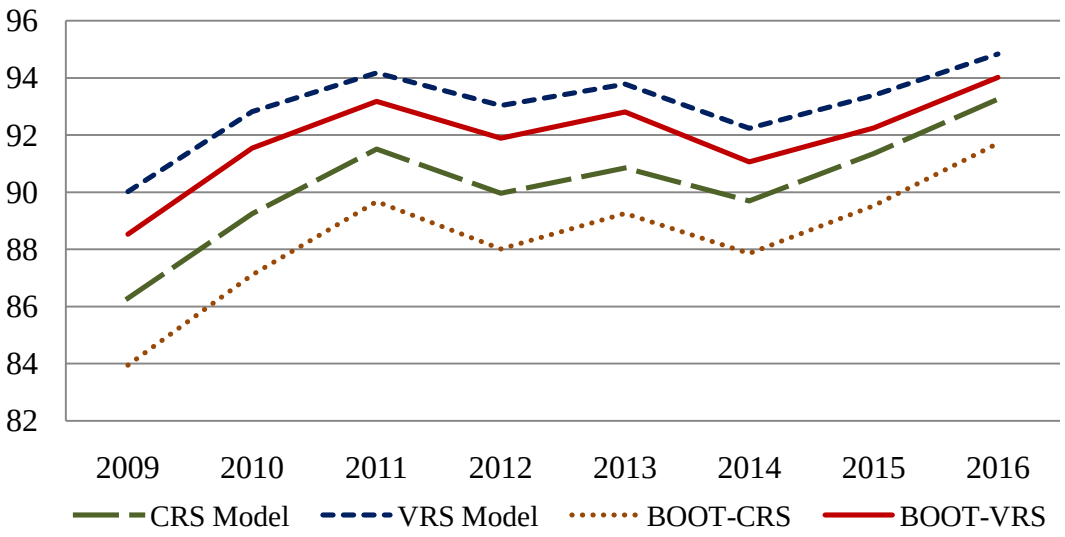
Results from the first stage are reported in Table 10 and graphed in Figure 17:

Table 10. DEA scores

	2009	2010	2011	2012	2013	2014	2015	2016
CRS Model								
Average	86.28	89.23	91.50	89.96	90.84	89.68	91.34	93.25
Median	86.60	88.92	91.12	91.06	92.13	90.06	92.34	94.09
Min	66.51	72.41	75.65	66.3	65.53	68.03	67.98	69.66
Max	100	100	100	100	100	100	100	100
Std. Dev	7.16	6.49	6.83	7.11	7.44	6.71	6.57	6.65
Number efficient units	4	4	11	8	11	5	4	19
VRS Model								
Average	90.01	92.81	94.17	93.02	93.77	92.23	93.38	94.83
Median	90.25	93.22	96.19	94.37	95.03	93.40	94.57	98.14
Min	68.52	74.56	79.19	70.22	67.96	68.03	70.09	70.28
Max	100	100	100	100	100	100	100	100
Std. Dev	7.54	5.99	6.08	6.7	6.66	6.42	6.26	6.39
Number efficient units	10	12	20	17	21	12	12	29

Over the period from 2009 to 2016, the overall efficiency of English acute foundation trusts increased in what appears to three phases: the first (2009-2011) mean scores increased, falling back slightly in the second phase (2012-2014) and then increasing again in 2015 and 2016. The efficiency trends are similar under both CRS and the VRS model. However, the VRS technical efficiency scores are constantly higher (as expected) than those of CRS confirming the existence of scale inefficiency. The VRS model also indicates that, on average, a large proportion of hospitals (76%) were operating at increasing returns to scale, 24% were at a constant return to scale and none experienced decreasing returns to scale.

Figure 17. The efficiency trend 2009 - 2016



Note: CRS – Constant returns to scale; VRS – Variable returns to scale; BOOT – Bootstrapped.

Examining the input and output data, it appears that the utilization of resources and responses to national policies could have been the main factors explaining the efficiency trends. Generally, over the period 2009-2016, input growth seemed to be lower than the respective growth in outputs (see Table 11). Specifically, a reduced number of beds and staff were the main factors contributing to the improved efficiency of the foundation trusts (over the whole period as well as for each specific phase). In both 2009-2011 and 2015-2016 when efficiency significantly improved growth in the number of employees and beds was much lower than in 2012-2014. Similarly, growth of inpatients, emergency, and other services in phases 1 and 3 were generally higher than phase 2.

Table 11. The average growth of inputs and outputs (%)

	Beds	Staff	Operating Expenses	Inpatients	Emergency	Outpatients	Others
2009-2011	-2.42	1.03	5.59	2.20	7.97	3.02	11.67
2012-2014	0.40	3.43	4.30	1.77	3.01	5.13	-0.57
2015-2016	0.28	-0.43	3.44	3.88	2.26	3.38	7.35
Average	**-0.49**	**2.74**	**4.73**	**2.43**	**4.39**	**3.66**	**5.66**

Table 12. Growth rates of employees

	Medical staff (12%)	Nurses & health visitors; Midwives (31%)	Scientific, therapeutic & technical staff (13%)	Support to clinical staff (28%)	NHS infrastructure support and others (16%)
2009-2011	1.73	0.91	3.08	1.01	-0.86
2012-2014	4.37	3.12	2.28	4.78	1.86
2015-2016	0.71	-2.14	-3.12	1.18	1.35
Average	**2.83**	**2.37**	**2.72**	**3.59**	**1.91**

Note: The number in () is the average proportion of each staff type.

Nurses, health visitors, midwives, and support to clinical staff made up almost 60% of the numbers of total staff (see Table 12). During the period from 2010 to 2014, while infrastructure related staff declined (around 20,000 FTEs), clinical and support to clinical staff increased (around 31,000 FTEs) which rebalanced the NHS staff structure toward clinical staff (Applyby *et al.*, 2015). A contributing factor may have been pressure arising from the Mid-Staffordshire scandal[18], which might have encouraged hospital management to recruit more staff to secure acceptable service standards (Applyby *et al.*, 2015; Powell and Mannion, 2016). In addition, increases in the number of hospital employees may have resulted from the Transforming Community Service program over 2010-2012 when Primary Care Trusts (PCTs) were required to separate commissioning functions and community services (Department of Health, 2010). As a consequence, some acute foundation trusts took over staff and assets when the provider arms were absorbed by the foundation trusts (Clover, 2011). Thus, the growth in staff numbers was the likely cause of the fluctuations in technical efficiency during the period 2012-2014.

In addition, NHS England during this period reduced the number of beds and increased the number of patients. Thus, greater utilization of beds to meet rising demand is likely to

[18]Mid-Staffordshire failure (dated back to late 2000s) related to unacceptable poor standards and high mortality rate at Stafford hospital which is considered to be the most notorious stigma in the NHS England history.

have been an important driving factor for efficiency improvement. As can be seen in Table 11, bed numbers slightly decreased or at least remained stable whilst output volumes constantly increased. Enhancing day surgery and shifting traditional treatment to community care might have helped to free up a number of beds (Baker, 2017). In addition, length of stay (LOS) declined consistently during the period 2009 to 2016.

Table 13. Length of stay (days)

	2009	2010	2011	2012	2013	2014	2015	2016
Investigated foundation trusts	4.42	4.26	4.21	4.25	4.25	4.22	4.15	4.17
Overall England	5.60	5.50	5.30	5.20	5.10	5.00	4.93	4.91

Sources: Authors' calculation and summarizing based on annual Hospital Episode Statistics (HES) published by NHS Digital.

The reduction in LOS could have been a response to meet the growing demand within the constraints of limited capacity where each hospital was required to actively respond to the introduction of the Referral to Treatment waiting time standard[19] (NHS England, 2017b). While the percentage of incomplete pathways within 18 weeks was around 90% before 2013, this indicator had increased to about 93% during 2013-2016. Some hospitals might have responded to the requirements by putting a higher priority on less complex patients as they require a shorter time to treat (Morris 2018), but others might have dealt with the issue by reducing patients' length of stay (Lewis and Edwards, 2015; Nuffieldtrust, 2014). Indeed, a variety of measures have certainly been deployed for this purpose, including the provision of recovery care at home through the use of "virtual wards"; improved pathways for frail patients; additional supply of seven-day support for discharged patients; and special arrangement for seniors to make early decisions about the treatments. Through

[19]Patients should be treated either as an inpatient or as an outpatient within 18 weeks of the referral. In 2013 additional target set was that no-one has to wait for more than 52 weeks to be treated. In June 2015, the admitted (90%) and non-admitted (95%) metrics were terminated, the only measure left is incomplete pathway standard.

these innovations, NHS hospitals shortened patients LOS, and avoided financial penalties, in order to cope with a limited number of beds.

4.4.2 Regression results

In the second stage, we identify the determinants of efficiency which we previously measured using DEA for English acute foundation trusts. Different panel data models (Fixed effects and Random effects) were estimated and since the Hausman test result did not support that the composite error term was uncorrelated with the explanatory variable (*Prob>chi2* = 0.000), we elected to adopt the Fixed effects model. In order to compare both the significance, and the relative size (and hence importance) of the explanatory variables, suggested by the extant literature, both the standard model and normalized fixed-effects model (using the standardized values of the original regressors) were estimated.

The results of both regression models are provided in Table 14. We also conducted a commonly employed multicollinearity test and found that the mean variable inflationary factor (VIF) was 2.48 (none of the VIFs is larger than 5), suggesting that the regression model does not suffer from multicollinearity problem (Berta *et al.*, 2010; Ding, 2014; Dong, 2016; Jindal *et al.*, 2018; Şamiloğlu and Akgün, 2016) (see Table 15).

Table 14. Regression results

Variables	Fixed effect model	Standardized model
LN(SCOPE)	-0.139***	-0.022*
	(0.045)	(0.013)
OCCU	0.476***	0.025***
	(0.155)	(0.009)
OLD	-0.442**	-0.062**
	(0.213)	(0.024)
LN(EQUIP)	-0.016	-0.007
	(0.036)	(0.024)

LN(BUILD)	-0.101***	-0.058***
	(0.035)	(0.022)
LN(INFOR)	0.022	0.021*
	(0.015)	(0.011)
LN(INVENT)	0.047	0.037
	(0.035)	(0.025)
LN(RECEIV)	0.012	0.004
	(0.023)	(0.019)
SATISFA	0.073	0.005
	(0.053)	(0.010)
LIQUID	-0.004***	-0.074***
	(0.001)	(0.010)
Constant	5.085***	4.657***
	(0.291)	(0.005)
Observations	483	483
R-squared	0.218	0.182
Number of hospitals	69	69

Standard errors in parentheses

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

Table 15. Variance inflation factor (VIF)

Variable	VIF	SQRT VIF	Tolerance	R-Squared
LN(SCOPE)	1.39	1.18	0.72	0.28
OCCU	1.13	1.07	0.88	0.12
OLD	1.36	1.16	0.74	0.26
LN(EQUIP)	3.82	1.96	0.26	0.74
LN(BUILD)	4.08	2.02	0.25	0.75
LN(INFOR)	1.70	1.30	0.59	0.41
LN(INVENT)	3.57	1.89	0.28	0.72
LN(RECEIV)	4.94	2.22	0.20	0.80
SATISFA	1.60	1.26	0.63	0.37
LIQUID	1.19	1.09	0.84	0.16
Mean VIF	**2.48**			

The results from our regressions suggest that scope, bed occupancy rate, proportion of elderly patients, effect of accommodation, information technology and liquidity are all statistically significant determinants of relative technical efficiency.

Economies of scope arise when a hospital could make the cost lower through diversification of the types of services provided by taking advantage of sharing the resources used. However, it is important to remain cognizant that diseconomies of scope may also arise – that is, that relative greater diversification can cause inefficiency. The negative coefficient in our analysis supports the (latter) case that hospitals with more types of services tend to have lower efficiency. As explained by the Monitor (2014) expanding the scope of services might, in fact, have deleterious effects on financial status, since additional income might be insufficient to offset the additional costs incurred.

A higher bed occupancy rate is generally assumed to be positively associated with efficiency. A theoretical and simple threshold of hospital occupancy rate is around 85%, at which point access block, waiting lists or seasonal bed crises can still be mitigated (Bain 2010; Jones 2001, 2011; Keegan 2010). The average occupancy rate of these acute foundation trusts over 2009-2015 was at about 87%, higher than the recommended cut off ratio but still lower than the maximum rate (90%) proposed by National Institute for Health and Care Excellence (NICE, 2017). Thus, a strong positive association between occupancy rate and hospital efficiency confirms the importance of this variable in the modelling.

Older patients normally require more intensive medical examinations and elicit higher resource consumption[20]. Thus, a higher proportion of older patients might be expected to inversely affect efficiency. The average percentage of older patients (aged 60^{+}) in the acute foundation trust between 2009 and 2015 was 56%. Although one might expect that

[20]In examining expenditure characteristics of England public hospitals, Kelly et al. (2016) point out that costs accelerate when the patients' age increase.

the costs incurred when treating old patients should be covered by the tariff, it is, however, reasonable to argue that the fixed tariff might only partially adjust for the variations in the patient ages and the inefficiency still occurs when the proportion of old patients increase. Thus, the negative coefficient in our empirical estimations confirms that a higher proportion of elderly patients tends to have a deleterious effect on relative technical efficiency.

The effect of accommodation (buildings and dwellings) suggests that with the same amount of assets, a hospital with a higher volume of services delivered yields higher efficiency. It is important to note that the nett building value excludes depreciation, but includes capitalization for new construction and work which extends the functionality or useful life of the asset (as well as other 'accounting treatments'). Therefore, accomodation is unlikely to be correlated with depreciation. Moreover, as indicated by Lord Carter of Coles (2016), inefficiency related to accommodation in acute trusts stems mainly from underutilization or inappropriately used buildings for non-clinical and unproductive purposes. Indeed, space that was not occupied by patients varied significantly across trusts, ranging from 12% to 69%. Other examples of wasted resources that might affect efficiency include buildings built in inappropriate locations, outdated buildings (due to changes in treatment), and over-specified inflexible spaces (Edwards, 2011). Since accommodation accounts for the majority of total fixed assets (74%) and given their substantial operating and maintenance costs, our finding implies that there are significant opportunities to improve efficiency through better use of these facilities.

The positive effect of information technology (IT; in the standardized model only) implies that investment in IT improves hospital efficiency. A large part of IT relates to the digitalization of the health system which aims at better services and lower cost. In fact, in an attempt to improve its efficiency, since 2002 NHS England has deployed the National Programme for Information Technology (NPfIT) which represents a £12.4 billion

investment. Although terminated in 2011, it was able to achieve "single national patient identifier, infrastructure to provide core services, and national electronic prescription" (Wachter *et al.*, 2016). It should be noted that NPfIT goals still remain and that the NHS continues to pursue digitization in secondary care (NHS England, 2014).

Liquidity is one of the financial indicators utilized by the Monitor to evaluate financial risk (a higher metric means lower risk) of each hospital. However, the regression result suggests that liquidity is negatively associated with hospital efficiency. The relationship between liquidity and profitability in the business sector has been examined with mixed evidence of both positive and negative correlation between these two factors (Şamiloğlu and Akgün, 2016; Umobong, 2015). Although it might not be strictly appropriate to apply business concepts to hospitals (and also because liquidity is often measured differently), foundation trusts are expected to generate a surplus which they are entitled to reinvest in services from retained earnings. Thus, when foundation trusts hold large nett working capital to enhance solvency, there might be a trade-off relationship between liquidity and profitability, as well as efficiency.

Notably, all of the coefficients in the standardized model are relatively small which suggests that public policy interventions designed to address the determinants can be expected to have relatively marginal effects on overall relative technical efficiency. When comparing the magnitude of potential impacts, the standardized coefficients suggest that liquidity and proportion of old patients have the greatest effects, followed by accommodation, occupancy rate, and information technology. Public policy might, therefore, be prioritized first for the variable that is most amenable to change and likely to produce the greatest effect such as liquidity, accommodation, and occupancy.

For the variables that were not statistically significant the signs of the coefficients were broadly in line with expectations. In order to meet the treatment timeline targets, NHS England invested more on medical equipment and as a result, there was an increase in the

number of MRI units and CT scanners (Cylus *et al.*, 2015). However, new medical technologies are generally expensive and may not be the most efficient use of capital (Sorenson *et al.*, 2013). Therefore, hospitals with a higher value of medical equipment might be less efficient (hence the negative coefficient in Table 14). The positive association for inventory was in line with the argument that "hospitals are likely to overstock to ensure high patient safety and bring the costs down when contracts with suppliers are negotiated" (Gebicki *et al.*, 2014, p. 219). Similarly, the coefficient for receivables was consonant with our expectations - generally, shorter receivable periods might be expected to improve profitability, however, a strict policy to promptly collect can divert resources from more efficient uses (Şamiloğlu and Akgün 2016). Finally, the positive coefficient for staff satisfaction was also in line with our expectations - staff are likely to be more dedicated and enthusiastic (e.g. lower absenteeism) when they have higher satisfaction (consistent with previous studies; Powell *et al.*, 2014).

4.5 Conclusion and policy implications

Our objective was to extend the literature on UK hospitals conducted prior to 2006 through an evaluation of how the technical efficiency of English foundation trusts changed over 2009 to 2016 under conditions of budget austerity and restructuring and examine the causes of variation in efficiency across hospitals.

Overall, technical efficiency of the acute foundation trusts investigated clearly improved, albeit with fluctuations over the period 2012-2014 period. Through our data analysis and other evidence, we have suggested that the efficiency gains appear to mainly result from bed optimization and initiatives to reduce the length of stay. We also found evidence to support the contention of scale economies given that most of the hospitals were operating at the increasing returns to scale.

The results obtained from the two-stage analysis with regard to determinants of efficiency

seem to be consistent with findings in the prior literature: Wider scope and a higher proportion of old patients reduce efficiency, while higher bed occupation rates increase efficiency. Although hospitals have less discretion in choosing what services to provide and thus have a lower degree of control over these factors, utilizing tools such as Service Line Report (SLR) and Patient Level Costing (PLICS) to identify the least efficient areas and to balance services provided seem to have potential to mitigate negative impacts associated with scope.

As the number of older patients is likely to continue to trend upwards, optimization of hospital beds will also be an essential means through which to improve efficiency. Better management of length of stay (LOS), especially adequate management of older patients LOS will benefit not only hospitals but also the patients themselves. In this regard, as a policymaker, the NHS Improvement has an important role to disseminate successful initiatives and good practice (in optimizing patient flow and shifting healthcare closer to home) so that less-efficient hospitals can emulate their peers.

In addition to the external factors which have been well established in the extant literature, our study has also made pioneering efforts to identify a number of internal factors which can be more easily targeted by hospital managers to improve efficiency. In particular, we found that the management of internal resources can play a vital role in improving efficiency. The relationship identified in this study between asset utilization and efficiency suggests that acute foundation trusts need to maintain an optimal level of specific assets to provide services with an acceptable level of quality, and that this might be achieved by exploiting resources such as buildings and dwellings and applying information technology more effectively. Specifically, attention should be paid to estate management planning in which hospitals adequately assess the current state of assets used and carefully forecast the demands for services in the future. This action might pave the way for more detailed measures such as reconfiguration of hospital sites and hospital services to reduce

unutilized spaces or additional measures to tackle maintenance backlogs and save on operating costs.

Measuring hospital efficiency is a complex and challenging task. This study provides evidence on the technical efficiency of English acute foundation trusts during a period of major restructuring and changes over recent years. In addition to the external factors which have been well examined in the literature, our study identified the determinants of efficiency, especially those related to internal factors which might provide more useful tools for hospital managers to directly target efficiency improvements. It thus serves to underline the importance of conducting second-stage regressions (hitherto widely neglected in the literature for UK hospitals) using a wide array of plausible patient, hospital, and asset variables to understand what is driving the observed efficiency.

CHAPTER 5: MEASURING EFFICIENCY OF THE NEW ZEALAND DISTRICT HEALTH BOARDS: AN EMPIRICAL RESEARCH USING TWO-STAGE DATA ENVELOPMENT ANALYSIS

Health care is one of the core public services in New Zealand and consumes the second largest share of government expenditure. It is thus important to monitor the technical efficiency of the health services. We employ a two-stage Data Envelopment Analysis of New Zealand District Health Board (DHBs) data over the period from 2013/14 to 2016/17, and found that DHBs generally improved their technical efficiency during this time. Our results suggest that further improvements to efficiency might be possible through greater attention to budget preparation, improvements to building asset utilization and optimization of patient flow at emergency departments.

5.1 Introductions

Public spending on healthcare services in New Zealand makes up about 21% of core government expenditure and ranks as the second largest component of government expenditure, behind social care (The Treasury, 2017b). In common with many countries, growing demand, rising health care cost, and constrained budgets have been a cause of concern to public policy-makers. The New Zealand Health Strategy (updated in 2016) suggests that without changes the services may not be financially sustainable into the future (Ministry of Health, 2016b). Indeed, it is predicted that the greatest growth in government spending will be in the area of healthcare cost and modelling suggests spending (as percentage of nominal GDP) may increase from 6.8% in 2010 to 10.8% in 2060 due to rising demand, prices and demographic changes (Ministry of Health, 2016; New Zealand Treasury, 2013). Against this background, improvements to technical efficiency defined as the conversion of inputs (e.g., staff and operational expenditure) into outputs (e.g., the volume of patients treated, life expectancy) become important, not only

for the financial sustainability of the sector but also for ensuring that there are sufficient resources to improve the range and quality of services (Knopf, 2017; Nolan, 2018).

The New Zealand health system provides universal access to health services through a network including public, private, and non-governmental organizations (World Health Organization, 2014). With respect to the public sector, twenty District Health Boards (DHBs) play the key role in implementing health policies and manage the majority of the daily health services. Apart from DHBs, there are other for-profit and not-for-profit entities which offer services in communities, residential facilities and private hospitals (Ministry of Health, 2017). Most health services in New Zealand are publicly funded in which eligible people (e.g. New Zealand citizens) can enjoy free public hospital services, subsidies on prescriptions, and support services for disabilities. In addition, other sources of funding for health services come from the Accident Compensation Corporation (for medical costs related to accidents), out-of-pocket payments, and private health insurance (Ministry of Health, 2018).

Unlike conventional hospitals, DHBs are responsible for improving the health status of people living within defined geographical locations. They not only run public hospitals (84 units, in total) but also allocate funds for primary care (e.g. general practitioners), secondary care, and aged-care services. Part of DHB back-office functions, such as administrative, financial, and information systems are managed by a shared-services entity.

DHBs consume around three-quarters of the public health budget (Ministry of Health, 2017) which is allocated from the central government based on the Population-Based Funding Formula (PBFF). PBFF is adjusted for social-economic factors such as age, gender, ethnicity, and tertiary cost structures aimed at facilitating horizontal equity in health provision. DHBs are charged with: (i) covering all their operating costs, (ii) prudent

management of assets and liabilities, (iii) maintaining liquidity, and (iv) ensuring long-term financial sustainability.

In New Zealand, the technical efficiency of public services has not been regularly measured and generally does not rate a mention in annual reports (New Zealand Productivity Commission, 2018). Moreover, most of the extant work has focused on specific aspects of health service provision and thus has not been able to provide comprehensive guidance to the sector at the regional level (Knopf, 2017). Indeed, the New Zealand Treasury (2016) has acknowledged that their method to assess the performance of DHBs does not accurately measure efficiency because it does not include non-hospital and non-surgical activity. In addition, Treasury work has been unable to eliminate costs related to other services and thus may lead to a mismatch between the resources used and the volume of services provided.

We seek to address these gaps in the extant literature through an evaluation of the technical efficiency of DHBs over the period from 2013/14 to 2016/17, inclusive. We employ the robust and well attested technique of data envelopment analysis (DEA) to estimate relative technical efficiency of each DHB in addition to monitoring the trend in efficiency over the period of analysis. In addition, we conduct a second stage bootstrapped truncated regression model to explain and identify the determinants of efficiency in order that we might provide public policy recommendations aimed at improving the efficiency of DHBs.

The remainder of the paper is organized as follows: In the next section, we review the extant literature with a particular focus on identifying the links between efficiency and operating environment. Following this, we describe our empirical methodology with particular emphasis on justifying the model specification. We then present the results of our empirical estimations. We conclude our essay with a review of the public policy implications and suggestions for further research.

5.2 Prior literature

Since its first application in 1983, DEA has been increasingly applied as a performance measurement and benchmarking tool to evaluate the efficiency of healthcare at different levels (Ozcan, 2014). An important assumption in DEA is the homogeneity assumption, which requires all decision-making units (DMUs) to be broadly similar in their organization and operating environment. However, strict homogeneity is rarely true and thus two-stage analysis is generally employed (Cantor and Poh, 2018). Moreover, two-stage analysis allows researchers to identify the determinants of technical efficiency.

Varabyova and Schreyögg (2013) compared the efficiency of the hospital sector for Organization for Economic Co-operation and Development (OECD) countries and found that good health outcomes are negatively associated with efficiency, which suggests that countries that have higher health expenditures *per capita* should look to improve hospital efficiency. Similarly, Kaya Samut and Cafrı (2016) provide evidence that suggests that countries with higher education attainment and relatively greater wealth have more efficient health systems. However, in a broader study which measured and compared the efficiency of entire healthcare systems in OECD countries (Hadad *et al.,* 2013), statistically significant evidence could not be found for associations between social-economic factors (e.g. GDP *per capita*) and efficiency.

Unhealthy lifestyles (e.g., smoking, drinking, and obesity) and the prevalence of morbidity (e.g., diabetes and cancer) can contribute to poor health and may indirectly reduce efficiency. Takundwa *et al.* (2017) investigated the efficiency of English Clinical Commission Groups (CCGs) and found no statistically significant association, which was somewhat contrary to expectations. However, a study of Canadian Health Regions by Allin *et al*. (2014) provided evidence to suggest that efficiency did indeed decrease in response to poor lifestyle choices and the incidence of morbidity. It can be contended that

inefficiency might be due to higher unplanned readmission rates and longer average length of stay. Notably higher average income, when considered on a regional basis, was inversely related to efficiency (Allin *et al.* 2014).

Extant evidence also exists which suggests that a higher proportion of elderly patients is inversely associated with efficiency while higher bed occupancy rates as well as reductions to the length of admission improved efficiency (Czypionka *et al*. 2013; Fixler *et al*. 2014). Additionally, the relationship between efficiency and other factors have also been examined, such as the effect of urbanity, ownership (public, private), and type of hospital (for-profit and non-profit hospital; teaching and non-teaching). Generally, the evidence suggests that private hospitals are more efficient than public hospitals (Czypionka *et al*. 2013; Kaya Samut and Cafrı 2016).

In the New Zealand context, little work has been done on the measurement of healthcare efficiency using DEA. One notable exception is Rouse and Swales (2006) who employed DEA as a method to set prices for hospital services (e.g. medical and surgical, pregnancy and childbirth, and community health). The specification for their model used expenditure as a single input which was hypothesized to be converted into a relatively large number of outputs (e.g., discharge volumes, outpatient attendances) commensurate with each service category. More recently, Sandiford *et al*. (2017) used DEA to measure the efficiency of DHBs in improving life expectancy of both Maori (indigenous) and European ancestry populations. By assuming that the population-based funding formula is fully adjusted for variances in socio-economic conditions, the input was therefore set to unity for each DHB. The authors found that while the proportion of Maori or the initial level of life expectancy did not impact on actual life expectancy, efficiency was significantly associated with financial performance (budget deficit).

Our review of prior literature suggests that the selection of explanatory factors varies by research context. Moreover, while some factors are found to have consistent effects, the association of several contextual factors with efficiency was found to be mixed. Although an exhaustive list of variables could be incorporated in the two-stage analysis, two broad categories identified are: (i) environmental factors, and (ii) internal factors (such as financial performance, asset utilization). While the former is largely beyond managers' control, the latter is normally dependent on the clinical governance (where managers have more discretionary power to decide how resources are used and services provided).

5.3 Model selected and data

5.3.1 DEA model

The underlying nature of the DEA model is a linear programming technique which optimizes the outputs/inputs ratio with unknown weights. Thus, assuming that DHBs (z_i) using k inputs (x_j) to provide m outputs (y_i), a basic DEA model is presented as follows:

$$MaxH_0 = \frac{\sum_{i=1}^{m} u_i y_{i0}}{\sum_{j=1}^{k} v_j x_{j0}}$$

$$\text{Subject to: } \frac{\sum_{i=1}^{m} u_i y_{iz}}{\sum_{j=1}^{k} v_j x_{jz}} \leq 1; z = 1, \dots, n$$

$$u_i, v_j \geq 0 \; for \; all \; i \; and \; j$$

Where:

H_0= the efficiency score of DHB_0;

x_{jz}= the amount of input j used by z_{th} DHB;

y_{iz}= the amount output i provided by z_{th} DHB;

u_i= weight assigned to output i;

v_j= weight assigned to input j

The above model in ratio form does not indicate whether the sums of weighted inputs are minimized (input-oriented) or if instead the sums of weighted outputs are maximized

(output-oriented). In the healthcare sector, the input-oriented model is often used as there is a wide acceptance that managers and policy-makers generally have more control over inputs than outputs and that more priority is put on costs than increasing health care demand (Cantor and Poh, 2018; O'Neill *et al*., 2008). Therefore we also employ an input orientation in the analysis that follows.

A second decision relates to whether to use the Charnes, Cooper, and Rhodes (CCR) model or the Banker, Charnes, and Cooper (BCC) model. While the former implies constant returns to scale (CRS) the latter includes variable returns to scale (VRS). As the debate on selection CCR or BCC is ongoing (Kohl *et al*., 2018), in the first stage, we estimate efficiency scores under both CRS and VRS assumptions. However, in the second stage analysis, since some explanatory factors are used as a proxy for size, the CCR model is more appropriate to avoid possible bias.

Unlike cross-sectional DEA analysis, an objective of the first part of this research is to investigate changes in efficiency with panel data. However, it is generally not appropriate to compare DEA scores between different cross-sectional analyses as they are calculated based on different reference groups. Therefore, we pool all of the four-year data into a single DEA sample for estimating technical efficiency (referred to as global intertemporal DEA). The results are then regrouped according to each year to calculate the mean scores.

Efficiency scores are sensitive to the selection of inputs and outputs. Based on the production process, inputs should capture all the resources consumed and outputs should fully cover the services provided. However, it seems that the variables used to measure efficiency in health care vary by circumstances, authors' rationale, and availability of data. For hospitals, capital investment, labour, and other operating expenses are the main inputs whereas inpatients and outpatients are the key outputs (Cantor and Poh, 2018; O'Neill *et al*., 2008). Regarding primary care, the outputs might be activities (number of visits or examinations), quality indicators or health outcomes (Pelone *et al*., 2015).

At regional level studies (such as England CCGs; Takundwa *et al*. (2017)) funding *per capita* and the number of general practitioners have been employed as inputs; the output side has included health status score, respiratory disease survival rate, cancer survival rate, cardiovascular score, cancer score, and chronic obstructive pulmonary disease score. Likewise, in the case of Canadian Health Regions, spending *per capita*, education, number of recent immigrants, and proportion of non-Aboriginals have been used as inputs whereas outputs have included potential years of life lost from treatable causes, mortality from treatable causes, and survival rate from treatable causes (Allin *et al*., 2014).

New Zealand DHBs operate as a combination of local authorities and hospitals. The set of variables used, therefore, should capture their operational characteristics. Accordingly, we define total expenditure (adjusting for inflation) as an aggregated input since it fully incorporates relevant resources consumed. However, payments to other DHBs are excluded from the total expenditure. The reason for this elimination is to ensure the resources consumed are commensurate with the outputs produced.

Outputs include inpatients, outpatients, elective surgeries, patient presentations to the Emergency department, infants aged eight months who have completed their primary course of immunization, the number of patients supported to quit smoking, and life expectancy. Since each patient requires different treatments depending on the morbidity and severity of the condition, inpatient volumes are case-weighted discharges. Unlike inpatients, there is no case-mix for outpatients, to enable comparison, aggregated outpatient discharge volumes are price weighted to reflect the differences in the level of resources needed. We also selected life expectancy at birth as an output since it reflects the overarching objective of the New Zealand healthcare system – "New Zealanders live longer, healthier, more independent lives" (Ministry of Health, 2014). The remaining outputs relate to national targets which have been used to measure the DHBs' performance. Although there are six metrics, we included only four of them due to missing data.

All DHBs have similar functions, operating under the same legal framework, hence the homogeneous assumption holds in this case. Moreover, concerning the number of variables employed, in order to have adequate numbers of degrees of freedom, a suggested rule of thumb formula (often referred to as Nunamaker's rule) is that that number of DMUs should be greater than the sum of the number of inputs and outputs multiplied by three times. Our analysis covers 20 DHBs and includes eight variables in total which have sufficient discriminatory power when estimating efficiency scores over a 4-year pool of data. Data were obtained from official databases and augmented with official information request[21] to DHBs and authorities. Specifically, we retrieved available data at the websites of the Ministry of Health and DHBs (e.g. volume of inpatients and outpatients, annual reports). In addition, we made requests to the Ministry of Health, The Treasury, and Statistics New Zealand for aggregated financial data, the performance against health targets, population, and life expectancy. Descriptive statistics of input and outputs are as follows:

Table 16. Variables used in the DEA model

Variable	Description	Mean	Std. Dev.
Input			
Expenditure	Total expenditure (NZ$1,000)	678,326	523,910
Outputs			
Inpatient	Volume of activities in treating inpatients	42,055	35,919
Outpatient	Volume of activities in outpatient service	13,228	9,578
Emergency	The number of patients at the Emergency unit	54,079	30,323
Surgery	Volume of elective surgeries	9,167	6,164
Smoke	The number of patients supported to quit smoking	85,930	60,638
Immunization	Number of eight-month infants get immunized	2,753	2,241
Life expectancy	Average life expectancy at birth	80.95	2.00

[21] The platform where we made requests for relevant data is at the website: https://fyi.org.nz/

5.3.2 Regression model

After reviewing studies where conventional DEA was combined with other techniques, Cantor and Poh (2018) conclude that second-stage regression analysis is the most appropriate approach to explain the variation in organizational performance. Geographical location, population, competition, and ownership are commonly employed independent variables or control variables. Since DHBs have high discretion in managing health services at their regions within the budget allocation, we have placed more emphasis on examining the effect of internal factors.

Under the New Zealand Public Health and Disability Act 2000, annual planning and strategic planning are statutory responsibilities. The annual plan is an important document presenting targets for two key areas (service delivery and financial status), which must align with its strategic plan and be broadly consistent with the New Zealand health strategy. As a result, the annual budget derived from the annual plan is expected to play a crucial role in managing and allocating resources effectively. Thus an unfavourable deviation from initial operating expenditure might be expected to reduce efficiency.

Fixed assets in good condition might be expected to contribute strongly to the delivery of health services. However, health assets in New Zealand have been consistently assessed as being in poor condition and this might well lead to service disruption (Controller and Auditor-General, 2016a). Moreover, indigenous Maori tend to have higher susceptibility to a number of health conditions (higher morbidity and lower life expectancy) and thus might be expected to attract greater resource need and hence lower efficiency. Thus, DHBs with insufficient investment in assets and a higher proportion of Maori persons are expected to be less efficient.

District Health Boards are also responsible for achieving the national health targets which include metrics used to improve the performance of health services that reflect public and government priorities. The performance for each target is used to rank DHBs against one another and is publicly published on the website of the Ministry of Health. Therefore,

although they are not themselves efficiency measures, we expect that achieving targets might improve efficiency (Knopf, 2017).

For identifying the determinants of efficiency (the second part of our work) different models have been proposed including ordinary least square (OLS) (Banker and Natarajan, 2008; McDonald, 2009), Tobit (Hoff, 2007), and bootstrapped truncated regression (Simar and Wilson, 2007). Given wide applications, the debate on assumptions underpinning regression models is ongoing. Banker *et al*. (2019, p. 368) indicate that the "Simar-Wilson model heavily depends on the assumption that the actual data generating process (DGP) exactly matches their assumed DGP and their approach does not yield correct inferences in environments characterized by stochastic noise". However, Simar and Wilson (2011, p. 205-206) contend that "conventional inference methods fail to give valid inference due to the fact that in the second-stage, true efficiency remains unobserved and must be replaced with DEA estimates of efficiency, and these are correlated by construction". In practice, the Simar-Wilson approach has been widely adopted as the norm in second stage analysis (Kohl *et al*., 2018).

In this study, a double bootstrap truncated regression (or "algorithm # 2") is used to investigate the potential determinants of DHBs' efficiency. See Simar and Wilson (2007) for required assumptions and Badunenko and Tauchmann (2018) for further detail on the procedure. Accordingly, we used one input and seven outputs in estimating biased corrected scores ($\widehat{\theta_i^{bc}}$) and seven independent variables (Z_i) in the bootstrap truncated regression model given by:

$$\widehat{\theta_i^{bc}} = \beta Z_i + \gamma_i, \quad i = 1, \dots, n$$

Where:

$\widehat{\theta_i^{bc}}$ is bias-corrected technical efficiency scores and internally estimated using one input and seven outputs as specified in the DEA model.

Z_i is the set of contextual factors.

Descriptive statistics for these variables are provided in Table 17.

Table 17. Explanatory factors used in the regression model

Variable	Description	Mean	Std. Dev.
Building	Nett value of building	201,763	165,244
Infortech	Nett value of information technology	2,620	4,022
Maori	The proportions of discharged Maori patients	18.42	10.47
Emergency	The performance in the target to treat patients within six hours since presentation at the emergency department	0.94	0.02
Exvar	The variance between actual expenditures and plan (the percentage of variance over the plan expenditures)	1.51	2.55
Outvar	The variances between actual and plan volume of outpatients	-138	842
Invar	The variances between actual and plan volume of inpatients	1,091	1,912

5.4 Results

5.4.1 Efficiency estimates

Average DEA scores under CRS and VRS assumptions are reported in Table 18:

Table 18. **Efficiency scores**

	2013	2014	2015	2016
CRS Model				
Average	94.14	93.57	96.18	94.99
Min	86.36	84.85	88.59	84.64
Std. Dev.	4.72	5.76	4.08	4.85
Number of efficient units	2.00	4.00	7.00	4.00
VRS Model				
Average	95.80	95.56	98.31	97.99
Min	87.07	85.01	90.41	90.31
Std. Dev.	4.03	4.79	3.08	3.42
Number of efficient units	5.00	7.00	13.00	13.00

Notably, for the most part, CRS and VRS were quite similar suggesting that scale plays a relatively small role in DHB efficiency. Moreover, overall DHBs improved their efficiency over the period under analysis despite isolated fluctuations. Efficiency trends were closely associated with growth in expenditure and volume of services provided. For instance, the decline in 2014 and 2016 could be explained by high growth in expenditure and lower volume of services. On the other hand, an apparent surge in technical efficiency in 2015 might have resulted from lower growth in costs and greater growth in services outputs relative to the other years examined (see Table 19).

Table 19. **Growth of input and outputs**

	Expenditure	**Inpatient**	**Outpatient**	**Emergency**	**Surgery**	**Smoke**	**Immunization**	**Life expectancy**
2014/13	2.43	1.80	4.23	2.32	3.19	22.89	0.81	0.20
2015/14	1.91	2.69	9.67	3.52	19.88	13.88	0.91	0.20
2016/15	2.80	2.01	-1.40	0.98	1.85	0.37	-0.01	0.06
Average	**2.38**	**2.17**	**4.07**	**2.27**	**8.01**	**12.00**	**0.57**	**0.16**

Further investigations, illuminated the factors driving technical efficiency trends. The formation of Health Benefits Limited (HBL) in July 2010 aimed to reduce the cost of non-clinical services to DHBs by optimizing the delivery of administrative, support and procurement services and the target set for the five years (2011-2016) was to achieve a gross savings of $700 million. Actual total savings by June 2014 was reported at $301.8 million (Controller and Auditor-General, 2015). Analysis of the cost structure (see Tables 20 and 21), revealed a reduction in non-clinical cost as a percentage of total expenditure and its actual growth was lower than most of the other main items, supporting claims for the effectiveness of this shared-service.

Table 20. **Expenditure growth in details**

	Actual/Plan (%)				Actual growth (%)			
	2013	**2014**	**2015**	**2016**	**2014/ 2013**	**2015/ 2014**	**2016/ 2015**	**Average**
Personnel	1.39	1.41	1.45	0.45	3.78	3.84	3.94	3.85
Outsourced	21.07	21.30	14.59	17.90	13.95	5.49	8.82	9.36
Clinical	1.64	4.37	2.77	3.63	4.35	0.97	4.52	3.27
Non-clinical &infrastructure	3.44	5.38	6.78	8.60	(1.48)	(0.72)	6.53	1.38
Providers payments	(4.12)	(4.89)	(1.69)	(3.02)	1.38	3.68	3.46	2.83
Others	12.32	17.91	1.95	15.67	12.12	(11.84)	13.55	3.92
Depreciation	(1.74)	(5.03)	(3.34)	(3.48)	3.68	3.27	3.20	3.38
Capital charge	0.04	4.68	10.07	(14.13)	5.07	12.10	(15.28)	(0.07)
Interest	(6.85)	(6.37)	(6.83)	(36.04)	4.14	(12.54)	(36.48)	(16.68)
Total	**(0.23)**	**(0.12)**	**0.95**	**(0.10)**	**3.11**	**3.21**	**3.63**	**3.32**

Notes: Some did not consistently report their Income statements, only 16/20 DHBs are incorporated to ensure the comparability.

Table 21. **DHBs' cost structure**

	2013	**2014**	**2015**	**2016**	**Average**
Personnel	37.51	37.75	37.99	38.10	37.84
Outsourced	3.92	4.33	4.43	4.65	4.33
Clinical	8.46	8.56	8.38	8.45	8.46
Non-clinical &infrastructure	4.69	4.48	4.31	4.43	4.48
Providers payments	39.94	39.27	39.44	39.38	39.51
- Non-DHBs providers	30.21	29.42	28.72	29.46	29.45
- DHBs providers	9.73	9.85	10.72	9.92	10.06
Others	0.91	0.99	0.84	0.92	0.92
Depreciation	2.44	2.46	2.46	2.45	2.45
Capital charge	1.32	1.35	1.46	1.20	1.33
Interest	0.80	0.81	0.69	0.42	0.68

DHBs also appear to have controlled staff expenses successfully. Although actual growth in personnel cost was slightly higher than the average, the actual cost closely aligned with

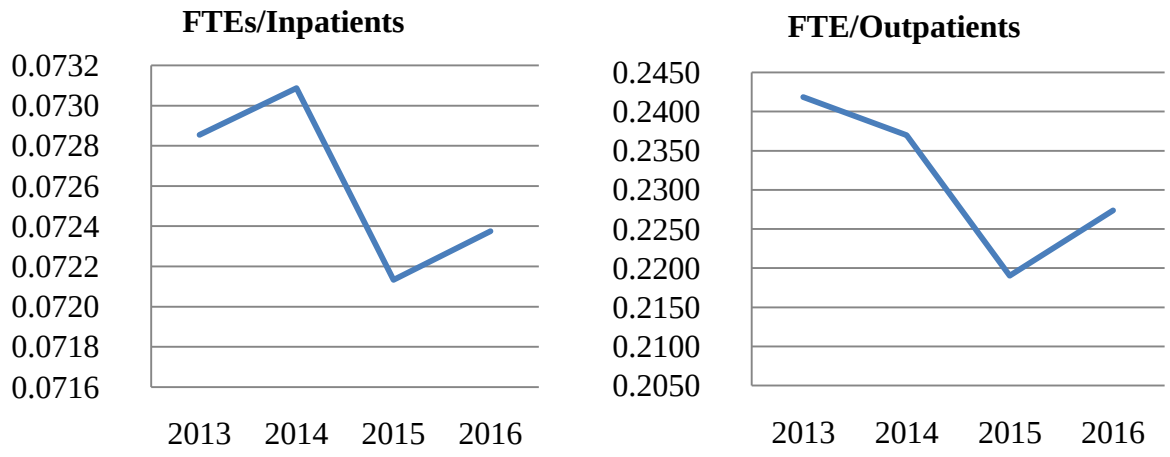
the budget (see Table 20). Moreover, the variance to the planned FTEs reduced significantly and the reduction in the ratio between FTEs and key outputs suggest that DHB tended to contain staff numbers (see Table 22 and Figure 18). Excepting medical staff and nurses, DHBs generally reduced allied personnel, support, and administrative staff (see Table 22). The same argument might also be applied to clinical costs, in which actual and budgeted amounts were closely aligned and, as a result of this, growth rate was contained.

Table 22. Full-time equivalent staff

	Actual FTE /Plan FTE (%)				Growth in FTE (%)				Growth in cost per FTE (%)			
	2013	**2014**	**2015**	**2016**	**14/13**	**15/14**	**16/15**	**Aver.**	**14/13**	**15/14**	**16/15**	**Aver.**
Medical	1.37	(0.47)	(0.69)	(1.39)	1.73	1.79	3.38	2.30	2.44	3.41	1.05	2.30
Nursing	1.22	1.66	1.78	1.73	2.29	2.08	2.16	2.18	1.80	(0.28)	1.85	1.12
Allied personnel	(2.15)	(1.24)	(1.89)	(1.82)	1.77	0.83	1.91	1.50	1.58	1.64	0.44	1.22
Support	3.81	(2.20)	(4.02)	(1.96)	4.55	(4.92)	3.25	0.87	1.87	2.12	1.08	1.69
Management	(1.15)	(0.68)	(0.52)	(1.21)	1.70	1.70	2.25	1.88	1.20	3.47	2.12	2.26
Total	**0.28**	**0.16**	**(0.01)**	**(0.13)**	**2.13**	**1.35**	**2.36**	**1.94**	**1.62**	**2.52**	**1.58**	**1.91**

Source: Authors' calculation using Annual schedule 4 - Average year to date consolidated accrued FTEs and Annual schedule 5 - Annualised average consolidated cost per FTE.

Figure 18. Ratio between FTEs and volume of activities



One might also argue that under-spending on payment to non-DHBs providers helped to drive efficiency (see Tables 20 and 21). However, reduction in funding for primary care might not be cost-effective in the long-term due to the likelihood of escalating demand for expensive treatments in secondary care (Controller and Auditor-General, 2018). A similar argument may be applied to the negative growth in capital charges which were mainly due to a reduction in charging rate in 2016/17 (Association of Salaried Medical Specialists, 2018). Interest also decreased markedly but it is important to be mindful that interest is a minor proportion of total costs (0.68%).

5.4.2 Bootstrap truncated regression with biased corrected scores

The second-stage regression result suggests that while budget accuracy has a negative effect, building, information technology, proportion of patients who identify as Maori, performance with respect to emergency targets, and outpatient variance are all positively associated with efficiency (see Table 23).

Table 23. **Regression results**

Variables	**Coefficients**	**Std. Err.**
Building	0.0443***	0.00616
Infortech	0.0220***	0.00590
Maori	0.0370***	0.00529
Emergency	0.0130**	0.00536
Exvar	-0.0086*	0.00505
Outvar	0.0131**	0.00531
Invar	0.0043	0.00518
Sigma	0.0307***	0.00323
Constant	0.9319**	0.00484
Observations	67	

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

The positive link between the nett value of building and information technology and efficiency suggests that more investment in these assets might be expected to better support service delivery. Effectively utilizing hospital buildings is critical not only because of their relative prominence as a fixed-asset item (72% of fixed assets are buildings) but also because of their potential impact on safety and capacity in providing care. Additionally, information systems and technology has become an important factor for DHBs to provide services more effectively and efficiently (Controller and Auditor-General, 2016b).

Although the nett value of buildings on the balance sheets increased, this does not necessarily mean that there was sufficient investment in buildings. In a review by the Controller and Auditor-General (2016a), it was found that buildings were mainly hospitals and that DHBs were planning more than $6 billion of capital expenditure on physical assets over the next 10 years (since 2015). The report also stressed that health assets had the lowest condition ratings in the public sector, and that only 80% of DHBs were actually carrying out their planned maintenance and renewal of buildings.

Indeed, we observe that the actual capital expenditure was only 65% of the forecast, on average. Collectively DHBs spent $1.106 billion less than budgeted for 2013/14 – 2016/17 suggesting that the value of fixed assets could have been greater if the plans had been executed. Buildings and plant, as well as information technology related assets were the main items having the largest unexecuted budgets (see Table 24). DHBs that inadequately upgrade, renew and replace fixed assets might not maintain a sound and sustainable service delivery and are likely to suffer from lower technical efficiency over time.

Table 24. **Capital expenditure**

Types of asset	Actual – Plan (NZ$1,000)				Actual/Plan (%)			
	2013/14	**2014/15**	**2015/16**	**2016/17**	**2013/14**	**2014/15**	**2015/16**	**2016/17**
Land	53	13,973	2,784	419			157.64	
Buildings and plant	(179,573)	(89,651)	(279,889)	(27,075)	49.76	69.34	47.51	93.83
Clinical equipment	(44,047)	(13,297)	(57,259)	(85,435)	72.40	88.27	63.61	47.36
Other equipment	(2,598)	(7,828)	(20,580)	(14,790)	81.32	48.55	25.56	39.69
Motor vehicles	1,757	(2,314)	(3,791)	(11,093)	150.67	68.56	56.82	32.78
Information technology	(25,028)	(17,448)	(34,708)	(15,369)	45.80	62.84	36.87	64.00
Software	(9,271)	(9,118)	(38,971)	(46,091)	82.28	80.01	43.58	38.98
Total	**(258,707)**	**(125,682)**	**(432,413)**	**(199,433)**	**59.12**	**75.87**	**49.48**	**73.77**

Sources: Authors' calculation based on Annual schedule 7- Capital expenditure

Since it impacts on the capacity to serve patients, we also examined the level of capital investment in fixed assets and healthcare resources in New Zealand in comparison with other countries having similar GDP *per capita* such as Italy, Spain, and Korea. Although the level of investment in fixed assets on healthcare sector was slightly higher than the OECD average (OECD, 2017), New Zealand was among the countries which appear to have relatively inadequate facilities for providing healthcare (see Table 25).

Table 25. New Zealand healthcare resources in comparison with OECD countries

Country	Hospital beds (per 1,000 population)	Computed Tomography scanners (per million population)	Magnetic Resonance Imaging units (per million population)	Positron Emission Tomography scanners (per million population)
United Kingdom	2.73	9.08	7.14	-
New Zealand	**2.77**	**16.74**	**12.37**	**1.24**
United States	2.88	42.09	36.87	5.08
Spain	2.99	17.55	15.19	1.58
Italy	3.31	33.25	26.14	2.96
Australia	3.78	55.88	11.90	2.54
OECD average	**4.80**	**25.54**	**15.10**	**2.05**
France	6.23	15.24	10.96	1.77
Korea	10.98	37.15	25.38	3.98
Japan	13.26	104.21	49.28	4.30

Sources: Authors' calculation using OECD's statistical databases (OECD.Stat, 2019b). The figures are average values over the period from 2010 to 2015 or 2016.

The emergency target requires the Emergency department (ED) to admit, treat, and discharge at least 95% of patients within six hours. A longer stay in ED might be associated with negative clinical outcomes (increased mortality and longer inpatient lengths of stay) and may indirectly reflect inefficiency in managing patient flow (Ardagh, 2015). In this respect, the positive association suggests that DHBs with higher performance in achieving the emergency targets could yield improved technical efficiency. Notwithstanding inevitable gaming, the emergency target does seem to have inspired positive initiatives towards improved performance by stimulating better cooperation with other services, developing new clinical practices and improved clinical governance (Ardagh, 2015; Tenbensel and Chalmers, 2016). Furthermore, there are unlikely to be any negative consequences arising from, say, compromised service quality

such as increased in-hospital mortality or growing patient admissions associated with achievement of this target (Ardagh, 2015).

Contrary to our assumption, the regression results show that a DHB with greater discharge of Maori patients appears to be relatively more efficient. McKendry *et al*. (1994, p. 54) found that: "while Maori use the services relatively more often, they are generally less resource consumptive than the non-Maori caseload". By conducting a two-sample t-test, we found that the average inpatient length of stay at DHBs with a higher proportion of Maori discharges were consistently shorter than those at the counterparts ($t(10) < -2.2$; $p<0.05$). Moreover, the data on hospital discharges show that Maori persons had the lowest proportion of patients aged over 65 when compared to other ethnic groups.

A high unfavourable variation to planned expenditures appears to reduce technical efficiency. It is reasonable to expect that DHBs used annual plans as an effective tool for management to achieve the targets set and therefore mitigated the variations to the budget. However, over the examined periods, more than 80% of DHBs overspent relative to their plans. Apart from additional services provided, which contribute to the mismatch between budgeted and actual results, DHBs might have failed to control the costs as forecasted. As the Controller and Auditor-General (2016) indicated financial management practices play an important role in sustainable health service provision. Without careful planning, organization and allocation, DHBs might struggle to ensure their viability, especially under conditions of uncertainty. Their examination of the variances between actual and planned results over the period 2008/09 to 2014/15 points to mixed evidence in the ability to operate as budgeted and diminished ability to manage uncertainty. Thus, those DHBs encountering unexpected events and underperforming with respect to saving targets are more likely to overspend relative to budget and therefore reduce technical efficiency.

While we assumed that efficiency increases when DHBs provide more services than planned, the regression result suggests that only an increase in outpatient volume to the

plan is statistically and positively associated with efficiency. Inpatients have relatively greater impacts on costs than do outpatients (Vitikainen *et al*., 2010) since they consume a premium of resources and thus an increase in the volume of inpatient services might generate greater uncertainty for management and therefore have deleterious implications for efficiency. On the other hand, DHBs might still be able to control their expenses within the budget or mitigate the magnitude of overspending when they treated more outpatients then planned. Moreover, when hospitals improved their clinical practices to admit and discharge patients in the same day, rather than increase the inpatients, the resources could be used relatively more efficiently (Gianino *et al*., 2018; Zeidler *et al*., 2008)

5.5 Conclusion

Healthcare service provision is a core function of the New Zealand government, consuming considerable financial resources. With budget constraints and growing demand, there is considerable interest in the question of how to achieve value for money without compromising service standards. Therefore, our empirical work should be of great interest to public policy architects and political decision-makers alike.

Generally, technical efficiency of DHBs increased over the examined periods. Moreover, it is likely that the formation of intermediary entities and containment of employee costs might have had a positive impact on efficiency. DHBs with greater investment in buildings and information technology, higher proportion of Maori patients, and better performance relative to emergency targets and outpatient plans were found to be more efficient. On the other hand, overspending relative to budget might be expected to reduce technical efficiency.

There thus exists potential scope for further improvements to efficiency. Specifically, DHBs should try to improve their ability to prepare budgets, manage expenditures closely

to the budget, and optimize patient flows at the Emergency department. Also, it appears that a greater reduction in back-office costs can be obtained when shared services entities have better communication and engagement with DHBs in delivering savings programmes (Controller and Auditor-General, 2015). In addition, asset management should be emphasized more strongly, particularly with respect to ensuring adequate investment in key assets such as buildings and information technology which are required to maintain sustainable service delivery and avoid escalating costs in the long-term. Our evidence suggests that DHBs should have a better understanding of their asset conditions and better planning for addressing maintenance, renewal and infrastructure backlogs. Indeed, it may prove fruitful for the Ministry of Health to direct DHBs to conduct regular and detailed long-term plans for asset stewardship. In addition, greater guidance on asset management may also enhance DHBs' performance (Controller and Auditor-General, 2016a).

While measuring the technical efficiency of DHBs as a whole provides an overarching picture, it would be interesting to disaggregate and evaluate the efficiency of each of the main services provided such as prevention services, early detection and management, intensive assessment and treatment, and rehabilitation and support. Future research might, therefore, be fruitfully applied in this direction. Another potential extension of this analysis might be to examine, in greater detail, the effects of the amenable mortality rate and factors related to lifestyle or prevalence of morbidity.

CHAPTER 6: INVESTIGATING THE TECHNICAL EFFICIENCY OF JAPANESE NATIONAL UNIVERSITIES FOLLOWING CORPORATIZATION: A TWO-STAGE DEA APPROACH

Following on from a radical restructuring in 2004, Japanese national universities have been operating in a challenging environment characterized by both reductions in government grants and reduced student enrollments. National universities consume considerable public resources and are a critical component of the knowledge economy, however little is known about the efficiency of these institutions. This study employed a two-stage Data Envelopment Analysis and found that both the size and disciplines of study offered by the universities were statistically significant determinants of efficiency. These findings suggest that efficacious public policy remedies might include measures to attract foreign student enrolments, merger of small universities and consolidation of inefficient departments.

6.1 Introduction

Incorporation of national universities in 2004 was one of the most radical reforms of the Japanese higher education since World War II (Yamada, 2014). Coupled with corporatization, national universities also experienced hard budget constraints and a steady decline in student enrolments (Ministry of Education, Culture, Sports, Science, and Technology (MEXT), 2010).

In essence, the corporatization program aimed at providing national universities with greater autonomy so that they could revitalize education and academic research (MEXT, 2010). Moreover, national universities were expected to be more efficient and gradually reduce their reliance on grants from the central government. In particular, the introduction of the National University Corporation Law of 2003 set up the legal framework for the reform including, *inter alia*, significant changes, such as: deregulation of institutional

budget, abolishing the public servant system[22], adopting private sector principles in management, and employing external people to participate on the board of directors (MEXT, 2003). The key motivations for these reforms were as responses to: (i) market-oriented policies and budget constraints, (ii) social critiques of failure in training competent human resources to cope with the economic recession, (iii) growing global competition in the university sector, and (iv) shrinking of Japanese youth populations (Mizobata and Yoshii, 2015).

Notwithstanding the fact that Japanese national universities now have greater discretion, they are still dependent on public funding. Although national universities are allowed to charge tuition fees (within the level stipulated by Ministry of Education, Culture, Sports, Science, and Technology), it is not always flexible and they still heavily rely on support from the government (around 43% of national university revenue takes the form of grants from the national government; Mizobata and Yoshii, 2015). This public funds support includes (i) management grants which cover necessary operational expenses for national universities to fulfill their functions, and (ii) capital grants to invest in facilities (MEXT, 2009).

Similar to many university sectors abroad, Japanese national universities have also been confronted with budget austerity in recent years. The deterioration of the Japanese economy over more than two decades (since the beginning of the 1990s) led to a decline in GDP per capita and an increase in the debt ratio, which engendered a reduction in public expenditure on education (Mizobata and Yoshii, 2015; OECD, 2018b). Indeed, the amount of national budget dedicated to national universities decreased by around one percent per annum over the period from 2004 to 2015 (Huang 2016). In an assessment of the situation after incorporation, a report prepared by Ministry of Education, Culture, Sports, Science,

[22] Prior to 2004, teachers and staff at national universities were civil servants.

and Technology noted that some national universities were operating under severe financial conditions (MEXT, 2010)

Japan has also been characterized by rapid aging and a shrinking population. Since the late 1970s, the youth population (aged under 15) has constantly decreased and this trend is projected to continue up to 2065 (National Institute of Population and Social Security Research, 2017). Specifically, the proportion of the population under 15 years of age in 2015 was just 12.6% - the lowest level on the globe (Statistics Bureau, 2015). Similarly, the 18-year-old population decreased by 42%, from its peak in 1992 (2.049 million) to 1.190 million in 2016. As a consequence, and further exacerbated by the rapid rise in the number of private universities in the 1990s[23] (Nemoto and Furumatsu, 2014), the higher education sector in Japan has been struggling to maintain sufficient enrollment numbers, especially given increasing national and global competition in the sector (Fukudome, 2019).

Japanese universities have been intensively scrutinized by the public on issues relating to learning and teaching quality, as well as on matters relating to cost-efficiency (Newby *et al.*, 2009). Indeed, there were various reports from the government as well as scholarly studies, which assess different aspects of national universities following corporatization (Christensen, 2011a; MEXT, 2010; Mirozumi, 2019).

Thus, in the context of more autonomy given from the reform, the budget constraints, the intensive competition, and growing public pressure, it is reasonable to expect that national universities were motivated to improve their efficiency. However, it seems that little is known about the technical efficiency of Japanese national universities. Indeed, while DEA has been widely applied to measure the performance of universities in many developed countries, there is a notable paucity of studies in the context of Japanese public universities.

[23] Private universities make up more than three-quarters of the total universities in Japan.

We fill this gap in the literature by evaluating the technical efficiency of national universities – a key component of public higher education in Japan. The publicly owned universities in Japan are divided into two categories on the basis of how they are founded: (i) national universities which are originally established and funded by the Japanese central government, and (ii) public universities which are established and funded by local public entities or public university corporations (MEXT, 2009). Although having an equivalent number of institutions, the number of students enrolled in national universities is more than four times higher than that of public universities. It thus represents an important avenue for scholarly inquiry.

Japanese national universities have a responsibility to prepare plans to achieve the targets set in the six-year plan approved by MEXT. The six-year mid-term plan (phase II) commenced after March 31st, 2010 (MEXT, 2010). Therefore, we chose the period between 2010 and 2016 for this study to avoid possible inconsistencies in policies applied in different periods. We employed the well-attested technique of global intertemporal data envelopment analysis (DEA) to estimate the relative technical efficiency of each university over the period of analysis. We also conducted a second stage analysis using a double bootstrapped truncated regression model to identify the possible determinants of efficiency. In addition, in order to provide more evidence to complement the findings and fortify the policy implications, content analysis is used to explore the assessment of the achievements of the incorporation of national universities and the performance following structural changes. More particularly, the research questions revolve the three key issues: (i) whether the expected goals of incorporation were achieved, especially the discretion in managing national universities and renovation of education and research? (ii) How did national universities respond to the reduction of government grants? and (iii) What are the suggestions for improving the performance of national universities? (See Appendix 3 for detailed questions)

The remainder of the paper is organized as follows: In the next section, we review the extant literature focused on measuring efficiency at the university level along with the environmental factors affecting efficiency. Following this, we describe our empirical methodology with particular emphasis on justifying the model specification. The results section presents our empirical estimations. We conclude by proposing some public policy implications and suggestions for further research.

6.2 Literature review

Data Envelopment Analysis has been used extensively to measure the technical efficiency of public services including education which is among the five most popular applications (Emrouznejad and Yang, 2018). DEA is particularly suitable for measuring the efficiency of education owing to its capacity to simultaneously incorporate multiple inputs and outputs without the requirement for price information that is often absent for public services (Johnes and Tone, 2017). Since DEA literature on education covers many different objectives (De Witte and López-Torres, 2017), we confine our discussion to relevant studies which measure efficiency at the level of the university and discuss possible factors associated with technical efficiency which we later use to inform our empirical strategy.

6.2.1 Prior research on university efficiency over time

A number of DEA studies have examined how efficiency of higher education institutions (HEIs) respond to policy changes, especially those related to financial pressure. Over the period 1980/81–1992/93, British universities suffered a considerable contraction in government grants in an operating environment of increased student enrolments (Fleg *et al.,* 2004). Studies found that technical efficiency substantially increased over the period of policy change and it has been suggested that fiscal pressure from cuts in grants strongly incentivized British universities to improve their efficiency (Flegg *et al.,* 2004). In a similar vein, Australian universities had also been under pressure to improve efficiency

since the late 1990s due to reductions in public grants per student, rising capital expenditure need, increased competition for research funding, and growth in the number of international students. By investigating the productivity growth of Australian universities from 1998 to 2003, Worthington and Lee (2008) estimated that the average productivity growth was 3.3% which mainly came from technology change, although increased research outputs also proved important. Likewise, Moradi-Motlagh *et al.* (2016) analyzed 37 Australian universities in the period 2007 to 2013 and also found a similar result. In an American context, Sav (2016) investigated 250 US universities covering the period before and after the Great Recession and found that efficiency increased slightly during this time (in 2007, 2010, and 2011, particularly). In sum, these studies suggest that universities might be expected to respond to austerity and hard budget constraints by generating income from other sources (e.g. tuition fees; Flegg *et al.,* 2004), utilizing information technology, and diversifying teaching activity (e.g. online and multi-campus delivery, out-of-semester enrolments; Worthington and Lee, 2008).

DEA has also been employed to evaluate the policies related to university consolidation. For example, in a quest to find empirical evidence to support the government policy of larger-sized and more-specialized universities, Glass *et al.* (2005) studied the efficiency of 98 non-specialist UK universities in 1996. The results suggest that UK universities might improve efficiency by placing greater emphasis on either teaching or research (rather than pursuing both goals), or by pursuing mergers (expected to yield efficiency gains through economies of scale). In more recent research, Papadimitriou and Johnes (2018) found that while mergers were positively associated with efficiency, this impact only occurred in the year following the merger. Another example is the work of Thanassoulis *et al.* (2011) who examined the cost structure of 363 UK universities at the time that the UK government was encouraging increased student enrolments (2000-02). The results suggested that universities might reap economies of scale by increasing the number of students by 20 to

27 percent. On the whole, these studies suggest the importance of measuring size effects in studies of university technical efficiency.

Although the universities analyzed from abroad also experienced budget restraints, financial crisis or reforms, they were not confronted with a rapid reduction in the student population over the examined periods (OECD, 2019). Also, it should be noted that many of the extant studies are based on Anglo-sphere countries which offer programs in English and thus have greater potential to attract overseas students. By way of contrast, Japanese national universities have been characterized by some rather unique challenges. First, student enrollments in national universities have decreased in all years, whilst public university enrollments have consistently increased and private universities have recovered following modest decreases which occurred between 2012 and 2014 (see Table 26). Second, due to the overexpansion of private universities in the 1990s (Nemoto and Furumatsu, 2014), the number of higher education institutions (HEIs) in Japan is relatively higher than that of most developed countries (Japan is among the countries that have the highest university density and the lowest number of student per HEI (see Table 27)). Finally, although the number of classes in English has been expanded (The Japan Association of National Universities, 2019), language barriers are likely to still be an obstacle to attract overseas students and promote internationalization (Yamato, 2018). Therefore, Japanese national universities provide an important case study to investigate how technical efficiency might be affected by these unique challenges.

Table 26. **Number of students by institutional type**

Year	National university		Public university		Private university	
	No. Institutions	Student enrolled	No. Institutions	Student enrolled	No. Institutions	Student enrolled
2010	86	625,000	95	143,000	597	2,120,000
2011	86	623,000	95	144,000	599	2,126,000
2012	86	618,000	92	146,000	605	2,112,000
2013	86	614,783	90	146,160	606	2,107,929
2014	86	612,509	92	148,042	603	2,094,978
2015	86	610,802	89	148,766	604	2,100,642
2016	86	610,401	91	150,513	600	2,112,710

Sources: Ministry of Education, Culture, Sports, Science, and Technology.

Table 27. **Number of students and HEIs at selected OECD countries**

Country	No. students	No. HEIs	No. students/ HEI	HEIs/million population (0-19)	HEIs/1,000 km^2
Australia	1,903,454	91	20,917	14.74	0.01
Belgium	504,745	64	7,887	24.85	2.10
Chile	1,221,774	64	19,090	12.81	0.08
Czech Republic	395,529	50	7,911	23.96	0.63
France	2,424,158	371	6,534	23.81	0.67
Germany	2,977,781	353	8,436	23.04	0.99
Italy	1,826,477	98	18,638	8.89	0.33
Japan	**3,845,395**	**782**	**4,917**	**35.20**	**2.07**
Korea	3,268,099	264	12,379	27.28	2.63
Mexico	3,515,404	1,600	2,197	35.78	0.81
Netherlands	842,601	60	14,043	15.80	1.43
Poland	1,665,305	395	4,216	52.75	1.26
Portugal	337,507	114	2,961	58.58	1.24
Spain	1,963,924	109	18,018	12.02	0.22
United Kingdom	2,330,334	254	9,175	16.37	1.05
United States	19,531,727	2,000	9,766	24.21	0.21
OECD16	**48,554,214**	**6,669**	**7,281**	**26.17**	**0.29**

Sources: Authors' calculation based on data sources: for student number (2016 data) (OECD.Stat, 2019a); for university numbers (2018 data) (UNESCO-WHED, 2018); for the population (2017 data) (United Nation-Population Divison, 2019); for the area (World Population Review, 2019)

6.2.2 Determinants of variation in the efficiency of universities

In addition to using DEA as an evaluation and benchmarking tool a semi-parametric method has also been employed, in second-stage analysis, to understand the determinants of efficiency scores. Explanatory variables found in the extant literature can be categorized into three groups: (i) characteristics of universities (e.g. faculty compositions, staff structure, and foundation year), (ii) social-economic indicators (e.g. GDP per capita, unemployment rate) and (ii) funding related factors (e.g. share of government funding, proportion of tuition fees).

The impacts of science faculties, such as medicine and engineering, have attracted the attention of scholars with mixed findings. As medical or engineering education normally requires a longer training time and consumes more resources, universities with these faculties are often found to be less efficient (see Agasisti and Pohl, 2012; Barra *et al.*, 2018; Kempkes and Pohl, 2010). However, Papadimitriou and Johnes (2018) and

Wolszczak-Derlacz and Parteka (2011) found that the presence of medical or pharmacy faculties are positively associated with efficiency since medical courses are vocational in nature and tend to attract good students (high entry scores) which might have a positive influence on efficiency (Papadimitriou and Johnes, 2018). This literature suggests that it will be important to test whether universities with science faculties might be less efficient given that they require relatively higher resources (especially with respect to expensive equipment).

Regarding staff structure, a higher ratio of full-time staff over part-time staff has been found to have a positive and significant effect on efficiency (Quirora-Martínez *et al.*, 2018). Thus it is important that investigations of the determinants of efficiency include the proportion of full-time teachers over the total number of teachers.

GDP per capita has been found to have both positive and negative effects on efficiency in the university sector literature (Kempkes and Pohl, 2010). While Barra *et al.* (2018), Kempkes and Pohl (2010), and Wolszczak-Derlacz (2017) found that GDP per capita had a positive link with efficiency, this impact was not statistically significant in the other studies (Agasisti and Pohl, 2012; Agasisti and Wolszczak-Derlacz, 2016). However, given the likelihood that local economic conditions may affect demand for university courses, it is important to test for an association between measures of income and university technical efficiency when investigating the determinants of the same.

The effect of sources of revenue on university technical efficiency has also been examined in the literature, focusing on the share of funding from the government. The idea behind this putative link is that the level of autonomy is inversely related to the dependence on public funding which is assumed to reduce efficiency (Aghion *et al.*, 2010; Christensen, 2011b). As indicated by Agasisti and Wolszczak-Derlacz (2016); Papadimitriou and Johnes (2018); Wolszczak-Derlacz (2017), an increase in the proportion of funding coming from the government, or greater share of the revenue from non-competitive

resources, is likely to be negatively associated with technical efficiency. However, in a study of US public universities, Sav (2013) found that efficiency could be improved by increasing government funding. Moreover, as we have noted earlier, recent reforms in the Japanese national universities have involved significant cuts to grant funding. We, therefore, include the proportion of grant funding as a regressor to investigate their effect on the relative technical efficiency of national universities.

6.3 Methodology

6.3.1 DEA model

DEA is a non-parametric method developed by Charnes, Cooper, and Rhodes in 1978 (CCR) which uses a linear programming technique to optimize the outputs/inputs ratio with unknown weights. The technical efficiency of each Decision-Making Unit (DMU) is estimated relative to other best-observed practices in the sample examined. In the educational context, a university is considered to be efficient when using the least resources to provide a given level of outputs or alternatively maximize the outputs with given available resources. Thus, assuming that a university (z_i) using p inputs (x_j) to provide q outputs (y_i), a ratio form of DEA is as follows:

$$MaxH_0 = \frac{\sum_{i=1}^{q} u_i y_{i0}}{\sum_{j=1}^{p} v_j x_{j0}}$$

$$\text{Subject to:} \frac{\sum_{i=1}^{q} u_i y_{iz}}{\sum_{j=1}^{p} v_j x_{jz}} \leq 1; z = 1, \dots, n$$

$$u_i, v_j \geq 0 \; for \; all \; i \; and \; j$$

Where:

H_0= the efficiency score of univeristy$_0$;

x_{jz}= the amount of input j used by z_{th} university;

y_{iz}= the amount output i provided by z_{th} university;

u_i= weight assigned to output i;

v_j= weight assigned to input j

Basically, the specification of a DEA model relates to the two assumptions ("orientation" and "returns to scale") and the choice of variables (inputs and outputs). The CCR model assumes constant returns to scale (CRS), implying that inputs and outputs increase or decrease proportionally. However, in 1984, Banker, Charnes, and Cooper (BCC) extended the CCR model by relaxing the CRS assumption and allowing for variable returns to scale (VRS) which recognizes the existence of economies or diseconomies of scale. With regard to "orientation", an input-oriented model means that inputs are minimized with a given level of outputs; on the contrary, the output-oriented model will aim at maximizing the outputs with a fixed level of inputs.

In practice, the choice of orientation depends on the ability of DMUs to control inputs or outputs while the assumption on returns to scale is contingent on the researcher's question. Since corporatization with a reduction in subsidies for operating expenses, Japanese national universities have been expected to make efforts to minimize the resources used and have relatively less control over outputs (MEXT, 2010). This suggests that an input-orientation is the most appropriate approach for measuring Japanese national university technical efficiency. Moreover, because we are interested in estimating the effect of size on university efficiency in our second-stage regressions, a CRS orientation is necessary (a VRS model would have already imputed size effects and thus introduce avoidable bias into the resultant estimates). We, therefore, estimate efficiency scores through a CRS input-orientated model. Also, since some national universities are not included in our sample, we bootstrapped our results with 1,000 replications to verify the accuracy and reliability of the efficiency scores estimated.

There were two components to the empirical work conducted. First, we ran a global intertemporal DEA to understand the trend in efficiency over time. Global intertemporal DEA essentially pools all DMUs for all years (after first inflating all pecuniary data to

2015 using the education sector Consumer Price Index (CPI)) into a single DEA analysis. DEA scores are then regrouped in each year to derive the mean scores so that comparisons can be made across time. Following this, we conducted a second-stage regression of national university efficiency so that we could understand the determinants of this aspect of performance.

6.3.2 Sample and variables selection

DEA estimates the relative efficiency of a DMU by comparing with other *homogenous* counterparts. Therefore, although there are 86 national universities, we excluded institutions that had distinctive remits (four medical universities, two specialized universities, and four universities that do not admit undergraduate students). Thus, after adjusting for missing data for two other universities, the sample employed was 74 national universities. Data regarding teachers, support staff, and students were collected from the documentation of each university, whereas financial data was purchased from Toyo Keizai Inc.

Higher education institutions use resources (e.g. employees and money) to fulfill two major missions – education and research. In a review of 223 papers on efficiency in education from 1977 to 2015, De Witte and López-Torres (2017) indicated that, at an institutional level, the most widely employed inputs include educational resources (e.g. grants), expenditures, and personnel (academics, administrators, and support staff) data. This study, therefore, employs three inputs including teacher cost, support staff cost, and other operating expenses. We disaggregated staff expenditure data because the mix of teaching staff to support staff is likely to have important effects on technical efficiency given that outputs are most directly related to the former.

Regarding teaching outputs, the number of graduates is the most widely used measure (Agasisti and Wolszczak-Derlacz, 2016; Barra *et al.*, 2018; Quirora-Martínez *et al.*, 2018).

Because students dropping out still receive teaching services enrolments rather than graduations is the most accurate measure of actual outputs (Carrington *et al.*, 2005). Moreover, Japanese universities have the highest graduation rate (93%) among the OECD countries (OECD, 2011). Therefore, in this study, the number of enrolled students disaggregated by the level of study (undergraduate, master, and doctoral), were employed as a proxy for teaching output to accurately reflect resources consumed. Notably, we disaggregated the different teaching outputs to reflect the fact that the resources deployed to teach the various levels of students tend to vary considerably.

With respect to research output, while Agasisti and Pohl (2012), Barra *et al.* (2018), and Kempkes and Pohl (2010) employed research grants, most other scholars used the number of publications (Quirora-Martínez *et al.*, 2018; Visbal-Cadavid *et al.*, 2017; Wolszczak-Derlacz and Parteka, 2011). In this study, we also use the number of publications[24] indexed in the Web of Science (specifically the number of journal articles, proceedings papers, and book chapters). Also, to adjust for the delay between the time of conducting research and the time of publication, we used a two-year moving average method to calculate the number of publications.

Apart from teaching and research, a third mission has also been considered as an output of higher education institutions (Johnes, 2015). Indeed, it is a striking characteristic of Japanese national universities that around half of them have affiliated hospitals. Therefore, we included income from attached hospitals as an output[25]. Descriptive statistics of inputs and outputs are as follows:

[24] We also ran estimations based on research income, by way of a robustness check, and found that the results (available from the corresponding author) were very similar.

[25] Although income from donation can be considered as a possible output, we did not include in our model since this is not the key activity of a university and we found that this source of income just made up a minor proportion over total income with about 2.3%.

Table 28. **Variables used in the DEA model**

Variable	Description	Mean	Std. Dev.
Inputs			
Teacher cost	The value of payment for teachers (million ¥)	9,984	9,466
Staff cost	The value of payment for support staff (million ¥)	7,193	7,328
Other operating expenses	The value of other operating expense (million ¥)	17,266	21,143
Outputs			
Undergraduate	Number of students at the undergraduate level	5,947	3,465
Master	Number of students at master degree level	1,343	1,419
Doctor	Number of doctoral students	655	1,026
Publications	The total number of publications (articles, proceeding papers, and book chapters)	891	1,468
Hospital income	The value of income from attached hospitals (million ¥)	11,098	12,024

6.3.3 Regression model

The relative efficiency scores derived from our global intertemporal DEA could help to identify efficient and inefficient units. However, from the managerial and public policy viewpoint, there is a need to also understand the determinants of efficiency. Accordingly, we also conducted two-stage regressions (Førsund 2018). Based on the prior literature (discussed earlier) and the context of Japanese national universities, we employed the following regressors (i) government grants as a proportion of total revenue, (ii) proportion

of full-time teaching staff, (iii) university size (measured by the total student numbers), (iv) whether or not the university had an attached hospital, (v) the presence of a science faculty (medicine, engineering), and (vi) resident income per capita at the prefecture where the universities are located.

Different approaches have been proposed for second-stage regressions of DEA scores, including ordinary least square (OLS) (Banker and Natarajan, 2008; McDonald, 2009), Tobit regression (Hoff, 2007), and bootstrapped truncated regression (Simar and Wilson, 2007). Moreover, there has been little consensus in choosing the regression model. In recent DEA studies on higher education institutions, it appears that the bootstrapped truncated regression model was the most common approach (Agasisti and Wolszczak-Derlacz, 2016; Andersson *et al.*, 2016; Barra *et al.*, 2018; Wolszczak-Derlacz, 2017). We, therefore, adopted a double bootstrap truncated regression model (or "algorithm # 2") to identify determinants of universities' efficiency (see, Simar and Wilson (2007) for required assumptions and Badunenko and Tauchmann (2018) for further explication). Accordingly, the biased corrected scores ($\widehat{\theta_i^{bc}}$) are regressed against independent variables (Z_i) in the double bootstrap truncated regression model given by:

$$\widehat{\theta_i^{bc}} = \beta Z_i + \gamma_i, \quad i = 1, \dots, n$$

Where:

$\widehat{\theta_i^{bc}}$ is bias-corrected technical efficiency scores and internally estimated from the specified inputs and outputs.

Z_i is the set of contextual factors.

Descriptive statistics for these variables are provided in Table 29.

Table 29. **Explanatory factors used in the regression model**

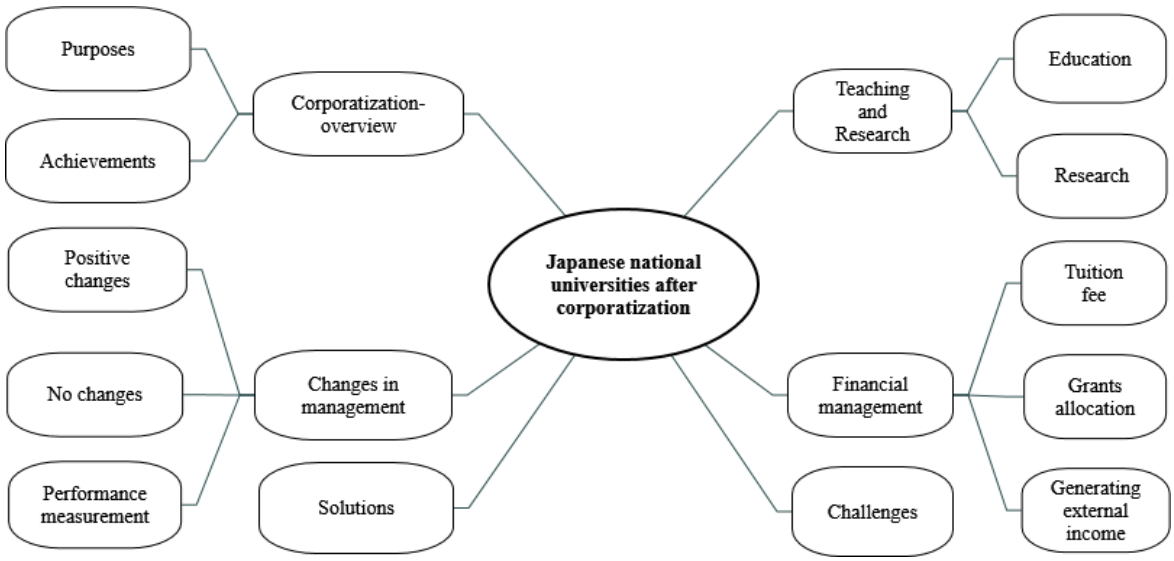
Variable	Description	Mean	Std. Dev.
Proportion of Full-time Teachers	The proportion of full-time teachers over the total number of teachers (%)	60.97	17.13
Proportion of Operating Grants	Share of revenue derived from government grants (%)	43.06	13.44
Resident Income per capita	Resident income per capita by prefecture (¥'000)	3,012	882
Tangible fixed assets	The average value of net fixed assets per student (¥'000)	12,140	6,990
Investment and other fixed assets	The average value of investment and other fixed assets per student (¥'000)	115	156
Current assets	The average value of current assets per student (¥'000)	1,265	848
Size	A dummy variable in which universities are classified into three categories: 0, 1 and 2 are assigned for small, medium, and large size, respectively		
Attached hospital	A dummy variable set to 1 if a university has an attached hospital		
Science Faculty	A dummy variable equals 1 if a university has a science faculty (engineering, pharmacy, etc), 0 otherwise		

6.3.4 Content analysis

We first constructed a list of questions focusing on the changes in management styles and the impacts on education and research. Also, as financial distress has been a striking feature, influences of the reduction in government grants and the responses of national universities are worth to examine. Moreover, under these circumstances, what are the lessons or the solutions to cope with possible challenges and to improve the performance of national universities? Questions were sent to the possible participants and asked for their acceptance to be as interviewees. Upon receiving their agreements, the arrangement for interviews is scheduled. The interviews were conducted at the interviewees' offices, audio-recorded, and then transcribed (Japanese transcripts were translated into English by

Google translate). The interview transcripts were then imported to the Nvivo software for analysis. Unlike the qualitative content analysis basing on a predetermined theory framework, this study applies an inductive approach to identify the answers for the research questions mentioned above. Accordingly, the interviews' contents are categorized as follows:

Figure 19. Categories of the interviews' contents



6.4 Results and discussion

6.4.1 Efficiency estimates

Mean technical efficiency scores from the first stage are reported in Table 30 and the efficiency trends are illustrated in Figure 20.

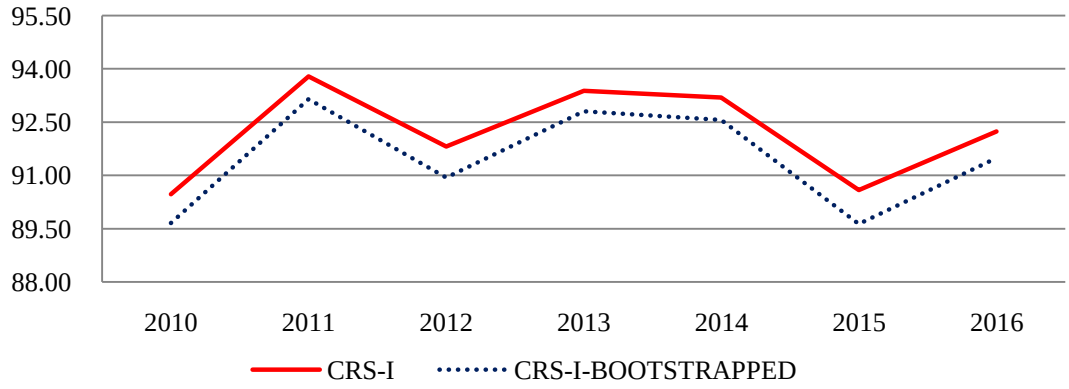
Table 30. Technical efficiency scores

Model	Measure	2010	2011	2012	2013	2014	2015	2016
CRS-I	Mean	90.47	93.78	91.81	93.38	93.19	90.59	92.24
	Min	59.22	61.43	57.49	61.12	57.90	60.36	62.20
	Std. Dev.	8.98	8.10	8.63	8.61	9.57	9.42	9.41
CRS-I-B	Mean	89.66	93.15	90.94	92.80	92.56	89.63	91.48
	Min	58.26	60.72	56.74	60.15	57.15	59.71	61.65
	Std. Dev.	9.03	8.25	8.65	8.81	9.76	9.42	9.55

Notes: CRS - Constant returns to scale; I-Input oriented; B- Bootstrapped

Generally, the technical efficiency of Japanese national universities fluctuated quite a lot over the period of analysis, with no discernable trend (and somewhat contrary to the evidence from abroad).

Figure 20. **Efficiency trends**



An analysis of growth in inputs and outputs suggests that fluctuating efficiency was probably mostly attributable to changes in teacher cost (see Figure 21). In addition, the relatively strong growth in support staff cost and other expenditures, along with a steady decline in the number of students in all courses, may have prevented universities from improving technical efficiency. Indeed, the number of students declined by 13,500 to 580,500 in 2016 from 2010 levels, which is equivalent to a decrease of 0.38% per year. Given the shrinking university-aged population in Japan, it seems that national universities may have struggled to streamline resources commensurate with the declining trend in student numbers.

Figure 21. **Efficiency trends and changes in teacher cost**

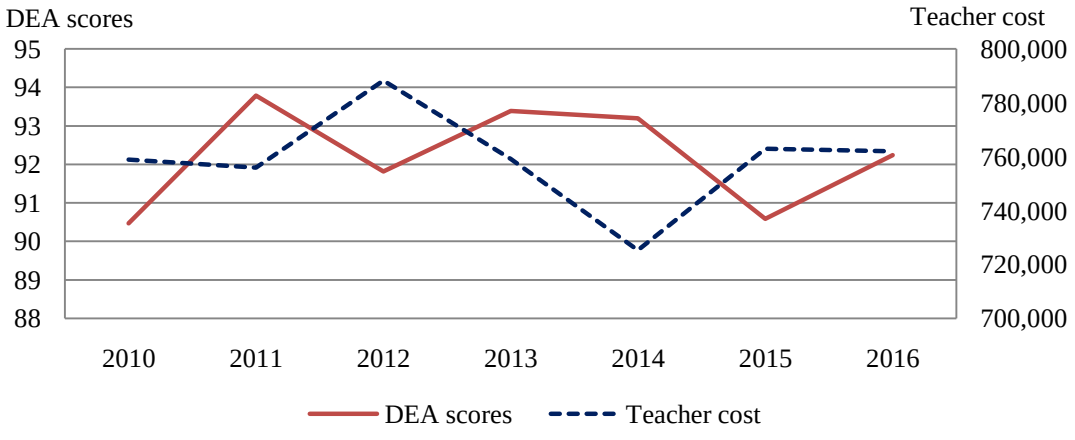


Table 31. **Input and output growth**

	2010	2011	2012	2013	2014	2015	2016	Growth
Inputs								
Teacher cost (mill. ¥)	758,878	755,932	788,185	759,124	725,297	762,887	761,934	0.07
Staff cost (mill. ¥)	520,326	533,969	547,786	550,766	551,591	572,358	578,615	1.79
Other operating expenses (mill. ¥)	1,219,279	1,235,511	1,282,079	1,316,236	1,388,633	1,421,346	1,420,180	2.57
Outputs								
No. Undergraduates	443,157	442,241	440,134	439,348	438,571	436,861	435,401	(0.29)
No. Master students	102,060	102,549	99,921	98,045	96,928	96,694	97,312	(0.79)
No. Doctoral students	48,830	49,302	48,757	48,234	47,913	47,885	47,835	(0.34)
Hospital income (mill. ¥)	729,397	808,572	840,388	877,135	894,653	904,942	928,969	4.11
No. Publications	63,271	64,307	65,952	66,741	66,612	68,905	71,342	2.02

Notes: Authors' calculation. Costs and revenue are deflated to 2015 value using the CPI in the education sector. The growth is the geometric mean of the entire examined period.

In order to get greater insight into national university cost-pressures, we further investigated the cost structure in Tables 32 and 33. As can be seen, medical expense is the second largest item behind employee cost and makes up about 21.30% of total cost or 42% of other operating costs. Moreover, medical expense had the second largest growth rate over the period of analysis, which suggests that national universities with hospitals were relatively disadvantaged in their efforts to contain costs over the period under analysis.

Although the contexts are different, it appears that institutions abroad were able to respond to cuts in grants through increases to efficiency (Flegg *et al.*, 2004; Worthington and Lee, 2008). However, it should be noted that, unlike Japan, the number of enrollments in higher education institutions in the UK and Australia rapidly increased (OECD, 2019). Moreover, probably thanks to the advantage of English speaking countries, the proportion of foreign students at these two countries was around 18% which is much higher than that of Japan (4%) (OECD, 2018a). It seems likely that the different context of dwindling student enrollments along with an inability in controlling costs might explain why Japanese

national universities were less successful in their efforts to respond to reductions in grants than their international peers.

Table 32. **Detail expenses of national universities (mill. ¥)**

	2010	2011	2012	2013	2014	2015	2016	Growth
Teacher cost	758,878	755,932	788,185	759,124	725,297	762,887	761,934	0.07
Staff cost	520,326	533,969	547,786	550,766	551,591	572,358	578,615	1.79
Medical expenses	483,128	518,769	543,621	567,672	590,747	607,079	619,884	4.24
Research expenses	382,278	393,898	409,427	408,446	438,806	453,557	457,588	3.04
Educational expenses	155,580	137,808	145,306	157,712	174,043	174,369	166,209	1.11
Others	198,293	185,037	183,725	182,406	185,038	186,341	176,500	(1.92)
Total operating expenses	2,498,484	2,525,412	2,618,050	2,626,126	2,665,521	2,756,591	2,760,729	1.68

Notes: Authors' calculation. The figures are deflated to 2015 value using CPI in the education sector.

Table 33. **Cost structures of national universities (%)**

	2010	2011	2012	2013	2014	2015	2016	Average
Teacher cost	30.37	29.93	30.11	28.91	27.21	27.68	27.60	28.79
Staff cost	20.83	21.14	20.92	20.97	20.69	20.76	20.96	20.90
Medical expenses	19.34	20.54	20.76	21.62	22.16	22.02	22.45	21.30
Research expenses	15.30	15.60	15.64	15.55	16.46	16.45	16.57	15.96
Educational expenses	6.23	5.46	5.55	6.01	6.53	6.33	6.02	6.02
Others	7.94	7.33	7.02	6.95	6.94	6.76	6.39	7.03

6.4.2 Regression results

There are three inputs and five outputs specified in the DEA model and five explanatory factors used in the second-stage analysis. Also, it should be noted that as we specified an input orientation in the regression model, a positive sign of the estimated coefficient means that, *ceteris paribus*, an increase in a variable will generate a higher efficiency, and vice versa (Badunenko and Tauchmann, 2018).

Table 34. **Regression results**

Variables	Coefficients	Standard errors
Proportion of Operating grants	-0.115***	0.015
Proportion of Full-time teachers	-0.004	0.005
Medium Size (dummy)	-0.006	0.014
Large Size (dummy)	0.10***	0.019
Attached hospital (dummy)	-0.129***	0.028
Science Faculty (dummy)	-0.038***	0.014
Resident Income per capita	0.024***	0.008
Tangible fixed assets	-0.053***	0.009
Investment and other fixed assets	0.019***	0.007
Current assets	0.015*	0.009
Sigma	0.069***	0.004
Constant	.098***	0.023
Observations (n)	421	

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

Table 34 details the results of our regression. The results suggest that as the proportion of operating grants increased, relative technical efficiency decreased (at the 1% level of statistical significance). Notwithstanding a general decline over the period of analysis, grant revenues were still the most important income item, suggesting a considerable reliance on financial support from the government.

Similar to findings in prior literature (Agasisti and Wolszczak-Derlacz, 2016; Wolszczak-Derlacz, 2017; Wolszczak-Derlacz and Parteka, 2011), our result implies that those universities which rely less on operating grants tend to be more efficient. Possibly, the reduction in funding from the government might have motivated universities to use the resources more efficiently. In addition, funding distributions (e.g. operating grants, financial support for leading initiatives) are not always transparent due to difficulties in performance evaluation and equivocal processes (Mirozumi 2019). For instance, other factors might affect funding allocations such as when some universities have privileged status (e.g. former imperial universities) or receive additional funding for implementing

national projects (e.g. Global 30, Top Global University Project) (Huang, 2018; MEXT, 2010). Therefore, apparent relative inefficiency might occur as either a result of the vagaries in funding allocations or due to reduced effort of the kind suggested by the well-known flypaper effect[26].

In order to examine the effect of size on efficiency we generated three strata of universities - small universities (the first quartile by student number; 19 universities had student populations under 4,500), medium (second and third quartiles; 36 universities) and large (fourth quartile; 19 universities with student populations over 10,000)[27]. The regression results suggest that large national universities (with enrolments greater than 10,000 students) are likely to be more efficient. This improved efficiency is likely to be a result of economies of scale. For instance, an increase in the number of students might be met by expanding the size of the class without recruiting more teachers. Indeed, Daraio *et al.* (2015) note that the extant literature confirms pervasive economies of scale in higher education, especially in undergraduate teaching. Moreover, in the context of Japanese higher education, Hashimoto and Cohn (1997) and Nemoto and Furumatsu (2014) both found evidence in support of the presence of economies of scales in private universities, (but none for public institutions). Thus our results are an important addition to the existing scholarly work.

National universities with attached hospitals were significantly less efficient than their peers. More than half of the examined national universities have affiliated hospitals, which is a rather unique characteristic of Japanese national universities. In the previous analysis on hospital services, we found that the cost growth associated with the hospitals outweighed income growth and this will have been a drag on overall efficiency. The regression analysis confirms our reasoning regarding the negative impact of the attached

[26]The "flypaper effect" - "money sticks where it hits" – suggesting that an additional government grants might result in a greater public spending (see Bailey & Connolly, 1998; Hines & Thaler, 1995)

[27]We used strata to avoid introducing bias (which may have occurred as a result of also employing disaggregated student numbers in the DEA).

hospitals. Along with training and researching missions, hospital universities are perceived as “a last bastion” which provide the most advanced medical treatments to treat intractable diseases in the regions where they located (MEXT 2010). This means that they have an implied responsibility to continuously innovate state-of-the-art medicine and medical technology which is generally very costly. A respondent in our interviews also stressed the difficulty in improving efficiency of attached hospitals: “We have been working desperately to improve efficiency where possible. Anyway, many medical schools and engineering-related things are expensive. But without it, it is a story that you cannot do research unless you put it in, or that you cannot do surgery, so it is inevitable” (NU3). Also, the interviewee expressed that they could not just put more priority on profitable areas since resources must be evenly allocated. In addition, while public hospitals enjoy high-level subsidies for maintaining facilities (50% of total cost), hospital universities are only subsidized for around 10% of their costs, thus making facility maintenance burdensome (National University Hospital Council of Japan, 2014b). Moreover, it seems that the fees for treatment may not have kept pace with increases in inputs (MEXT, 2010).

Similar to the findings by Agasisti and Pohl (2012) and Barra *et al.* (2018), science faculties are negatively associated with efficiency. It seems that relatively longer course duration time and much higher resources requirements (for experiments and the like) might reasonably explain the negative association for science faculties. Likewise, the positive impact of income per capita is consistent with the findings by Barra *et al.* (2018), Kempkes and Pohl (2010), and Wolszczak-Derlacz (2017). Probably, those national universities locating in the regions with more favorable economic conditions might be more efficient thanks to the advantages in attracting more students or collaborating with other institutes in doing research.

Types of assets are found to be significantly associated with technical efficiency in which tangible fixed assets have a negative influence while investment and other fixed assets and current assets have a positive impact. Notably, tangible fixed assets make up 98% of the

total fixed assets value suggesting that this kind of assets is the main source driving the depreciation and other maintenance expenses. Therefore, a university having higher value of tangible fixed assets per student appears to be less efficient. As a complement analysis, we also examined the link between accumulated depreciation and efficiency scores. The DEA scores for the period from 2012 to 2016[28] and each year will be categorized into two groups: low efficiency and high efficiency to compare the differences in the accumulated depreciation ratio. Similarly, accumulated depreciation ratio is ranked to categorize into high and low groups to identify the differences in efficiency scores. The results suggest that national universities that are more efficient having a higher accumulated depreciation ratio and vice versa. Thus, universities have assets that have been already fully depreciated will have lower net value and lower depreciation costs; on the contrary, those have low accumulated depreciation implying that they have recently invested in fixed assets so the depreciation cost might affect the technical efficiency.

Table 35. **T-test results based on efficiency scores**

	2012	**2013**	**2014**	**2015**	**2016**	**2012-2016**
High efficient group	45.63	48.77	48.91	51.03	55.21	49.62
Low efficient group	43.62	46.85	46.70	48.85	51.36	47.14
Difference	2.01	1.92	2.21	2.18	3.85	2.44
T-test (unequal)	Pr(T > t) = 0.0708	Pr(T > t) = 0.0620	Pr(T > t) = 0.0379	Pr(T > t) = 0.0292	Pr(T > t) = 0.0002	Pr(T > t) = 0.0058

Table 36. **T-test results based on accumulated depreciation ratio**

	2012	**2013**	**2014**	**2015**	**2016**	**2012-2016**
High accumulated depreciation ratio	93.91	95.46	95.56	93.20	95.07	93.66
Low accumulated depreciation ratio	89.70	91.30	90.81	87.98	89.40	89.55
Difference	4.20	4.16	4.75	5.22	5.67	4.11
T-test (unequal)	Pr(T > t) = 0.0176	Pr(T > t) = 0.0185	Pr(T > t) = 0.0161	Pr(T > t) = 0.0082	Pr(T > t) = 0.0044	Pr(T > t) = 0.0025

[28] The data from 2010 to 2011 are missing.

6.4.3 Results from content analysis

We conducted the interviews at three national universities, which are summarized as follows:

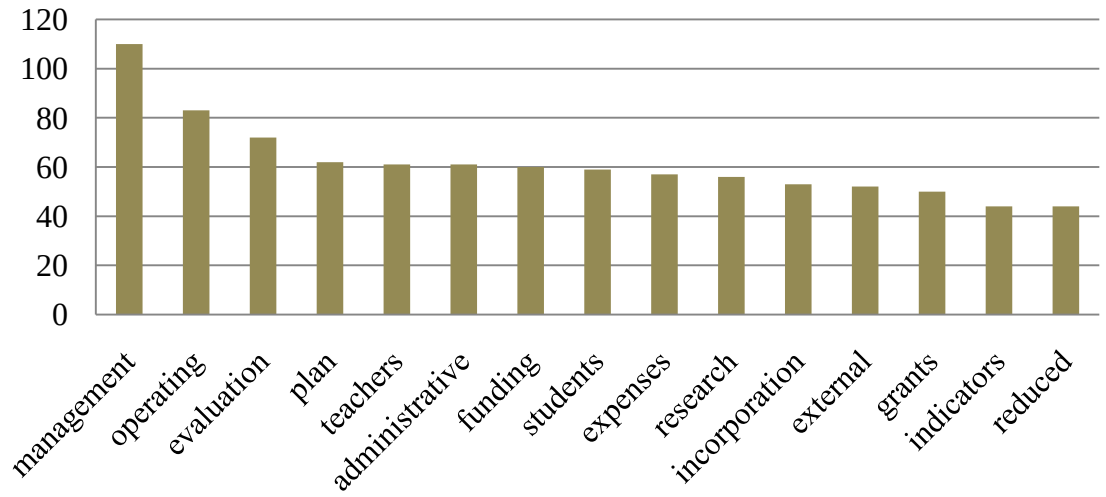
Table 37. Summary of the interviews

	University	Position	Participant	Place	Date	Duration
1	National university 1 (NU1)	Vice president	1	On-site	23rd December 2019	0h:50m:30s
2	National university 2 (NU2)	Vice president	1	On-site	25th December 2019	0h:42m:27s
3	National university 3 (NU3)	Vice president	1	On-site	04th March 2020	0h:43m:42s

By exploring the 100 words that frequently used in the interviews, it is suggested that the executives at national universities are concerned about the management, operating expenses, performance measurement and evaluation, research, and funding issues as illustrated in the following word frequency cloud.

Figure 22. **The one hundred most frequent words used in the interviews**

Figure 23. 15 most frequently used words



The themes arising from the interviews include: (i) The purposes of incorporation national universities seems to have not been significantly achieved despite the transformation have experienced almost 15 years; (ii) National universities are quite unsatisfied with policies applied for funding allocation and the response to financial restraint vary by university; and (iii) The ideas of accepting more overseas students and delivering courses in English are well advocated to cope with the shrinking student numbers and motivate the internationalization.

The first common view from the interviews is that the goals of corporatization have not been delivered as expected. According to Ministry of Education, Culture, Sports, Science, and Technology, the rationales for transforming the operational structure of Japanese national universities were to provide more autonomy in governance and promote the competitiveness to meet the social needs and renovate education and research (MEXT, 2010). These ideas were also confirmed by the interviewees when they suggested that the basic reasons were to increases the discretion of the universities (becoming more independent and autonomous) to stimulate education and research. Given that the purposes of reforms have been well grasped, the realization of these expectations might be still

limited as expressed in the following overall assessment on the achievements of corporatization:

"I think it's very difficult to answer. After all, the philosophy was good, and in any form, I think it was good to think that a traditional national university should seek some kind of change. Unfortunately, the high philosophy of incorporation, and perhaps some of the traditional people at national universities and the Ministry of Education have not been able to keep up with such people. If we couldn't keep up, we didn't have enough motivation or willingness or direction to achieve our ideals. I believe that the fact that they did not actively participate in the so-called "telling from above" story may be the reason that the current national university is changing halfway because it has various scopes and cannot be applied uniformly. There were a lot of university reforms and a lot of push reforms from MEXT" (NU 1)

Also, the interviewees suggested that probably the people at universities as well as at the Ministry of Education have not positively accepted the new concepts due to their old perceptions have entrenched in decades. Indeed, as indicated by Cummings (2010) that, as a nature of educational reform (even in a radical reform), the former ideology and practices will still exist and have an influence on the new and even probably take over the new at some time in the future. Moreover, there have been no incentives for them to achieve the goals of incorporation specified. Therefore, universities did not sufficiently prepare to stand on their own feet. These manifestations are also observed in a study by Fukudome (2019, pp.50-51) indicating that "…universities are basically conservative and often resist change. Changes are often externally enforced. Particularly in Japan, government policies have the greatest influence on institutions and compel such institutions to implement reforms. In more than two decades, autonomous governance and management by higher education institutions have been thought to be critical. However, it is rather difficult to change the mentality of government officials and academicians, both

of whom have formulated their traditions over a long period."

Regarding the management aspect, there were certain discernible improvements. For instance, the management has become more solid when compared to the prior model; the implementation of medium-plans allowed universities to better envisage the future. In addition, more open discussions and the use of consultancy have promoted the initiatives and suggestions (e.g. adopting good practices from private companies or other universities) for the daily management and reforms of the universities. However, adverse effects might be unavoidable since teachers are normally nominated as executives but they are not professional managers suggesting insufficient competence in management experiences and skills. On the other hand, it might not suitable to recruit managers from the private sector as the governance of a national university is dissimilar from those applied for-profit companies. In addition, it is suggested that given that national universities have had more discretion on finance and human resource management, the level of independence has not been achieved as expected. What we found is quite similar to the argument that efficient and effective management is one of the key challenges of Japanese universities since managers are academics with little management skills and lack the impetus to be professional managers; it has become worse in the context of national universities even after corporatization (Homma, 2012).

As a management tool, the six-year plan has been introduced following the incorporation is a good initiative to manage the university in the medium term. Accordingly, the Ministry of Education sets medium-term goals; the universities have the obligation to prepare a medium-term plan for six years, having the approval of the Minister of Education, and then prepare an annual plan for each year. However, it appears that the implementation was problematic since the university found it is difficult to specify the relationship between the evaluation of the medium-term plan and the evaluation of the single year, and what should be focused on daily management. While national universities

are diversified, the evaluation that using the same indicators or metrics might also cause the disparity. Moreover, it is stressed that not only in the evaluation of the medium-term plan but also in various evaluations which resulted in increasingly complicated paperwork. In addition, while the midterm evaluation was expected to strongly associate with financial reward or penalty, it turned out that no major financial consequences occurred (the reduction in subsidy or additional reward was less than 0,5% of the total subsidy) or in other words, the effectiveness of the evaluation exercise was not significant and the incentive was weak (Kaneko, 2012). The interviewees also expressed the difficulties and necessity in measuring the performance of national universities as follows:

"After all, education has been difficult to evaluate until now, and in the end, that area has come to this day without being clearly defined. However, it is not unclear that there are such aspects, but on the other hand, as a national university, there are responsibilities that respond to the demands of society, so what kind of form is evaluated and the results of the evaluation? It is important how to reflect this, so I don't think that such an approach is necessarily all useless." (NU1)

"There are some difficulties in university, so education is particularly difficult. However, when abstract language alone is not enough to convince, we sometimes try to set numerical targets. As you may be especially aware, I guess that it is the individual use of functional enhancement costs that have been around for the past three years or so, because it is useless if we do not set numerical targets, so each university has had a lot of trouble. In the medium-term plan, I was able to get what I could, but it was a case-by-case basis." (NU3)

"In short, the Ministry of Finance has been talking about individual evaluation indicators for each university so far, and that is, in short, not objective, because the universities are separate. The evaluation is the goal itself. As with [Key Performance Indicators] KPIs, the parts are quite loose and their evaluations are weak. The Ministry of

Finance says that, in short, it is incompatible with the careful allocation of universities that are doing their best, so a common indicator would be needed, so we did this last year. But, in the end, accounting and real management, apart from this, would not be the essence of university education and research as I mentioned earlier." (NU3)

Stimulating education and academic research was a key principle in the reform of national universities. A report on the situation of national universities after reform indicated that more than half of the respondents mentioned that the incorporation has a positive influence in revitalizing educational activities, research, and the services to support students were also improved (MEXT, 2010). Although the interviewees did not clearly indicate any specific assessments, they provided some examples that might partly support the positive changes. For instance, regarding education, the university collaborated with companies to develop a joint education program that the curriculum focuses on fieldwork. Also, it seems that students have taken classes more seriously recently since universities more frequently report on the attendance or the results of students. Therefore, the quality of teaching methods and the related standards have been gradually improved. However, it should be noted that these kinds of reform might not completely attribute to the incorporation but rather the influence of initiatives during the daily operation or discussions (e.g. those proposed by the university council). Similarly, in the academic research mission, joint research with external organizations is a typical example that not only expands the scope of research but also generates their own income.

The second concern arising from the interviews relates to financial management, especially the regulations and responses to the reduction of government grants. In the policy aspect, a common complaint was about the ambiguity in allocating funding. More particularly, they do not understand how the basis, norms, or criteria are applied in distributing funding and the current policies have not fully taken into account the

diversities in characteristics or peculiarities of national universities, which is expressed in the following quotation:

"It is not unclear how to make use of evaluations in allocating funds….In the future, how to organize the evaluation as a whole and how to allocate such resources based on the evaluation will be a national issue." (NU 1)

"It's a bit annoying, but for example, a college that only trains teachers, such as music, has a harder time in terms of research than science and engineering, so it is an evaluation that can deal with that. The president of my university has expressed my opinion at the National University Association. Regarding that point, we have not yet received a formal communication, but we will coordinate the next fiscal year's budget between the government and national universities in order to introduce an evaluation scale for each academic department in education and research" (NU 2)

An interesting point is the difference in responses to the reduction in government grants by either focusing on containing expenses or increasing external revenues but not increasing tuition fees. Also, the responses appeared to depend on the magnitude of reliance on government grants. While the subsidies reduce or remain unchanged, the labor cost increases and the other operating expenses will also be affected. Therefore, an approach was to generate income by collaborating with the private sector and receive the donation. In addition, the interviewee (at the university with lower proportion of government grants) also suggested that the university should not solely rely on government grants. In another university (that still heavily relies on funding from government), the main measure to cope with funding reduction was to cut the expenses where possible. For instance, payment for retired employees was frozen and the number of administrative staff was also be streamlined (e.g. a section manager can work in two departments), or salaries for specially appointed professors were set lower than those of full-time teacher. The key explanation

given for a stable level of tuition fees was that to support the students with low income have the opportunities to enroll higher education.

Finally, the third main thread arising from the interviews was the internationalization issue. Due to the low proportion of international staff and international students, the Japanese higher education system, as well as national universities, is expected to expand more widely and deeply international engagement (e.g. student exchange, transnational programs, joint research). The underlying motivation is to increase international competitiveness and secure the number of students (Goodman, 2016). This tendency was well reflected in the interviews with various initiatives. Indeed, an interviewee recognized that:

"… we must fight not only Japanese universities and nationals but also public and private universities around the world. As globalization progresses, perhaps a society where young people can play an active role around the world will probably come soon. For that reason, I think there is competition on the level of where to choose an educational institution to go to society and which university institution to choose" (NU1).

Therefore, universities collaborated with other international universities to provide doubled-degree programs and promote international joint research. In addition, more efforts have been made to recruit international students and this was viewed critically amidst the shrinking of the 18-year-old population. In parallel, interviewees indicated that training courses in English are necessary not only to attract overseas students but also to improve the competence of Japanese students after graduation to work in an international environment. However, it should be noted that the lack of incentives (e.g. payment is fixed and depending on the years of working experiences but not the value of the contribution to the university) hampers the recruitment of foreign teachers and restrains the delivery of English courses.

6.5 Public policy implications and conclusion

Japanese national universities experienced considerable structural reform following corporatization introduced in 2004, along with financial constraints and contraction in student numbers. Moreover, national universities have received increasing attention as they have an important role in social-economic development and consume a considerable amount of public funding. Therefore, there is a need to have an adequate understanding of whether national universities are efficient in using the allocated resources to execute their assigned missions.

The purpose of this study was to measure the technical efficiency of national universities over the period 2010 to 2016. In addition, we also sought to identify the determinants of efficiency, especially those amenable to public policy intervention (such as grant allocations and university size).

The results suggest that, over the examined period, the technical efficiency of national universities did not systematically decrease or increase despite corporatization and a reduction in grant allocations. This is somewhat contrary to what might be expected from a review of the extant literature. The most likely reasons for this absence of a trend were that while teacher cost changed in the opposite direction to efficiency trends, support staff cost and other operating expenses increased, at the same time that student enrolments declined. In addition, the content analysis using the interviews with top administrators implied that the structural reform of national universities has not yet achieved the goals initially designed.

Our analysis of the determinants of efficiency casts some light on likely efficacious public policy interventions. First, the negative impact of high reliance on grants could be interpreted to recommend a policy to reduce the amount of government funding allocated to national universities. However, an alternate public policy response might be to instead

provide greater incentives for universities to generated own-source revenue through commercializing their research, establishing greater links with the corporate sector or encouraging donations. In addition, overhauling procedures for allocating grant funds (e.g. target-setting, evaluation standards) might be expected to improve both transparency and efficiency.

The government might also motivate universities to use resources more efficiently (e.g. universities get rewards or penalties based on their performance against the targets) and universities should be more proactive in eliminating waste by improving the budget plan to better control the resources employed. In doing so, the actual autonomy in management and the management skills of executive managers should be enhanced. The universities should also put more emphasis on asset management (e.g. planning, acquiring, and maintenance) to utilize the fixed assets more efficiently, especially in the context of shrinking student numbers.

Demographic changes, especially dwindling student enrolments, might be expected to continue to act as a drag on efficiency for small-scale universities. One way to address this problem is to recruit students from abroad, in a similar manner to how British and Australian universities responded to hard budget constraints. To do so it will be important to enhance the competitiveness of Japanese universities by delivering more programs in English and further improving international rankings. If these measures are not possible then a potential solution is to merge existing universities to optimize scale. Indeed, a survey at the end of 2018 revealed that 16% of national universities were in the process of exploring mergers (Mainichi, 2019). Therefore, responses to the problem of diminishing enrolments might either be found through the recruitment of international students or by executing mergers of universities.

More emphasis might also be placed on the attached hospitals. Although an investigation in greater detail is beyond the scope of this study, renovating hospital governance (e.g.

implementing long-term financial plans), enhancing the collaboration to share best practices and consolidating the entire structure (including education, research, and medical services) might be expected to improve the efficiency (National University Hospital Council of Japan, 2014a). Similarly, streamlining science faculties might mitigate their negative impacts on efficiency. Accordingly, it might be helpful for government authorities to provide guidelines that direct and facilitate the acquisition of departments among national universities.

Future research might consider examining and comparing the other types of Japanese universities. In addition, if appropriate indicators for quality of teaching and research later become available this would also likely lead to more nuanced insights into the sector. The public policy recommendations that we have made in response to the empirical evidence presented in this paper will clearly require further development prior to implementation. Nevertheless, this study is important for both addressing a gap in the extant literature and identifying possible determinants to efficiency which are amenable to remedial public policy intervention.

Appendix 3: Semi-structured interview questions for Japanese national universities

1.1 Introductory questions

- The information collected will be used for doing research only. Respondent's information is anonymous and confidential. Please ask the interviewer to clarify if the respondent does not understand any questions.

- Could you please first state your position and responsibility?

- Could you please briefly summarize your experiences in the higher education sector?

1.2 Main questions

* *Management*

Q1. In your opinion, what are the main purposes of the incorporation of national universities?

Q2. Could you please describe the key changes in the governance at national universities following the corporatization? Do national universities have more discretionary power in management as expected?

Q3. What do you think about the effectiveness of the six-year plan?

Q4. Do you know any key indicators used to evaluate the performance of national universities? What do you think about the effectiveness of the implementation in reality?

* *Education and Research*

Q5. In what aspects has the incorporation affected the teaching and research activities?

Q6. Has the incorporation motivated the performance obtained by national universities?

* *Funding*

Q7. How has the reduction in government grants influenced the operation of universities?

Q8. Have your university increased tuition fees recently and what are the main reasons?

Q9. What do you think about the possibilities for national universities to generate other sources of income (except for tuition fees and grants)?

Q10. What is your opinion on the fairness and transparency in allocating government grants (including operating grants and research funding)?

* *Challenges and measures*

Q11. How is your assessment on the success of corporatization (e.g. management system, revitalizing education and research, activities)?

Q12. In your opinion, what are the main challenges for national universities (e.g. shrinking student population, reduction in government grants, internationalization …)?

Q13. Are there any defects or inconsistencies in government policies that have negative impacts on the performance of national universities? What should be changed to motivate the performance of national universities?

Q14. In comparison with other countries such as UK, Australia, the number of universities in Japan is very high and the number of students per university is quite small. How do you think about the consolidation of the faculties or merger of the universities?

Q15. According to statistic figures provided by MEXT, the number of students has decreased. Can you think of the solutions to cope with this issue?

1.3 Conclusion and thanks

CHAPTER 7: POLICY IMPLICATIONS AND CONCLUSION

The overarching theme of this thesis is efficiency measurement of public service organizations, specifically English acute foundation trusts, New Zealand DHBs, and Japanese national universities. Two-stage Data Envelopment Analysis - a well-attested technique - is the primary tool employed to estimate how efficiency changed and identify the determinants of efficiency. The two first sections of this chapter summarize the research findings and policy implications for the cases under examination while the limitations and recommendations for future work are indicated in the final section.

7.1 Summary of research findings

7.1.1 Efficiency trends under structural reform and austerity

One of the key purposes of this study is to measure efficiency of specific public service organizations in a period characterized by structural reforms, tightened budgets, and increased pressure stemming from other social-economic factors (e.g. aging population, rising cost, growing demand). Under these circumstances, we suppose that public service providers might be motivated to improve their efficiency as a result of reform and to cope with the reduction in public funding. However, despite the institutions under investigation having experienced similar conditions such as reforms and budget constraints, the results obtained from efficiency estimated seem to vary by case.

While both English acute foundation trusts and New Zealand DHBs have improved their efficiency, we could not observe a systematic increase or decrease in the efficiency of Japanese national universities. This might be used to indicate that reform and budget tightening may not always be effective, but rather might vary based on the particular context under examination. For instance, the healthcare sector in New Zealand and England faced growing demand with requirements to not compromise the service

standard, suggesting that they were required to “do more with less”. In contrast, Japanese national universities experienced intense competition and a consequent reduction in student numbers. Thus, whilst the service outputs of the healthcare sector in England and New Zealand increased, student numbers – an important output for universities - decreased.

The difference in the efficiency trend between the Japanese case and its counterparts can be explained by the responses to the reform and fiscal hardship faced. Upon further analysis of the utilization of resources, we found that English hospitals and New Zealand DHBs did make an effort to constrain their input usage. This is evident as the growth of inputs appears to be lower than those of outputs over the examined period; especially in relation to the reduction in the number of hospital beds and staff which may be seen as key factors contributing to efficiency improvement of English hospitals. Similarly, New Zealand DHBs seem to be successful in controlling the personnel cost and reducing non-clinical services expenses through the use of shared-service arrangements to save back-office costs. On the contrary, in the case of Japanese national universities, it seems that they were unable to contain resources usage in response to the contraction in student numbers given support staff cost and other operating expenses increased while student numbers constantly declined. In addition, it appears that the objectives targeted through the incorporation of Japanese national universities (e.g. providing more autonomy, renovating education, and research) have not been significantly achieved as there still exit resistance to change, lack of management skills, and ineffective management tools (e.g. six-year plan, funding allocation).

7.1.2 Determinants of efficiency

While measuring efficiency of these public service organizations can provide an overall picture of the performance of these chosen public organizations, understanding the factors that explain the variability in efficiency among the service providers might be the crucial

concern of managers, policy-makers, and other stakeholders alike. Therefore, our studies also put an emphasis on identifying the determinants of efficiency, especially the internal factors which are under the control of managers.

With respect to healthcare services, our findings obtained from the two-stage analysis are consistent with the current literature. In particular, while wider scope (various types of services provided) is negatively associated with efficiency, higher bed occupancy can improve hospital efficiency. Patient characteristics were also found to have an impact on efficiency. For example, since treating older patients is more costly compared to younger patients, a higher proportion of patients over 60 years was found to have a deleterious effect on the efficiency of English acute hospitals. Likewise, when examining the efficiency of New Zealand DHBs, we suspected that a higher proportion of Maori people might require greater resources due to a higher incidence of unfavorable health conditions, thus lower efficiency. However, contrary to our assumption, DHBs with a higher proportion of Maori patients tend to be more efficient as Maori patients were found to use relatively fewer resources and had the lowest proportion of patients aged over 65 when compared with other ethnic groups

Moreover, we also found that internal factors may provide useful avenues for managers to improve efficiency. Firstly, we focused on factors related to asset utilization. In the case of English acute foundation trusts, it is suggested that the value of buildings have a negative effect on efficiency, which may result from underutilization, use of buildings for unproductive purposes, or the buildings being outdated and inflexible. On the other hand, in the context of New Zealand DHBs, buildings were found to be positively associated with efficiency. We observed that DHBs did not sufficiently invest in buildings when 35% of the budget on capital expenditure was not executed. Therefore, DHBs with inadequate investment in buildings might not be able to maintain sustainable service levels and seem to be less efficient. Consequently, the impact of buildings in healthcare services for the

levels of technical efficiency obtained might vary, depending on the condition of assets, the utilization, and management of service providers. Secondly, information technology was found to have a positive impact on the efficiency of both English hospitals and New Zealand DHBs, suggesting the benefits of digitalization for the improvement of efficiency (better services and lower costs). Thirdly, the impact of performance against plans or targets was specifically investigated in the case of New Zealand DHBs. The result suggests that DHBs that outperformed emergency targets and outpatient plans were found to be more efficient while technical efficiency might be reduced if overspending occurred relative to the budget. Thus, there is evidence to suggest that efficiency may also be influenced by the ability of organizations to controlling spending levels in accordance with the budget and delivering services in line with the targets set.

Utilizing a different approach, through an investigation of Japanese national universities, we found evidence that variables related to public funding or characteristics of universities were associated with efficiency. The level of autonomy was assumed to be inversely linked with the dependence on public funding and thus a higher proportion of funding from the government might reduce efficiency. We, therefore, examined the impact of the reduction in government grants for Japanese national universities utilizing the proportion of grants over total income. Also, we hypothesized that other factors such as the decline in the number of student enrolments, affiliated hospitals, and faculty structures might be associated with efficiency. The result suggests that a higher reliance on government grants can reduce efficiency, which might result from the relative lack of transparency in funding allocation or "flypaper effect". Conversely, universities with enrolments greater than 10,000 students are likely to be more efficient suggesting the presence of economies of scale. Since running attached hospitals is often more costly, and science faculties typically consume higher resources compared with other faculties (such as economics or humanities), these two factors were found to have deleterious effects on the efficiency of

national universities. Similar to the cases of English hospitals and New Zealand DHBs, we also attempted to identify the impact of tangible fixed assets and found that universities that have a higher value of tangible fixed assets per student tend to be less efficient on average. In this research, along with factors that well established to be associated with efficiency in the literature, we also identified the impact of other factors specific to Japanese national universities such as the attached hospital, student numbers, and the value of assets, thus making a vital contribution to the literature.

7.2 Policy implications

Although it appears that reform and a tightening of public budgets can provide a motivation to improve efficiency, public services organizations can also play a vital role in realizing the motivation and thus increasing efficiency, subject to their responses and their ability to eliminate waste in the use of the available resources (e.g. asset utilization). Therefore, based on these results, it may be recommended that a greater emphasis is placed on internal factors that can be controlled by the organizations (in this instance hospitals and universities) when any structural reform plan is proposed. Based on these findings, detailed policy implications derived might draw the attention of both policymakers and the managers to avenues which might provide fruitful for the further improvement of efficiency in the public service organizations under investigation

From the policymakers' perspective, in the case of English healthcare, NHS Improvement should widely propagate good practices and initiatives in treating and optimizing patient flows (e.g. provision of recovery care at home) so hospitals can learn from their peers. Similarly, with respect to Japanese authorities, it is recommended that MEXT needs to overhaul and revise the policies in funding allocation and the implementation of six-year plans to improve transparency and efficiency. Also, more directions and instructions may be necessary for Japanese national universities to facilitate mergers between universities or the acquisitions of faculties to address the shrinking number of enrollments. In addition,

greater incentives (e.g. tax exemption or tax deduction) for universities to generate own-source revenue through the commercializing of research output, establishing greater links with the corporate sector, or encouraging donations, which indirectly reduce the reliance on government grants and increase autonomy.

Given we stress the impact of internal factors within managers' control, most of the proposed solutions relate to the responsibility of managers. For example, as a decline in the length of stay (LOS) was associated with efficiency improvement of English hospitals, managers should motivate the engagement of front-line staff in finding initiatives to optimize the treatment processes (e.g. avoiding over-diagnosis or over-prescription) and shorten LOS (e.g. reviewing LOS of patients who have been in hospitals for, say, one week) to free up capacity and reduce costs.

With regard to asset management, the association identified in our studies between asset utilization and efficiency suggests that fixed assets such as buildings and information technology are critical in improving efficiency since they constitute a significant share of total assets and largely influenced available capacity. Therefore, it is advised for healthcare providers to maintain an optimal level of assets to provide services with an acceptable level of quality, and this might be achieved by exploiting resources such as buildings and applying information technology more effectively. Specifically, in order to minimize unutilized or unproductive spaces, measures such as reconfiguration of buildings and services can potentially save on operating costs. Along with enhancing the management of assets currently used, attention should be paid to sufficient investment in assets in the long-term (to avoid infrastructure backlogs and transferring costs to the future) through a comprehensive assessment of the assets' conditions and careful forecast the demands for services. Similarly, Japanese national universities should also put more emphasis on asset management (e.g. planning, acquiring, and maintenance) to utilize the fixed assets more efficiently, especially in the context of shrinking student numbers

Under the pressure of squeezed public funding, public service organizations were required to achieve savings targets, and the managers were confronted with the challenge to maintain financial sustainability and provide services efficiently. Therefore, our findings also provide recommendations to improve financial management. The implementation of tools such as Service Line Report (SLR) and Patient Level Costing (PLICS) is recommended since they can monitor and identify the least efficient areas in an organization and thus might potentially mitigate negative outcomes when English acute foundation trusts expand the scope of services. Likewise, as preparing and controlling budgets are the two key elements of financial management, it is suggested that New Zealand DHBs should prepare the budgets more deliberately and make greater efforts to ensure that expenditure is more closely aligned to the budget. In particular, possible measures might include enhancing the participation of financial experts at executive boards, greater use of digital technologies, and collaborating with other DHBs to develop both short-term and long-term plans. In this regard, Japanese national universities should pay more attention to the enhancement of management skills for executive managers in order to improve the budget plan to better control resources employed and thus eliminate waste. In addition, as efficiency relates to the ability to utilize the available resources more efficiently, only in accrual basis are the cost fully recognized; accrual accounting and accrual budgeting are, therefore, expected to provide more adequate information for better decision making and thus improve efficiency. In this respect, our in-depth analysis of cost structure and the incorporation of internal factors (e.g. assets utilization, performance against the budget) in the two-stage analysis also indirectly underlines and advocates the implementation of accrual accounting and accrual budgeting, which could be a suggestion for Japanese public sector to further accelerate the adoption of accrual basis in both accounting and budgeting.

Reorganization or consolidation of the organization is specifically proposed in the case of Japanese national universities, in which mergers of universities or streamlining of inefficient faculties might be possible to exploit economies of scale and economies of scope. Although there is a need to have greater involvement in the form of directions or instructions from governments as mentioned previously, universities should also be more proactive to promote this process in a timely manner. In addition, an alternative measure to cope with shrinking student numbers is to increase the enrolments from overseas students by delivering more programs in English and further improving international rankings. Attention should also be paid for redesigning governance of affiliated hospitals (e.g. implementing long-term financial plans), enhancing collaboration to share best practices which might be expected to improve efficiency.

7.3 Limitations and further research

As a deterministic approach, Data Envelopment Analysis is rather sensitive to the data and sample selected to estimate the efficiency. For example, efficiency scores might be influenced by the accuracy of data, model specification (e.g. number of inputs and outputs), the number of units under evaluation, and the period of time examined. Moreover, while measuring the efficiency of public service organizations is a complex task, a study might only look at specific areas of efficiency or performance due to time constraints or other limitations. Indeed, previous efforts involving efficiency measurement of the public sector have encountered difficulties in capturing aspects of service quality. Therefore, we clarify a few limitations and possible considerations that can be improved in future research.

Regarding the study on English hospitals, we have carefully adjusted the outputs for the patient complexities. However, these models still lack controls for the factors associated with hospital quality such as waiting time, the four-hour targets, patient experiences, delayed transfer, mortality, and readmission, which may also be worth examining either as

outputs or explanatory variables in future studies. Moreover, while we found a certain improvement in efficiency over the examined period, an extension of time period might provide additional evidence as to whether the acute foundation trusts can still deliver efficiently or they have been stifled under prolonged austerity.

Likewise, research on the efficiency of New Zealand DHBs can be expanded by evaluating the efficiency of DHBs in different service areas such as prevention services, early detection and management, intensive assessment and treatment, and rehabilitation and support. Similarly, the efficiency of a DHB can be examined based on the composition of two tiers – primary care and secondary care. In this regard, different DEA methods such as network DEA can be applied to get more insights into efficiency of component divisions. In addition, the amenable mortality rate and factors related to lifestyle or prevalence of morbidity are additional factors worthy of further examination.

In the case of Japanese public higher education, apart from national universities, public universities are also publicly-owned entities. Therefore, it would be interesting to compare the efficiency between these two types of Japanese public universities. Moreover, if data proxies for quality of teaching and research are available, an analysis with these factors will provide valuable insights into the sector.

While the debate on the selection of the regression model applied in the second-stage analysis is divisive and ongoing, future research might extend this research by examining the validity and robustness of the results obtained under different model specifications. We have identified possible determinants of efficiency which may be used as a basis for public policy interventions. However, caution should be taken that the policy recommendations or the managerial solutions that we have made definitely require elaborate development and impact assessment prior to implementation.

All in all, efficiency measurement of the public service organizations is part of the performance measurement and performance management which is a key component in public administration and public management. Our studies of efficiency measurement complement and contribute to the current literature on measuring the efficiency of the public sector in different ways. While there is a scarcity of recent research conducted in the contexts that we examined, our empirical research complements the existing literature regarding the application of DEA methodologies in the education and healthcare sectors. In this way, we have made a pioneering approach in applying two-stage DEA to providing an updated picture of efficiency and identify possible determinants of efficiency in the cases of English acute foundation trusts, New Zealand DHBs, and Japanese national universities. In addition, unlike the extant DEA studies that measured efficiency in isolation, we also attempted to provide more detailed insights by explaining the changes in efficiency over time through a deeper investigation of government policies and analysis of the structure and growth of resources used as well as services provided. Moreover, the adjustment for complexity in calculating the hospital outputs made in our analyses might better secure the accuracy when estimating efficiency for hospitals. Likewise, the impacts of internal factors that are under the control of managers can provide additional insights when conducting the two-stage analysis. Nevertheless, measuring the efficiency of public service organizations is a challenging task; what we have found might reflect only a small part of a complex and dynamic picture, and thus there still exist fruitful avenues for the improvement and further development of efficiency measurement in the public sector.

REFERENCES

Agasisti, T., & Pohl, C. (2012). Comparing German and Italian Public Universities: Convergence or Divergence in the Higher Education Landscape? *Managerial and Decision Economics*, *33*(2), 71–85.

Agasisti, T., & Wolszczak-Derlacz, J. (2016a). Exploring efficiency differentials between Italian and Polish universities, 2001-11. *Science and Public Policy*, *43*(1), 128–142.

Agasisti, T., & Wolszczak-Derlacz, J. (2016b). Exploring efficiency differentials between Italian and Polish universities, 2001-11. *Science and Public Policy*, *43*(1), 128–142.

Aghion, P., Dewatripont, M., Hoxby, C., Mas-Colell, A., & Sapir, A. (2010). The governance and performance of universities: evidence from Europe and the US. *Economic Policy*, *25*(61), 7–59.

Allen, B., Berman, E., Cantal, C., Eppel, E., Jackson, B., Löfgren, K., … Plimmer, G. (2017). *Independent Review of the Performance Improvement Framework*. Wellington, New Zealand: Victoria School of Government. Retrieved from http://www.ssc.govt.nz/performance-improvement-framework

Allen, B., & Eppel, E. (2019). Holding on tight – NPM and the New Zealand performance improvement framework. *Australian Journal of Public Administration*, (October), 1–16. https://doi.org/10.1111/1467-8500.12405

Allen, R., Hemming, R., & Potter, B. H. (2013). Introduction: The Meaning, Content and Objectives of Public Financial Management. In R. Allen, R. Hemming, & B. Potter (Eds.), *The International Handbook of Public Financial Management*. Basingstoke: Palgrave Macmillan.

Allin, S., Grignon, M., & Wang, L. (2014). The determinants of efficiency in the Canadian health care system. *Health Economics, Policy and Law*, *11*(1), 39–65. https://doi.org/10.1017/S1744133115000274

Andersson, C., Antelius, J., Månsson, J., & Sund, K. (2016). Technical efficiency and productivity for higher education institutions in Sweden. *Scandinavian Journal of Educational Research*, *61*(2), 205–223. https://doi.org/10.1080/00313831.2015.1120230

Andrews, R., & Entwistle, T. (2014). *Public Service Efficiency: Reframing the Debate*. New York: Routledge.

Applyby, J., Baird, B., Thompson, J., & Jabbal, J. (2015). *The NHS under the coalition government. Part two: NHS performance*. Retrieved from https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/the-nhs-under-the-coalition-government-nhs-performance-kings-fund-mar15.pdf

Aragon Aragon, M. J., Castelli, A., & Gaughan, J. (2017). Hospital Trusts productivity in the English NHS: Uncovering possible drivers of productivity variations. *PLoS ONE*, *12*(8), 1–14. https://doi.org/10.1371/journal.pone.0182253

Ardagh, M. (2015). What have five years of the shorter stays in the emergency department health target done to us? *New Zealand Medical Journal*, *128*(1421), 47–54.

Ashton, T., & Tenbensel, T. (2012). Health reform in New Zealand: short-term gain but long-term pain? *Expert Review of Pharmacoeconomics & Outcomes Research*, *12*(5),

579–588.

Association of Salaried Medical Specialists. (2018). *Should district health boards pay a capital charge ?* Retrieved from https://www.asms.org.nz/wp-content/uploads/2018/05/Research-Brief-Capital-Charge_169877.2.pdf

Auckland District Health Board. (2017). *Annual Report 2016/2017*. Retrieved from https://www.adhb.health.nz/assets/Documents/About-Us/Planning-documents/Auckland-DHB-2016-17-Annual-Report-online-version.pdf

Australian Institute of Health and Welfare. (2018a). *Australia's hospitals 2016–17 at a glance*.

Australian Institute of Health and Welfare. (2018b). *Hospital resources 2016-17 : Australian hospital statistics*. Retrieved from https://www.safetyandquality.gov.au/sites/default/files/migrated/ACSQHC_Annual_Report_2017-2018.pdf

Badunenko, O., & Tauchmann, H. (2018). *Simar and Wilson two-stage efficiency analysis for Stata* (FAU Discussion Papers in Economics No. 08/2018). Erlangen. Retrieved from https://www.econstor.eu/handle/10419/179503

Bailey, S. J., & Connolly, S. (1998). The flypaper effect: Identifying areas for further research. *Public Choice*, *95*(3–4), 335–361.

Bain, C. A. (2010). Myths of ideal hospitals occupancy. *Medical Journal of Australia*, *192*(1), 42–43.

Baker, B. C. (2017). *NHS Indicators: England, October 2017*. Retrieved from http://researchbriefings.files.parliament.uk/documents/CBP-7281/cbp07281-oct17.pdf

Banker, R. D., & Natarajan, R. (2008). Evaluating Contextual Variables Affecting Productivity Using Data Envelopment Analysis. *Operations Research*, *56*(1), 48–58.

Banker, R., Natarajan, R., & Zhang, D. (2019). Two-Stage Estimation of the Impact of Contextual Variables in Stochastic Frontier Production Function Models Using Data Envelopment Analysis: Second Stage OLS versus Bootstrap Approaches. *European Journal of Operational Research*, *278*(2), 368–384

Barnum, D. T., Walton, S. M., Shields, K. L., & Schumock, G. T. (2011). Measuring hospital efficiency with Data Envelopment analysis: Nonsubstitutable vs. substitutable inputs and outputs. *Journal of Medical Systems*, *35*(6), 1393–1401.

Barra, C., Lagravinese, R., & Zotti, R. (2018). Does econometric methodology matter to rank universities? An analysis of Italian higher education system. *Socio-Economic Planning Sciences*, *62*, 104–120.

Behn, R. (2003). Why Measure Performance? Different Purposes Require Different Measures. *Public Administration Review*, *63*(5).

Bergmann, A. (2009). *Public Sector Financial Management*. Essex: FT Prentice Hall.

Berta, P., Callea, G., Martini, G., & Vittadini, G. (2010). The effects of upcoding, cream skimming and readmissions on the Italian hospitals efficiency: A population-based investigation. *Economic Modelling*, *27*(4), 812–821.

Bevan, G., & Hood, C. (2006). What's measured is what matters: Targets and gaming in the English public health care system. *Public Administration*, *84*(3), 517–538.

Bird, S. M., Cox, D., Farewell, V. T., Goldstein, H., Holt, T., & Smith, P. C. (2005). Performance indicators: Good, bad, and ugly. *Journal of the Royal Statistical Society*, *168*(1), 1–27.

Blackett, D. J., Carslaw, A., Lees, D., Fargher, S., Rao, S., & Pohl, M. (2014). The impact of the 6-month waiting target for elective surgery: A patient record study. *New Zealand Medical Journal*, *127*(1405), 45–53.

Blondy, G., Cooper, J., Irwin, T., Kauffmann, K., & Khan, A. (2013). The Role of Fiscal Reorting in Public Fiancial Management. In Cangia (Ed.), *Public Financial Management and Its Emerging Architecture*. Washington, D.C: International Monetary Fund.

Boyne, G. (2010). Performance management: does it work? In R. M. Walker, G. A. Boyne, & G. A. Brewer (Eds.), *Public Management and Performance: Research Directions*. Cambridge: Cambridge University Press.

Braithwaite, J., Hibbert, P., Blakely, B., Plumb, J., Hannaford, N., Long, J. C., & Marks, D. (2017). Health system frameworks and performance indicators in eight countries: A comparative international analysis. *SAGE Open Medicine*, *5*, 1-10.

Bruijn, H. De. (2007). *Managing Performance in the Public Sector*. New York: Routledge.

Byrkjeflot, H., du Gay, P., & Greve, C. (2018). What is the 'Neo-Weberian State' as a Regime of Public Administration? In E. Onrago & S. van Thiel (Eds.), *The Palgrave Handbook of Public Administration and Management in Europe* (pp. 991–1009). London: Palgrave Macmillan.

Cantor, V. J. M., & Poh, K. L. (2018). Integrated Analysis of Healthcare Efficiency: A Systematic Review. *Journal of Medical Systems*, *42*(8), 1–23.

Carrington, R., Coelli, T. I. M., & Rao, D. S. P. (2005). The Performance of Australian Universities: Conceptual Issues and Preliminary Results. *Economic Papers*, *24*(2), 145–163.

Castelli, A., Street, A., Verzulli, R., & Ward, P. (2015). Examining variations in hospital productivity in the English NHS. *European Journal of Health Economics*, *16*(3), 243–254.

Chalmers, L. M., Ashton, T., & Tenbensel, T. (2017). Measuring and managing health system performance: An update from New Zealand. *Health Policy*, *121*(8), 831–835.

Chan-Tiberghien, J. (2010). Academic Capitalism in Japan: National University Incorporation and Special Zones for Structural Reform. In J. Zajda & M. A. Geo-JaJa (Eds.), *The Politics of Education Reforms* (pp. 41–53). New York: Springer Science + Business Media.

Chang, H., Chang, W. J., Das, S., & Li, S. H. (2004). Health care regulation and the operating efficiency of hospitals: Evidence from Taiwan. *Journal of Accounting and Public Policy*, *23*(6), 483–510.

Charnes, A., Cooper, W., Lewin, A., & Seiford, L. (1994). *Data Envelopment Analysis:*

Theory, Methodology, and Application. New York: Springer Science+Business Media.

Chilingerian, J., & Sherman, H. D. (2011). Health-Care Applications: From Hospitals to Physicians, from Productive Efficiency to Quality Frontiers. In W. W. Cooper, L. M. Seiford, & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (Second Edi, Vol. 44, pp. 445–493). Boston, MA: Springer US.

Chou, T.-H., Ozcan, Y. A., & White, K. R. (2012). Technical and scale efficiencies of Catholic hospitals: Does a system value of stewardship matter? In E. T`anfani & A. Testi (Eds.), *Advanced DecisionMaking Methods Applied to Health Care* (pp. 83–102). Milan: Springer-Verlag Italia.

Christensen, T. (2011a). Japanese university reform-hybridity in governance and management. *Higher Education Policy*, *24*(1), 127–142.

Christensen, T. (2011b). University governance reforms : potential problems of more autonomy ? *Higher Education*, *62*, 503–517.

Christensen, T., & Laegreid, P. (2011). Post-Npm reforms: Whole of government approaches as a new trend. In S. Groeneveld & S. Van de Walle (Eds.), *New Steering Concepts in Public Management* (pp. 11–24). Bingley: Emerald Group Publishing Ltd.

Clover, B. (2011). Monitor directors say FTs “opportunistic” in acquiring community services. Retrieved from https://search.proquest.com/central/docview/875630680/fulltext/2928B2FBB254B77PQ/1?accountid=39809

Coelli, T. J., Rao, D. S. P., O’Donnell, C. J., & Battese, G. E. (2005). *An introduction to efficiency and productivity analysis*. Springer Science & Business Media.

Controller and Auditor-General. (2015). *Inquiry into Health Benefits Limited*. Retrieved from https://www.oag.govt.nz/2015/inquiry-hbl/docs/health-benefits-ltd.pdf

Controller and Auditor-General. (2016a). *District health boards’ response to asset management requirements since 2009*. Retrieved from https://www.oag.govt.nz/2016/dhbs-assets/docs/dhbs-assets.pdf

Controller and Auditor-General. (2016b). *Health sector: Results of the 2014/15 audits*. Retrieved from https://www.oag.govt.nz/2016/health-audits/docs/health-audits.pdf

Controller and Auditor-General. (2018). *Health sector: Results of the 2016/17 audits*. Retrieved from https://www.oag.govt.nz/2018/health-sector-audits/docs/health-sector-audits.pdf

Cook, W. D., Tone, K., & Zhu, J. (2014). Data envelopment analysis: Prior to choosing a model. *Omega (United Kingdom)*, *44*, 1–4.

Cooper, W. W., Seiford, L. M., & Joe, Z. (2011). Data Envelopment Analysis: History, Models, and Interpretations. In W. W. Cooper, L. M. Seiford, & J. Zhu (Eds.), *Handbook on Data Envelopment Analysis* (2nd ed., pp. 1–40). New York: Springer Science+Business Media.

Crawford, R., & Stoye, G. (2015). *Challenges for health spending*. Retrieved from https://www.ifs.org.uk/uploads/gb/gb2015/ch8_gb2015.pdf

Cummings, W. K. (2010). How Educational Systems Form and Reform. In J. Zajda & M. A. Geo-JaJa (Eds.), *The Politics of Education Reforms* (pp. 19–40). New York: Springer Science + Business Media.

Curristine, T., & Flynn, S. (2013). In Search of Results: Strengthening Public Sector Performance. In M. Cangiano, T. Curristine, & M. Lazare (Eds.), *Public Financial Management and Its Emerging Architecture* (pp. 225–258). Washington, D.C: International Monetary Fund.

Cylus, J., Richardson, E., Findley, L., Longley, M., O'Neill, C., & Steel, D. (2015). United Kingdom: Health system review. *Health Systems in Transition*, *17*(5), 1–125. https://doi.org/http://www.euro.who.int/__data/assets/pdf_file/0008/85391/E93667.pdf

Czypionka, T., Kraus, M., Mayer, S., & Röhrling, G. (2013). Efficiency, ownership, and financing of hospitals: The case of Austria. *Health Care Management Science*, *17*(4), 331–347.

Daraio, C., Bonaccorsi, A., & Simar, L. (2015). Efficiency and economies of scale and specialization in European universities : A directional distance approach. *Journal of Informetrics*, *9*(3), 430–448.

De Bruijn, H., & Van Helden, G. J. (2006). A Plea for Dialogue Driven Performance-Based Management Systems: Evidence From the Dutch Public Sector. *Financial Accountability and Management*, *22*(4), 405–423.

De Vries, M., & Nemec, J. (2013). Public sector reform: an overview of recent literature and research on NPM and alternative paths. *International Journal of Public Sector Management*, *26*(1), 4–16.

De Witte, K., & López-Torres, L. (2017). Efficiency in education: a review of literature and a way forward. *Journal of the Operational Research Society*, *68*(4), 416–430.

Department of Health. (2010a). *Equity and Excellence: Liberating the NHS*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/213823/dh_117794.pdf

Department of Health. (2010b). *The Operating Framework 2011/12*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/216187/dh_122736.pdf

Department of Health. (2011). *Operational guidance to the NHS: Extending patient choice of provider*. Retrieved from http://www.dh.gov.uk/prod_consum_dh/groups/dh_digitalassets/documents/digitalasset/dh_128462.pdf

Department of Health. (2012a). *A simple guide to Payment by Results*. Retrieved from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/213150/PbR-Simple-Guide-FINAL.pdf

Department of Health. (2012b). *Overview of the Health and Social Care Act fact sheet*. Retrieved from https://www.gov.uk/government/publications/health-and-social-care-act-2012-fact-sheets

Dickinson, H. (2016). From New Public Management to New Public Governance: The implications for a 'new public service.' In J. Butcher & D. Gilchrist (Eds.), *The Three Sector Solution* (pp. 41–60). Australian National University: ANU Press.

Ding, D. X. (2014). The effect of experience, ownership and focus on productive efficiency: A longitudinal study of U.S. hospitals. *Journal of Operations Management*, *32*(1–2), 1–14.

Dong, G. N. (2016). Earning Management in US Hospitals. *Journal of Health and Human Services Administration*, *39*(1), 41–71.

Dooren, W. Van, Bouckaert, G., & Halligan, J. (2015). *Performance Management in the public sector* (2nd ed.). New York: Routledge.

Drechsler, W. (2014). The rise and demise of the New Public Management: Lessons and opportunities for South East Europe. *Central European Public Administration Review*, *7*(3), 7–27.

Drechsler, W., & Randma-liiv, T. (2014). *The New Public Management Then and Now : Lessons from the Transition in Central and Eastern Europe*. *Working Papers in Technology Governance and Economic Dynamics no. 57*.

Drew, J., & Gamage, S. (2018). Just Do It? A Cautionary Tale on Implementing Performance Management Regimes. In R. Pilcher & D. Gilchrist (Eds.), *Public sector accounting, accountability and governance—Globalising the experiences ofAustralia and New Zealand* (pp. 105–118). UK: London: Routledge.

Drew, J., & Grant, B. (2017). Means, Motive, and Opportunity – Local Government Data Distortion in a High-Stakes Environment. *Australian Journal of Public Administration*, *76*(2), 237–250.

Drew, J., O'Flynn, J., & Grant, B. (2018). Performing what? Exploring and expanding the notion of synecdoche in performance management practice. *Public Administration Quartely*, *42*(3), 395–424.

Drew, J. and O'Flynn, J. (2019). A Test of Wills? Exploring Synecdoche and Gaming in the National Literacy and Numeracy Performance Monitoring Regime. In Blackman, D., Buick, F., Gardner, K., Johnson, S., O'Donnell, M., and Olney, S. (Eds), *Handbook of Performance Monitoring in the Public Sector*. Cheltenham: Edgar Elgar.

Drisko, W. J., & Maschi, T. (2016). *Content Analysis*. New York: Oxford University Press Inc.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001a). Pitfalls and protocols in DEA. *European Journal of Operational Research*, *132*(2), 245–259.

Dyson, R. G., Allen, R., Camanho, A. S., Podinovski, V. V., Sarrico, C. S., & Shale, E. A. (2001b). Pitfalls and protocols in DEA. *European Journal of Operational Research*, *132*(2), 245–259.

Edwards, A. N. (2011). *Perspectives NHS buildings : obstacle or opportunity ?* Retrieved from https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/perspecti

ves-estates-nhs-property-nigel-edwards-jul13.pdf

Elo, S., Kääriäinen, M., Kanste, O., Pölkki, T., Utriainen, K., & Kyngäs, H. (2014). Qualitative Content Analysis: A Focus on Trustworthiness. *SAGE Open*, *4*(1), 1-10.

Emrouznejad, A., & Yang, G. liang. (2018). A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016. *Socio-Economic Planning Sciences*, *61*, 4–8.

Erlingsson, C., & Brysiewicz, P. (2017). A hands-on guide to doing content analysis. *African Journal of Emergency Medicine*, *7*(3), 93–99.

Ferrari, A. (2006). Market oriented reforms of health services: A non-parametric analysis. *Service Industries Journal*, *26*(1), 1–13.

Ferreira, D. C., Marques, R. C., & Nunes, A. M. (2018). Economies of scope in the health sector: The case of Portuguese hospitals. *European Journal of Operational Research*, *266*(2), 716–735.

Fixler, T., Paradi, J. C., & Yang, X. (2014). A data envelopment analysis approach for measuring the efficiency of Canadian acute care hospitals. *Health Services Management Research*, *27*(3–4), 57–69.

Flegg, A. T., Allen, D. O., Field, K., & Thurlow, T. W. (2004). Measuring the efficiency of British universities: a multi-period data envelopment analysis. *Education Economics*, *12*(3), 231–249.

Flynn, S., Moretti, D., & Cavanagh, J. (2016). *Implementing Accrual Accounting in the Public Sector*. (M. Cangiano, T. Curristine, & M. Lazare, Eds.), *Technical Notes and Manuals*. International Monetary Fund.

Førsund, F. R. (2018). Economic interpretations of DEA. *Socio-Economic Planning Sciences*, *61*(1), 9–15.

Frederickson, H. G., Smith, B. K., Larimer, C., & Licari, M. (2012). *The public administration theory primer*. (2nd). Boulder: Westview Press.

Fukudome, H. (2019). Higher Education in Japan: Its Uniqueness and Historical Development. In Y. Kitamura, T. Omomo, & M. Katsuno (Eds.), *Education in Japan: A Comprehensive Analysis of Education Reforms and Practices* (pp. 41–52). Springer Nature Singapore.

Gauld, R. (2016). A view from abroad: a New Zealand perspective on the English NHS health reforms. In M. Exworthy, R. Manion, & M. Powell (Eds.), *Dismantling the NHS ? Evaluating the impact of health reforms* (pp. 343–364). Bristol: Policy Press.

Gebicki, M., Mooney, E., Chen, S. J. (Gary), & Mazur, L. M. (2014). Evaluation of hospital medication inventory policies. *Health Care Management Science*, *17*(3), 215–229.

Giancotti, M., Guglielmo, A., & Mauro, M. (2017). Efficiency and optimal size of hospitals: Results of a systematic search. *PLoS ONE*, *12*(3).

Gianino, M. M., Siliquini, R., Damiani, G., Lenzi, J., Ricciardi, W., Bonaudo, M., & Fantini, M. P. (2018). The switch between cataract surgical settings: Evidence from a time series analysis across 20 EU countries. *Plos ONE*, *13*(2), e0192620.

Gill, P., Stewart, K., Treasure, E., & Chadwick, B. (2008). Methods of data collection in

qualitative research: Interviews and focus groups. *British Dental Journal*, *204*(6), 291–295.

Glass, J. C., McCallion, G., McKillop, D. G., Rasaratnam, S., & Stringer, K. S. (2005). Implications of variant efficiency measures for policy evaluations in UK higher education. *Socio-Economic Planning Sciences*, *40*(2), 119–142.

Goodman, R. (2016). Forword. In J. Mock, H. Kawamura, & N. Naganuma (Eds.), *The Impact of Internationalization on Japanese Higher Education* (pp. vii–ix). Rotterdam: Sense Publishers.

Grant, D., & Chapman, J. (2010). New millennium, new public management and the New Zealand model. *Australian Journal of Public Administration*, *69*(3), 301–313.

Griffiths, S., & Kippin, H. (2017). Public Services after Austerity : Zombies , Suez or Collaboration ?, *88*(3), 417–424.

Hadad, S., Hadad, Y., & Simon-Tuval, T. (2013). Determinants of healthcare system ' s efficiency in OECD countries. *European Journal of Health Economics*, *14*(2), 253–265.

Ham, C. (2014). *Reforming the NHS from within*. Retrieved from http://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/reforming-the-nhs-from-within-kingsfund-jun14.pdf

Ham, C., Baird, B., Gregory, S., Jabbal, J., & Alderwick, H. (2015). *The NHS under the coalition government. Part one: NHS reform*. *The King's Fund*. Retrieved from http://www.kingsfund.org.uk/sites/files/kf/field/field_publication_file/the-nhs-under-the-coalition-government-part-one-nhs-reform.pdf

Hashimoto, K., & Cohn, E. (1997). Economies of Scale and Scope in Japanese Private Universities. *Education Economics*, *5*(2), 37–41.

Health Department of Da Nang. (2020). Report on the quality assessment and examination of hospitals in Da Nang city in 2019 [Kết quả kiểm tra, đánh giá chất lượng Bệnh viện năm 2019]. Retrieved from https://soyte.danang.gov.vn/chi-tiet-tin-tuc?dinhdanh=31102&cat=32518

Hines, J. R., & Thaler, R. H. (1995). Anomalies: The Flypaper Effect. *Journal of Economic Perspectives*, *9*(4), 217–226. https://doi.org/10.1257/jep.9.4.217

HM Treasury. (2017). Public Expenditure Statistical Analyses (PESA). Retrieved from https://www.gov.uk/government/collections/public-expenditure-statistical-analyses-pesa

Hoff, A. (2007). Second stage DEA: Comparison of approaches for modelling the DEA score. *European Journal of Operational Research*, *181*(1), 425–435.

Hollingsworth, B. (2008). The measurement of efficiency and productivity of health care delivery. *Health Economics*, *17*(10), 1107–1128.

Hollingsworth, B., & Parkin, D. (2003). Efficiency and productivity change in the English National Health Service: Can data envelopment analysis provide a robust and useful measure? *Journal of Health Services Research and Policy*, *8*(4), 230–236.

Homma, M. (2012). Current Challenges Facing Japanese Universities and Future

Perspectives. In H. G. Schuetze, W. Bruneau, & G. Grosjean (Eds.), *University Governance and Reform: Policy, Fads, and Experience in International Perspective* (pp. 197–208). New York: Palgrave Macmillan. https://doi.org/10.1017/CBO9781107415324.004

Hood, C. (1991). All Seasons ? *Public Administration*, *69*(1), 3–19.

Hood, C. (2006). Gaming in targetworld: The targets approach to managing british public services. *Public Administration Review*, *66*(4), 515–521.

Huang, F. (2016). *Changes and challenges to higher education financing in Japan*. *Center of Global Higher Education working paper series*. Retrieved from https://www.researchcghe.org/perch/resources/publications/wp10.pdf

Huang, F. (2018). Higher education financing in Japan: Trends and challenges. *International Journal of Educational Development*, *58*, 106–115.

Hughes, O. (2010). Does governance exist? In S. P. Osborne (Ed.), *The New Public Governance?: Emerging Perspectives on the Theory and Practice of Public Governance* (pp. 87–104). New York: Routledge.

Hurst, J., & Williams, S. (2012). Can NHS hospitals do more with less?, (January), 92. Retrieved from http://www.nuffieldtrust.org.uk/sites/files/nuffield/can-nhs-hospitals-do-more-with-less_full-report-120112.pdf

Hyndman, N., & Connolly, C. (2011). Accruals accounting in the public sector: A road not always taken. *Management Accounting Research*, *22*(1), 36–45.

Hyndman, N., & Lapsley, I. (2016). New Public Management: The Story Continues. *Financial Accountability and Management*, *32*(4), 385–408.

Hyndman, N., & Liguori, M. (2016). Public Sector Reforms: Changing Contours on an NPM Landscape. *Financial Accountability and Management*, *32*(1), 5–32.

Jacobs, R. (2001). Alternative Methods to Examine Hospital Efficiency: Data Envelopment Analysis and Stochastic Frontier Analysis. *Health Care Management Science*, *4*(2), 103–115.

Jacobs, R., Smith, P., & Street, A. (2006). *Measuring efficiency in health care: Analytic techniques and health Policy*. Cambridge: Cambridge University Press.

Jindal, R. P., Gauri, D. K., Singh, G., & Nicholson, S. (2018). Factors influencing hospital readmission penalties: Are they really under hospitals' control ? *Decision Support Systems*, *110*, 58–70.

Johnes, G., & Tone, K. (2017). The efficiency of higher education institutions in England revisited: comparing alternative measures. *Tertiary Education and Management*, *23*(3), 191–205.

Johnes, J. (2015). Operational research in education. *European Journal of Operational Research*, *243*(3), 683–696.

Jones, P., Wells, S., Harper, A., Le Fevre, J., Stewart, J., Curtis, E., … Ameratunga, S. (2017). Impact of a national time target for ED length of stay on patient outcomes. *New Zealand Medical Journal*, *130*(1455), 15–34.

Jones, R. (2011). Hospital bed occupancy demystified. *British Journal of Healthcare*

Management, *17*(6), 242–248.

Jones, Rodney. (2001). Bed occupancy in acute and mental health hospitals . *Health Service Journal*, *111*(5752), 28–31.

Lampe, H. W., & Hilgers, D. (2015). Trajectories of efficiency measurement: A bibliometric analysis of DEA and SFA. *European Journal of Operational Research*, *240*(1), 1–21.

Liu, J. S., Lu, L. Y. Y., & Lu, W. M. (2016). Research fronts in data envelopment analysis. *Omega*, *58*, 33–45.

Liu, J. S., Lu, L. Y. Y., Lu, W. M., & Lin, B. J. Y. (2013). Data envelopment analysis 1978-2010: A citation-based literature survey. *Omega*, *41*(1), 3–15.

Kaneko, M. (2012). Incorporation of National Universities in Japan: An Evaluation Six Years On. In H. ans G. Chuetze, W. illiam Bruneau, & G. Grosjean (Eds.), *University Governance and Reform: Policy , Fads , and Experience in International Perspective* (pp. 1689–1699). New York: Palgrave Macmillan.

Kaya Samut, P., & Cafrı, R. (2016). Analysis of the Efficiency Determinants of Health Systems in OECD Countries by DEA and Panel Tobit. *Social Indicators Research*, *129*(1), 113–132.

Keegan, A. D. (2010). Hospital bed occupancy: More than queuing for a bed. *Medical Journal of Australia*, *193*(5), 291–293.

Keene, L., Bagshaw, P., Nicholls, M. G., Rosenberg, B., Frampton, C., & Powell, I. (2016). Funding New Zealand's Public healthcare system: time for an honest appraisal and public debate. *New Zealand Medical Journal*, *129*(1435), 10–20.

Kelly, E., Soye, G., & Vera-hernadez, M. (2016). Public hospital spending in England: Evidence from National Health Service administrative records. *Fiscal Studies*, *37*(3–4), 433–459.

Kempkes, G., & Pohl, C. (2010). The efficiency of German universities - some evidence from nonparametric and parametric methods. *Applied Economics*, *42*(16), 2063–2079.

Kettl, D. F. (2005). *The Global Public Management Revolution* (2nd ed.). Virginia: The brookings institution.

Kirigia, J. M., & Asbu, E. Z. (2013). Technical and scale efficiency of public community hospitals in Eritrea: An exploratory study. *Health Economics Review*, *3*(1), 1–16.

Klijn, E. H., & Koppenjan, J. (2012). Governance network theory: Past, present and future. *Policy and Politics*, *40*(4), 587–606.

Knopf, E. (2017). *History of Efficiency Measurement by the New Zealand Health Sector: Post 2000*. Retrieved from https://www.productivity.govt.nz/sites/default/files/Knopt%282017%29 History of efficiency measures in Health sector.pdf

Kohl, S., Schoenfelder, J., Fügener, A., & Brunner, J. O. (2018). The use of Data Envelopment Analysis (DEA) in healthcare with a focus on hospitals. *Health Care Management Science*. https://doi.org/10.1007/s10729-018-9436-8

Kounetas, K., & Papathanassopoulos, F. (2013). How efficient are Greek hospitals? A

case study using a double bootstrap DEA approach. *European Journal of Health Economics*, *14*(6), 979–994.

Krippendorff, K. (2004). *Content Analysis: An Introduction to Its Methodology* (2nd ed.). London: Sage Publications, Inc.

Lampropoulou, M., & Oikonomou, G. (2018). Theoretical models of public administration and patterns of state reform in Greece. *International Review of Administrative Sciences*, *84*(1), 101–121.

Lanfond, S. (2015). Funding overview: Current NHS spending in England. *The Health Foundation*. Retrieved from https://www.health.org.uk/publication/current-nhs-spending-england

Latham, R., & Prowle, M. (2012). *Public Services and Financial Austerity: Getting Out of the Hole?* London: Palgrave Macmillan.

Lau, E., Lonti, Z., & Schultz, R. (2017). Challenges in the Measurement of Public Sector Productivity in OECD 180 Countries. *International Productivity Monitor*, *32*, 180–195.

Le Grand, J. (2007). *The Other Invisible Hand: Delivering Public Services through Choice and Competition*. *Journal of Chemical Information and Modeling* (Vol. 53). New Jersey: Princeton University Press.

Lee, B. L. (2011). Efficiency of Research Performance of Australian Universities: A Reappraisal using a Bootstrap Truncated Regression Approach. *Economic Analysis and Policy*, *41*(3), 195–203.

Levy, R. (2010). New public management: End of an era? *Public Policy and Administration*, *25*(2), 234–240.

Lewis, J. M. (2015). The politics and consequences of performance measurement. *Policy and Society*, *34*(1), 1–12.

Lewis, Richard, Smith, J., & Harrison, A. (2009). From quasi-market to market in the National Health Service in England: What does this mean for the purchasing of health services? *Journal of Health Services Research and Policy*, *14*(1), 44–51.

Lewis, Ruth, & Edwards, N. (2015). *Improving length of stay : what can hospitals do ?* Retrieved from https://www.nuffieldtrust.org.uk/files/2017-01/improving-length-of-stay-hospitals-web-final.pdf

Licchetta, M., & Stelmach, M. (2016). *Fiscal sustainability analytical paper: Fiscal sustainability and public spending on health*. Retrieved from http://obr.uk/docs/dlm_uploads/Health-FSAP.pdf

Lodge, M., & Gill, D. (2011). Toward a New Era of Administrative Reform? The Myth of Post-NPM in New Zealand. *Governance: An International Journal of Policy, Administration, and Institutions*, *24*(1), 141–166.

Lodge, M., & Hood, C. (2012). Into an age of multiple austerities? Public management and Public Service Bargains across OECD countries. *Governance*, *25*(1), 79–101.

Lord Carter of Coles. (2016). *Operational productivity and performance in English NHS acute hospitals: Unwarranted variations*. Retrieved from

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/499229/Operational_productivity_A.pdf

Lovell, C. A. K., Walters, L. C., & Wood, L. L. (1994). Stratified Models of Education Production Using Modified DEA and Regression Analysis. In A. Charnes, W. Cooper, A. Lewin, & L. Seiford (Eds.), *Data Envelopment Analysis: Theory, Methodology, and Application* (pp. 329–351). Springer Science+Business Media New York.

Mainichi, T. (2019). 16 % of nat ' l universities looking to mergers to streamline management : survey. Retrieved from https://mainichi.jp/english/articles/20190106/p2a/00m/0na/003000c

Maniadakis, N., & Thanassoulis, E. (2000). Assessing productivity changes in UK hospitals reflecting technology and input prices. *Applied Economics*, *32*(12), 1575–1589.

Marshall, L., Charlesworth, A., & Hurst, J. (2014). *The NHS payment system: evolving policy and emerging evidence Research report*. Retrieved from https://www.nuffieldtrust.org.uk/files/2017-01/2014-nhs-payment-research-report-web-final.pdf

Matranga, D., Bono, F., Casuccio, A., Firenze, A., Marsala, L., Giaimo, R., … Vitale, F. (2014). Evaluating the effect of organization and context on technical efficiency: A second-stage dea analysis of Italian hospitals. *Epidemiology Biostatistics and Public Health*, *11*(1), 1–11.

Mayring, P. (2000). Qualitative Content Analysis. *Forum: Qualitative Social Research*, *1*(2).

Mayring, P. (2015). Qualitative Content Analysis: Theoretical Background and Procedures. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Approaches to qualitative research in mathematics education: examples of methodology and methods* (pp. 365–380). Dordrecht: Springer.

Mccallion, G., Glass, J. C., Jackson, R., Kerr, C. A., & Mckillop, D. G. (2000). Investigating productivity change and hospital size: A nonparametric frontier approach. *Applied Economics*, *32*(2), 161–174.

McCallion, G., McKillop, D. G., Glass, J. C., & Kerr, C. (1999). Rationalizing northern ireland hospital services towards larger providers: Best-practice efficiency studies and current policy. *Public Money and Management*, *19*(2), 27–32.

McDonald, J. (2009). Using least squares and tobit in second stage DEA efficiency analyses. *European Journal of Operational Research*, *197*(2), 792–798.

McKendry, C. ., Howard, P. ., & Carryer, B. . (1994). *New Zealand hospital sector performance 1983-1992*. Retrieved from http://www.moh.govt.nz/notebook/nbbooks.nsf/0/0c35611388f46d0f4c2565d70018e458/$FILE/Hosp Sector.pdf

MEXT. (2003). White Paper on Education, Culture, Sports, Science and Technology. Retrieved from http://www.mext.go.jp/b_menu/hakusho/html/hpac200301/index.html

MEXT. (2009). *Higher Education in Japan*. Retrieved from http://www.mext.go.jp/en/policy/education/highered/title03/detail03/__icsFiles/afieldfile/2012/06/19/1302653_1.pdf

MEXT. (2010). *The present conditions and problems after incorporation of national university*. Retrieved from http://www.mext.go.jp/a_menu/koutou/houjin/1295896.htm

Miles, M., & Huberman, M. (1994). *Qualitative data analysis: An expanded sourcebook* (2nd ed.). London: Sage Publications, Inc.

Ministry of Health. (2008). *Health targets: Moving towards healthier futures, 2007/08 the results*. Retrieved from https://www.health.govt.nz/system/files/documents/publications/health-targets-0708-results-nov08_0.pdf

Ministry of Health. (2014). *Statement of Intent 2014 to 2018*. Wellington: Ministry of Health.

Ministry of Health. (2016a). Criteria on quality assessment of Vietnamese hospitals, 2.0 version [Bộ tiêu chí đánh giá chất lượng Việt Nam, phiên bản 2.0). Retrieved from http://kcb.vn/wp-content/uploads/2016/11/Bộ-Tiêu-chí-Chất-lượng-bệnh-viện-Việt-Nam-năm-2016_-BẢN-CHÍNH-THỨC-pdf.pdf

Ministry of Health. (2016b). *New Zealand Health Strategy: Future direction*. Retrieved from https://www.health.govt.nz/publication/new-zealand-health-strategy-2016

Ministry of Health. (2017). The New Zealand Health and Disability System : Handbook of Organisations and Responsibilities. Retrieved from https://www.beehive.govt.nz/sites/default/files/2017-12/Health - Organisations and Responsibilities_0.pdf

Ministry of Health. (2018). New Zealand heath system. Retrieved from https://www.health.govt.nz/new-zealand-health-system

Ministry of Health. (2019). About the health targets. Retrieved from http://www.health.govt.nz/new-zealand-health-system/health-targets/about-health-targets

Mirozumi, A. (2019). Higher Education Reform: Focusing on National University Reform. In Y. Kitamura, T. Omomo, & M. Katsuno (Eds.), *Education in Japan: A Comprehensive Analysis of Education Reforms and Practices* (pp. 197–210). Springer Nature Singapore.

Mizobata, S., & Yoshii, M. (2015). Restructuring of the Higher Educational System in Japan. In J. C. Brada, W. Bienkowski, & M. Kuboniwa (Eds.), *International Perspectives on Financing Higher Education* (pp. 25–50). London: Palgrave Macmillan.

Monitor. (2014). *Facing the future: smaller acute providers*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/320075/smalleracuteproviders-report.pdf

Monitor. (2015). *2015 / 16 National Tariff Payment System : A consultation notice*. Retrieved from

https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/379076/NTPS_ConsultationNotice_26Nov.pdf

Monitor. (2017). NHS foundation trust accounts. Retrieved from https://www.gov.uk/government/publications/nhs-foundation-trust-accounts-consolidation-ftc-files-201617

Moradi-Motlagh, A., Jubb, C., & Houghton, K. (2016). Productivity analysis of Australian universities. *Pacific Accounting Review*, *28*(4), 386–400.

Morris, J. (2018). The growing problem of treatment waiting times. Retrieved from https://www.nuffieldtrust.org.uk/files/2017-01/improving-length-of-stay-hospitals-web-final.pdf

Moynihan, D. P. (2008). *The dynamic of performance managemnt: Constructing information and reform*. Washington, D.C: Georgetowm University Press.

National Institute of Population and Social Security Research. (2017). *Population Projections for Japan*. Retrieved from http://www.ipss.go.jp/index-e.asp

National NHS Staff Survey Co-ordination Centre. (2018). NHS staff survey. Retrieved from http://www.nhsstaffsurveys.com/Page/1019/Past-Results/Staff-Survey-2016-Detailed-Spreadsheets/

National University Hospital Council of Japan. (2014). Searching for the Future Vision of NUH. Retrieved May 25, 2019, from http://www.univ-hosp.net/english/vision/hospital/

Nedelea, I. C., & Fannin, J. M. (2013). Technical efficiency of Critical Access Hospitals: An application of the two-stage approach with double bootstrap. *Health Care Management Science*, *16*(1), 27–36.

Nemoto, J., & Furumatsu, N. (2014). Scale and scope economies of Japanese private universities revisited with an input distance function approach. *Journal of Productivity Analysis*, *41*(2), 213–226.

New Zealand Productivity Commission. (2018). Improving State Sector Productivity, (August). Retrieved from https://www.productivity.govt.nz/sites/default/files/Improving State Sector Productivity_Final Report_FINAL ONLINE.pdf

New Zealand Treasury. (2013). *Affording Our Future: Statement on New Zealand's Long-Term Fiscal Position*. Retrieved from https://treasury.govt.nz/publications/ltfp/affording-our-future-statement-new-zealands-long-term-fiscal-position

Newby, H., Weko, T., Breneman, D., Johanneson, T., & Maassen, P. (2009). *OECD Reviews of Tertiary Education Japan*. Paris: OECD Publishing.

NHS Choice. (2016). The NHS in England. Retrieved from https://www.nhs.uk/NHSEngland/thenhs/records/healthrecords/Pages/overview.aspx

NHS Digital. (2017). Hospital Activity. Retrieved from https://digital.nhs.uk/search/document-type/publication/publicationStatus/true?area=data&sort=date

NHS England. (2014). *Five years forward review*. Retrieved from https://www.england.nhs.uk/wp-content/uploads/2014/10/5yfv-web.pdf

NHS England. (2017a). Bed availability and overnight occupancy. Retrieved from https://www.england.nhs.uk/statistics/statistical-work-areas/bed-availability-and-occupancy/

NHS England. (2017b). *Referral to treatment (RTT) waiting times statistics for consultant-led elective care: 2016-17 Annual Report*. *NHS England*. Retrieved from https://www.england.nhs.uk/statistics/wp-content/uploads/sites/2/2017/06/RTT-Annual-Report-2016-17-v0.9-final.pdf

NHS Improvement. (2016). *NHS foundation trusts: Consolidated Accounts 2015/16*. Retrieved from http://www.monitor-nhsft.gov.uk/sites/default/files/NHS Foundation Trusts Consolidated Accounts 1011 website file.pdf

NHS Improvement. (2017a). National schedule of reference costs. Retrieved from https://improvement.nhs.uk/resources/reference-costs/

NHS Improvement. (2017b). *NHS foundation trusts: Consolidated Accounts 2016/17* (Vol. 331). Retrieved from https://improvement.nhs.uk/about-us/corporate-publications/publications/nhs-foundation-trusts-consolidated-accounts-201617/

NICE. (2017). *Chapter 39 Bed occupancy*. Retrieved from https://www.nice.org.uk/guidance/ng94/documents/draft-guideline-39

Nolan, P. (2018). Measuring Productivity in the Service Sector. *Policy Quarterly*, *14*(3), 40–45. Retrieved from https://www.victoria.ac.nz/__data/assets/pdf_file/0004/1686136/Nolan.pdf

Nuffieldtrust. (2014). *Length of stay case study*. Retrieved from https://www.nuffieldtrust.org.uk/public/event/reducing-length-of-stay

O'Neill, L., Rauner, M., Heidenberger, K., & Kraus, M. (2008). A cross-national comparison and taxonomy of DEA-based hospital efficiency studies. *Socio-Economic Planning Sciences*, *42*(3), 158–189.

OECD.Stat. (2019a). Enrolment by type of school. Retrieved from https://stats.oecd.org/

OECD.Stat. (2019b). OECD's statistical databases. Retrieved from https://stats.oecd.org/Index.aspx?DatasetCode=HEALTH_STAT

OECD/EU. (2016). *Health at a Glance: Europe 2016 - State of Health in the EU Cycle*. https://doi.org/10.1787/9789264265592-en

OECD. (2006). *Whole-of-Government Approaches to Fragile States*. Paris: OECD Publishing. https://doi.org/10.1787/journal_dev-v8-art39-en

OECD. (2011). *OECD Economic Surveys: Japan 2011*. OECD Publishing. https://doi.org/https://dx.doi.org/10.1787/eco_surveys-jpn-2011-en

OECD. (2017a). *Health at a Glance 2017: OECD Indicators*. Paris: OECD Publishing. https://doi.org/http://dx.doi.org/10.1787/health_glance-2017-en

OECD. (2017b). OECD Health Statistics 2017. Retrieved from https://stats.oecd.org/Index.aspx?DataSetCode=HEALTH_REAC

OECD. (2018a). *Education at a Glance 2018: OECD Indicators*. Paris: OECD Publishing.

https://doi.org/10.1787/eag-2018-en

OECD. (2018b). *Education Policy in Japan: Building Bridges towards 2030*. Paris: OECD Publishing.

OECD. (2019). Number of students (indicator). https://doi.org/doi: 10.1787/efa0dd43-en

OECD, Eurostat, & World Health Organization. (2017). *A System of Health Accounts 2011* (Revised ed). Paris: OECD Publishing. https://doi.org/http://dx.doi.org/10.1787/9789264270985-en

Ongaro, E., Ferré, F., & Fattore, G. (2015). The fiscal crisis in the health sector: Patterns of cutback management across Europe. *Health Policy*, *119*(7), 954–963.

Osborne, S. P. (2006). The new public governance? *Public Management Review*, *8*(3), 377–387.

Ozcan, Y. A. (2014). *Health Care Benchmarking and Performance Evaluation: An Assessment using Data Envelopment Analysis (DEA)* (Second Edi, Vol. 210). NewYork: Springer US.

Papadimitriou, M., & Johnes, J. (2018). Does merging improve efficiency? A study of English universities. *Studies in Higher Education*, 1–21. https://doi.org/10.1080/03075079.2018.1450851

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).

Peters, B. G. (2011). Governance responses to the fiscal crisis-comparative perspectives. *Public Money and Management*, *31*(1), 75–80.

Pidd, M. (2012). *Measuring the Performance of Public Services: Principles and Practice*. Cambridge: Cambridge University Press.

Poister, T. H. (2003). *Measuring Performance in Public and Nonprofit Organizations*. San Francisco: Jossey-Bass.

Poister, T. H., Aristigueta, M. P., & Hall, J. L. (2014). *Managing and measuring performance in public and nonprofit organizations: An intergrated approach* (2nd ed.). San Francisco: Jossey-Bass.

Pollitt, C. (2003). Joined-up Government : a Survey. *Political Studies Review*, *1*, 34–49.

Pollitt, C. (2010). Cuts and reforms – public services as we move into a new era. *Society and Economy*, *32*(1), 17–31.

Pollitt, C. (2013). The logics of performance management. *Evaluation*, *19*(4), 346–363.

Pollitt, C. (2016). Be prepared? An outside-in perspective on the future public sector in Europe. *Public Policy and Administration*, *31*(1), 3–28.

Pollitt, C., & Bouckaert, G. (2017). *Public Management Reform* (4th ed.). London: Oxford University Press.

Powell, M., Dawson, J., Topakas, A., Durose, J., & Fewtrell, C. (2014). Staff satisfaction and organisational performance: evidence from a longitudinal secondary analysis of the NHS staff survey and outcome data. *Health Services and Delivery Research*,

2(50), 1–306.

Powell, M., & Exworthy, M. (2016). Never again? A retrospective and prospective view of English health reform. In M. Exworthy, R. Mannion, & M. Powell (Eds.), *Dismantling the NHS ? Evaluating the impact of health reforms* (pp. 365–380). Bristol: Policy Press.

Powell, T. (2016). *The structure of the NHS in England*. Retrieved from https://researchbriefings.parliament.uk/ResearchBriefing/Summary/CBP-7206

Quirora-Martínez, F., Fernández-Vázquez, E., & Alberto, C. L. (2018). Efficiency in public higher education on Argentina 2004 – 2013: institutional decisions and university - specific effects. *Latin American Economic Review*, *27*(14), 1–18.

Reiter, R., & Klenk, T. (2018). The manifold meanings of 'post-New Public Management' – a systematic literature review. *International Review of Administrative Sciences*, *0*(0), 1–17.

Roberts, A., Marshall, L., & Charlesworth, A. (2012). *A decade of austerity? Nuffield Trust*. Retrieved from http://www.nuffieldtrust.org.uk/publications/decade-austerity-funding-pressures-facing-nhs

Robertson, R., Wenzel, L., Thompson, J., & Charles, A. (2017). *Understanding NHS financial pressures - How are they affecting patient care?* Retrieved from https://www.kingsfund.org.uk/sites/default/files/field/field_publication_file/Understanding NHS financial pressures - full report.pdf

Rosenberg, B., & Keene, L. (2016). Did the 2017 Budget provide enough for Health ?, (16), 1–13. Retrieved from http://www.union.org.nz/wp-content/uploads/2016/06/Did-the-Budget-provide-enough-for-Health-2016.pdf

Rouse, P., & Swales, R. (2006). Pricing public health care services using DEA: Methodology versus politics. *Annals of Operations Research*, *145*(1), 265–280.

Şamiloğlu, F., & Akgün, A. İ. (2016). The Relationship between Working Capital Management and Profitability: Evidence from Turkey. *Business and Economics Research Journal*, *7*(2), 1–14.

Sandiford, P., Consuelo, D. J. J. V., & Rouse, P. (2017). How efficient are New Zealand's District Health Boards at producing life expectancy gains for Māori and Europeans? *Australian and New Zealand Journal of Public Health*, *41*(2), 125–129.

Sav, G. T. (2013). Effects of Financial Source Dependency on Public University Operating Efficiencies: Data Envelopment Single-Stage and Tobit Two-Stage Evaluations. *Review of Economics & Finance*, *3*, 63–73.

Sav, G. T. (2016). Recession and Post-Recession Efficiency and Productivity Changes in United States Public Universities: The Good, Bad, and Ugly. *Advances in Management & Applied Economics*, *6*(3), 1–15.

Schick, A. (2013). Reflections on Two Decades of Public Fianancial Management Reforms. In M. Cangiano, T. Curristine, & M. Lazare (Eds.), *Public Financial Management and Its Emerging Architecture* (pp. 21–78). Washington, D.C: International Monetary Fund.

Scott, R., & Boyd, R. (2016). Results, Targets and Measures to Drive Collaboration:

Lessons from the New Zealand Better Public Services reforms. In D. Gilchrist & J. Butcher (Eds.), *The three sector solution: Delivering public policy in collaboration with not-for-profits anf business* (pp. 235–260). Australian National University: ANU Press.

Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of Econometrics*, *136*(1), 31–64.

Simar, L., & Wilson, P. W. (2011). Two-stage DEA: Caveat emptor. *Journal of Productivity Analysis*, *36*(2), 205–218.

Smith, P. (1995). On the unintended consequences of publishing performance data in the public sector. *International Journal of Public Administration*, *18*(2–3), 277–310.

Sorenson, C., Drummond, M., & Khan, B. B. (2013). Medical technology as a key driver of rising health expenditure: disentangling the relationship. *ClinicoEconomics and Outcomes Research*, *5*, 223–234.

Starke, P. (2010). Why institutions are not the only thing that matters: Twenty-five years of health care reform in New Zealand. *Journal of Health Politics, Policy and Law*, *35*(4), 487–516.

Statistics Bureau. (2015). *2015 Population Census*. Retrieved from https://www.stat.go.jp/english/data/kokusei/index.html

Sueyoshi, T., Yuan, Y., & Goto, M. (2017). A literature study for DEA applied to energy and environment. *Energy Economics*, *62*, 104–124

Takundwa, R., Jowett, S., McLeod, H., & Peñaloza-Ramos, M. C. (2017). The Effects of Environmental Factors on the Efficiency of Clinical Commissioning Groups in England: A Data Envelopment Analysis. *Journal of Medical Systems*, *41*(6), 1–7.

Talbot, C. (2010). *Theories of performance: Organizational and Service Improvement in the Public Domain*. New York: Oxford University Press Inc.

Tenbensel, T., & Chalmers, L. (2016). Comparing the implementation consequences of the immunisation and emergency department health targets in New Zealand A tale of two targets, *30*(6), 1009–1024.

Tenbensel, T., Chalmers, L., Jones, P., Appleton-Dyer, S., Walton, L., & Ameratunga, S. (2017). New Zealand's emergency department target - Did it reduce ED length of stay, and if so, how and when? *BMC Health Services Research*, *17*(1), 1-15.

Thanassoulis, E., Kortelainen, M., Johnes, G., & Johnes, J. (2011). Costs and efficiency of higher education institutions in England: A DEA analysis. *Journal of the Operational Research Society*, *62*(7), 1282–1297.

Thanassoulis, Emanuel. (2001). *Introduction to the Theory and Application of Data Envelopment Analysis*. New York: Springer.

The Japan Association of National Universities. (2019). Internationalization. Retrieved from https://www.janu.jp/eng/national_universities/Internationalization.pdf

The Ministry of Health. (2014). New Zealand Health System, *4*(2). Retrieved from http://www.health.govt.nz/new-zealand-health-system

The Treasury. (2016). *Analysis of District Health Board Performance to 30 June 2015*.

Retrieved from http://www.treasury.govt.nz/publications/informationreleases/health/dhb-performance/dhb-performance-jun16.pdf

The Treasury. (2017a). *District Health Board Financial Performance to 2016 and 2017 Plans*. Retrieved from https://treasury.govt.nz/sites/default/files/2017-05/dhb-performance-feb17.pdf

The Treasury. (2017b). *Financial Statements of the Government of New Zealand for the Year Ended 30 June 2017*. Retrieved from https://treasury.govt.nz/publications/year-end/financial-statements-30-june-2017-html

TheKing'sFund. (2017). How the NHS is funded. Retrieved from https://www.kingsfund.org.uk/projects/nhs-in-a-nutshell/how-nhs-funded

Timmins, N. (2018). *' The World ' s Biggest Quango ': The First Five Years of NHS England*. Retrieved from https://www.kingsfund.org.uk/publications/worlds-biggest-quango-nhs-england

Umobong, A. (2015). Assessing the impact of liquidity anf profitability ratios on growth of profits in pharmaceutical firms in Nigeria. *European Journal of Accounting, Auditing and Finance Research*, *3*(10), 97–114.

UNESCO-WHED. (2018). *International Handbook of Universities*. UNESCO.

United Nation-Population Divison. (2019). Total population - both sexes. Retrieved from https://population.un.org/wpp/Download/Standard/Population/

Valdmanis, V., Rosko, M., Mancuso, P., Tavakoli, M., & Farrar, S. (2016). Measuring performance change in Scottish hospitals: a Malmquist and times-series approach. *Health Services and Outcomes Research Methodology*, *17*(2), 113–126.

VanThiel, S., & Leeuw, F. L. (2002). The Performance Paradox in the Public Sector. *Public Performance & Management Review*, *25*(3), 267-281.

Varabyova, Y., & Schreyögg, J. (2013). International comparisons of the technical efficiency of the hospital sector: Panel data analysis of OECD countries using parametric and non-parametric approaches. *Health Policy*, *112*(1–2), 70–79.

Verzulli, R., Jacobs, R., & Goddard, M. (2018). Autonomy and performance in the public sector: the experience of English NHS hospitals. *European Journal of Health Economics*, *19*(4), 607–626.

Visbal-Cadavid, D., Martínez-Gómez, M., & Guijarro, F. (2017). Assessing the efficiency of public universities through DEA. A case study. *Sustainability*, *9*(8), 1–20.

Vitikainen, K., Linna, M., & Street, A. (2010). Substituting inpatient for outpatient care: What is the impact on hospital costs and efficiency? *European Journal of Health Economics*, *11*(4), 395–404.

Vizard, P., & Obolenskaya, P. (2015). *The Coalition ' s Record on Employment : Policy , Spending and Outcomes 2010-2015*. Retrieved from http://sticerd.lse.ac.uk/dps/case/spcc/WP15.pdf

Wachter, R., Slee, A., & Brailer, D. (2016). *Making IT work: harnessing the power of health information technology to improve care in England*. Retrieved from

https://www.gov.uk/government/publications/using-information-technology-to-improve-the-nhs/making-it-work-harnessing-the-power-of-health-information-technology-to-improve-care-in-england

Willing, E. (2016). Hitting the target without missing the point: New Zealand's immunisation health target for two year olds. *Policy Studies*, *37*(6), 535–550.

Wolszczak-Derlacz, J. (2017). An evaluation and explanation of (in)efficiency in higher education institutions in Europe and the U.S. with the application of two-stage semi-parametric DEA. *Research Policy*, *46*(9), 1595–1605.

Wolszczak-Derlacz, J., & Parteka, A. (2011). Efficiency of European public higher education institutions: A two-stage multicountry approach. *Scientometrics*, *89*(3), 887–917.

World Health Organization. (2014). *New Zealand Health System Review*. *New Zealand Health System Review* (Vol. 4). Retrieved from https://apps.who.int/iris/handle/10665/207738

World Population Review. (2019). The largest countries in the world by area. Retrieved from http://worldpopulationreview.com/countries/countries-in-world-by-area/

Worthington, A. C., & Lee, B. L. (2008). Efficiency , technology and productivity change in Australian universities , 1998 – 2003. *Economics of Education Review*, *27*, 285–298.

Yamada, R. (2014). Japanese Higher Education: Policies and Future Issues. In R. Yamada (Ed.), *Measuring Quality of Undergraduate Education in Japan* (pp. 17–33). Springer Singapore.

Yamamoto, K. (2009). Public sector management reform in Japan. In S. F. Goldfinch & J. L. Wallis (Eds.), *International Handbook of Public Management Reform* (pp. 336–350). Northampton: Edward Elgar.

Yamato, B. (2018). The Internationalization of Japanese Higher Education: Incremental Change in a Dynamic Global Environment. In A. Yonezawa, Y. Kitamura, B. Yamato, & T. Tokunaga (Eds.), *Japanese Education in a Global Age* (pp. 221–240). Singapore: Springer.

Zeidler, J., Mittendorf, T., Vahldiek, G., Zeidler, H., & Merkesdal, S. (2008). Comparative cost analysis of outpatient and inpatient rehabilitation for musculoskeletal diseases in Germany. *Rheumatology*, *47*(10), 1527–1534.

Zhu, J. (2016). *Data Envelopment Analysis: A Handbook of Empirical Studies and Applications*. New York: Springer